EARTHJUSTICE
BECAUSE THE EARTH NEEDS A GOOD LAWYER

BIODIVERSITY DEFENSE PROGRAM
INFO@EARTHJUSTICE.ORG
EARTHJUSTICE.ORG

September 27, 2024

Horseshoe Crab Management Board
Atlantic States Marine Fisheries Commission
1050 N. Highland Street, Suite 200 A-N
Arlington, VA 22201
comments@asmfc.org

VIA ELECTRONIC MAIL

Re: ASMFC's "Technical Response to External Review of the 2022 ARM Framework Revision"

Dear Members of the Horseshoe Crab Management Board:

New Jersey Audubon and Defenders of Wildlife urge the Atlantic States Marine Fisheries Commission ("ASMFC" or the "Commission") to maintain the prohibition on the bait harvest of female Delaware Bay-origin horseshoe crabs. The attached report by Dr. Kevin Shoemaker reaffirms that the Commission's adaptive resource management ("ARM") model fails to represent the relationship between red knots[1] and horseshoe crabs, underestimates the risks to both species, and is not suitable for determining bait harvest quotas. The ARM model therefore cannot legitimately serve as a basis for resuming the female bait harvest, and its recommendation for a female harvest should not be adopted.

Dr. Shoemaker has prepared two prior analyses of the ARM model: first during the public comment period in 2022, which was held before the model's computer code was publicly available, and again in 2023 after the computer code was released and analyzed by Dr. Shoemaker. Both of his analyses identified critical flaws demonstrating the grave risks that utilizing the ARM model would pose for the fragile Delaware Bay ecosystem. ASMFC responded to the merits of those analyses for the first time in April 2024, and Dr. Shoemaker addresses that response in his new report attached to this letter.[2]

In addition to Dr. Shoemaker's analyses, more than 34,000 members of the public opposed adopting the new model and resuming a female horseshoe crab harvest during the 2022 comment period, compared to only seven commenters in support. The public expressed concern about horseshoe crabs and the species that rely upon them, including the red knot, a

---

[1] In these comments, "red knot" refers to the *rufa* subspecies unless otherwise noted.

[2] Dr. Shoemaker's new report is attached as Exhibit A. Dr. Shoemaker's 2022 and 2023 analyses (hereinafter "Shoemaker 2022" and "Shoemaker 2023") are available at https://earthjustice.org/wp-content/uploads/2023/09/nj-audubon-defenders-of-wildlife-2023-comments-to-hsc-board.pdf. ASMFC's "Technical Response to External Review of the 2022 ARM Framework Revision" appeared in the Horseshoe Crab Management Board's spring 2024 meeting materials.

shorebird that migrates up to 17,000 miles every year and requires horseshoe crab eggs as a crucial energy source. In 2015, red knots were listed as a threatened species under the federal Endangered Species Act ("ESA"), with the overharvest of horseshoe crabs identified as a key contributor to their decline. If ASMFC authorized a bait harvest of female horseshoe crabs that reduced the food source available to migrating red knots, it would risk violating the ESA by depriving red knots of essential nutrition and thereby committing "take" of this threatened shorebird.

In his attached report, Dr. Shoemaker has carefully assessed ASMFC's response and demonstrated that it does not undermine his core conclusions. Critically, the model fails to accurately represent red knots' reliance on horseshoe crabs. It would not predict a decline in red knots even under a collapse of the horseshoe crab population, and it ignores horseshoe crab egg surveys, which are much more closely linked to red knot survival than the data inputs used by the model. The model also significantly overestimates red knots' survival rate—and ASMFC has misread or misconstrued many of the studies that it relies on to support its erroneously high estimate. In the few instances where ASMFC's claims provided a legitimate basis for Dr. Shoemaker to update his prior analyses, he has done so. Nevertheless, his updated analysis continues to demonstrate significant flaws in the ARM model.

This cover letter describes key points from Dr. Shoemaker's analysis and raises other concerns with the ARM model, including ASMFC's shifting strategies for gap-filling the extremely low estimates of newly mature female horseshoe crabs, which offer additional reasons that the model-generated female harvest recommendation should not be adopted. While elements of Dr. Shoemaker's analysis are summarized below, please refer to his attached report for his complete response.

I. The ARM model would fail to predict a decline in red knots even under a collapse of the horseshoe crab population.

At the outset, a key conclusion that Dr. Shoemaker reached two years ago holds true today and continues to counsel against relying on the ARM model to set harvest quotas: the model fails to accurately reflect the relationship between the red knot and horseshoe crab populations. In his 2022 analysis, Dr. Shoemaker evaluated the weak relationship between red knots and horseshoe crabs in the ARM model and calculated that the model would predict an increase in red knots passing through Delaware Bay even if horseshoe crabs disappeared entirely from the region.[3] This finding raised concerns about the model's ability to predict future declines in red knot abundance in Delaware Bay, including under new proposed horseshoe crab harvest scenarios, as it would not have predicted the historical decline that occurred in the wake of severe horseshoe crab overharvest in the late 20th century. Because ASMFC held its 2022 public comment period on the model at a time when the federal government was denying repeated requests to release the model's computer code to the public for independent review, Dr. Shoemaker by necessity based this finding on a back-of-the-envelope calculation, as he repeatedly noted in his analysis.[4]

---

[3] Shoemaker 2022 at 6-12.

[4] *Id.* at 7, 9.

ASMFC's April 2024 response nevertheless criticizes Dr. Shoemaker based on technical information that was not available to the public when he conducted his analysis. Regardless, the points raised in the April 2024 response are misguided. The response contains two principal contentions. First, with the benefit of the computer code, it is evident that an increase in red knots when there are zero horseshoe crabs in Delaware Bay is "mathematically impossible."[5] But this argument misses the point. The importance of Dr. Shoemaker's critique is not merely that the ARM model would be inadequate if horseshoe crab numbers actually reached zero, but that the model fails to represent red knots' response generally across a wide range of horseshoe crab abundance, including abundance figures that have been historically observed. Further, while the model would not predict an increase in red knots if the horseshoe crab population were literally zero, ASMFC has not—and could not—deny that the model *would* predict an increase in red knots at breathtakingly low horseshoe crab abundance levels indicating an ecosystem collapse.

ASMFC neglected to provide the precise horseshoe crab abundance threshold at which the model would begin to predict a decline in red knots at Delaware Bay, so Dr. Shoemaker reran his analysis using the model's computer code to answer that question. He calculated that the model would not predict a decline in red knot abundance unless the number of mature female horseshoe crabs in Delaware Bay fell below approximately 300,000—less than a tenth of the lowest number ever estimated from empirical data. Of course, red knot abundance plummeted when the relevant crab population actually reached that prior low. Yet the ARM model predicts that red knot abundance would remain stable even if the horseshoe crab population plunged dramatically lower still. Thus, for management purposes, whether the model begins to show a decline in red knots at zero or 300,000 female horseshoe crabs is immaterial. The material fact is that the model cannot accurately predict the red knot population response to horseshoe crab harvest scenarios such as the female harvest recommendation that is now being considered.

ASMFC's second argument is to accuse Dr. Shoemaker of conducting a "dangerous exercise"[6] for running a scenario well outside of the ARM model's training data. Furthermore, ASMFC forecasts unanimous support for curtailing the horseshoe crab harvest under such dire conditions in which the horseshoe crab population plummeted. Again, this misses the point, which is that the model would fail to predict a decline in red knots even under conditions that have been historically observed to cause such a decline. If the model is intended to be functional only within limited bounds of female horseshoe crab abundance, ASMFC should specify as much—especially if the model cannot function within the full range of historically observed conditions. Speculation that fisheries managers would intervene under catastrophic circumstances, even if well founded, does not alter the conclusion that the ARM model fails to accurately represent the environmental conditions that it purports to reflect.

II. The ARM model significantly overstates red knot survival rates.

The ARM model is also plagued by critical reliance on an assumed survival rate for red knots that is insupportably high. Dr. Shoemaker explained that the ARM model's finding that red

---

[5] ASMFC Response 26.

[6] *Id.*

knots have a 93% survival rate is likely erroneously high.[7] He hypothesized that this error resulted from relatively rare but consequential mistakes in the dataset. Specifically, the survival rate formula is based largely on resighting observations—birds that are spotted over multiple years, as identified by leg flags bearing unique codes that can be read from a distance without requiring physical recapture. However, the difficulty of reading leg flags from afar gives rise to the possibility of error. If the same leg flag is spotted more than once in a season, the subsequent sightings help to verify the initial identification, and there is a high likelihood that the bird was truly present in Delaware Bay. Conversely, flag codes spotted only once in a season (approximately 9% of total resighting observations) lack that verification and carry a higher probability that they were misreads. These misreads are likely to bias the estimated survival rate higher because the birds bearing those flag codes may be dead and are mistakenly recorded as living longer than they did, potentially by many years.

In his 2023 analysis, Dr. Shoemaker recalculated the red knot survival rate with the same dataset used by ASMFC but excluded birds that were resighted only once in a season. He found that the survival rate plunged to around 80%. He also calculated the survival rate exclusively from birds whose leg bands were read upon recapture—when misreads are likely to be negligible—and again calculated around 80%. The difference in survival rates has profound consequences: with ASMFC's likely erroneous survival rate, the average red knot would live nearly 14 years, but using the more realistic survival rate, the average lifespan drops to less than 5 years.

ASMFC's April 2024 analysis makes no attempt to refute or explain the discrepancy between the ARM model's survival rate and the survival rate calculated with more verified data. Instead, it undermines its own position by presenting data that directly support Dr. Shoemaker's findings. ASMFC's response states, "[O]bservations of birds more than 5 years old are common in the mark-recapture data set (approximately 20% of birds), with a maximum of 17 years between physical recaptures."[8] But as Dr. Shoemaker explains in his attached report, those figures are consistent with (if not lower than) what would be expected with an 80% survival rate. In contrast, under a 93% survival rate—as assumed by ASMFC—70% of birds would survive to age 5, and more than 2% would survive past 17 years. Yet ASMFC does not report any such results from the mark-recapture data, because they do not exist. Instead, ASMFC appears to have inadvertently raised the question of why, if the survival rate is 93%, there are so few red knots that are confirmed to be at least 5 years old.

ASMFC's next defense of its high survival rate estimate in the April 2024 response is to point to scientific publications, including Piersma et al. (2016), which studied a different subspecies of red knot (*Calidris canutus piersmai*) in Australia. While studies of a different subspecies across the world cannot substitute for a rigorous interpretation of the data collected at Delaware Bay, they may be informative. But Piersma et al. does not support ASMFC's

---

[7] In its April 2024 response, ASMFC implies that the ARM model found a survival rate of 90%, but the actual figure is 93%. ASMFC Response 6; ASMFC, *Revision to the Framework for Adaptive Management of Horseshoe Crab Harvest in the Delaware Bay Inclusive of Red Knot Conservation (Draft for Board Review)* 74 (2021). While the discrepancy may seem trivial, it amounts to a four-year difference in red knots' mean expected lifespan.

[8] ASMFC Response 6.

conclusions, and ASMFC appears to have misinterpreted the study. ASMFC asserts that Piersma et al. found "annual apparent survival for red knots in Western Australia were well above 90% in most years of their study."[9] Yet the study says no such thing: for most years of the study, the annual apparent survival percentage rate hovered in the 80s; it never reached 90%, and in the final two years, it plummeted to 76% and 67%.[10] (ASMFC may have confused *annual* survival rates with *seasonal* survival rates, which were also discussed in the study.) Moreover, Piersma et al. attributed the plunging survival rate observed in its study to habitat loss in a key staging area. Thus, the study found that red knot survival rates were *never* as high as ASMFC stated, and in fact the study supports the conclusion that problems at a staging area—like Delaware Bay for the *rufa*—can harm the species.

ASMFC's April 2024 response then references another scientific study (also of non-*rufa*), Boyd & Piersma (2001), for the proposition that some red knots have long lifespans—which, as explained above, is not in dispute and would be expected even under lower survival rates. Confoundingly, ASMFC's response fails to disclose that the study also estimated mean adult survival of red knots using two different methods, both of which yielded estimates below 80% over the duration of the study.[11] Again, ASMFC's response erroneously claims support from a scientific publication that does not support ASMFC's conclusions, and, to the contrary, supports Dr. Shoemaker's analysis. More fundamentally, the Commission fails to square its defense of a 93% red knot survival rate with the contrary data reported in the very studies cited in ASMFC's own response.

A third article that ASMFC cites, Tucker et al. (2022), was authored predominantly by researchers who collaborated to create the ARM model[12] and used the same method of counting singlet observations that Dr. Shoemaker critiques. The study and the ARM model made the same error and thereby generated similar results. The study therefore does not provide independent validation of the ARM model's methodology or estimated survival rate. All told, of the five studies that ASMFC cites to support a higher survival rate, three of them either directly refute ASMFC's position or replicate the ARM model's contested approach.

In addition, the scientific evidence for a red knot survival rate far lower than 93% continues to grow. A new study of red knots wintering in Texas, Louisiana, and Florida found mean apparent annual survival rates of 76.8%, 81.9%, and 79.0%, respectively.[13] Further, Amie MacDonald of Birds Canada recently presented research estimating that the true annual survival for adult red knots staging in Canada's James Bay is 81%.[14] Concerningly, both of

---

[9] *Id.*

[10] Theunis Piersma et al., *Simultaneous declines in summer survival of three shorebird species signals a flyway at risk*, Journal of Applied Ecology vol. 53, 479, at 486 tbl. 5 (Apr. 2016).

[11] Hugh Boyd & Theunis Piersma, *Changing Balance Between Survival and Recruitment Explains Population Trends in Red Knots Calidris Canutus Islandica Wintering in Britain, 1969-1995*, Ardea vol. 89(2) 301, at 307 tbl. 2 (Jan. 2001).

[12] *Compare* ASMFC Response 1 (listing contributors to response) *with id.* at 31 (listing authors of Tucker et al.). Anna Tucker, Conor McGowan and James Lyons appear in both places.

[13] David J. Newstead et al., *Survival of red knots in the northern Gulf of Mexico*, Frontiers in Ecology and Evolution, at 7 tbl. 2 (Apr. 9, 2024) (attached as Exhibit B).

[14] Amie MacDonald et al., *Uniting* rufa *Red Knot resighting data throughout the western Atlantic Flyway offers myriad opportunities for survival analysis* 24, PowerPoint presentation (2024) (attached as Exhibit C).

these studies, like Piersma et al. (2016), found survival rates declining significantly over time. The red knot survival rate utilized in the ARM model is out of step with these research findings.

III. ASMFC provides no compelling reason to exclude horseshoe crab egg density surveys from the ARM model.

ASMFC's April 2024 response does not dispute Dr. Shoemaker's analysis that egg density—the concentration of horseshoe crab eggs on the beach—has a significant positive correlation to red knot survival. And ASMFC expressly (and accurately) "does not deny that eggs are the true link between horseshoe crabs and red knots."[15] Nevertheless, to explain the omission of egg density from the ARM model, ASMFC's response states, "Ultimately, egg density data could not be considered in the ARM Revision because they were not provided to the ARM Subcommittee when requested."[16]

Whatever data availability issues may have arisen previously, ASMFC presents no evidence that they persist. Moreover, ASMFC may have been at least partly responsible for any past availability issues: when excluding egg density data from the prior version of the ARM model, ASMFC made no mention of data availability and wrote, "We do not foresee using the egg survey data in our models or in our decision analysis in the foreseeable future, and we place low priority on continuing this survey and researching/improving survey methodologies."[17] In 2013, at the request of the state of Delaware, ASMFC dropped egg density surveys as a compliance requirement.[18] Thus, any data availability issues that ASMFC previously encountered may have arisen at least partly from ASMFC's own actions. In light of that history, it is especially inappropriate for ASMFC to criticize Dr. Shoemaker's analysis on the basis that it did not include egg density data from Delaware.[19] But regardless, now that the data availability concerns appear resolved, ASMFC's statement that it "is not opposed to using the egg density data" is welcome.[20]

While ASMFC describes challenges associated with incorporating egg density data into the ARM model, there is no justification for continuing to rely exclusively on measures (horseshoe crab trawl surveys) that bear minimal correlation to red knot survival while ignoring measures (egg surveys) that bear a strong correlation. The technical challenges raised by ASMFC center on the lack of a modeled connection between egg density and female horseshoe crab abundance, which the Commission acknowledges "must ultimately be linked."[21] More research into this system would be beneficial, but that should not prevent or

---

[15] ASMFC Response 11.

[16] *Id.*

[17] ASMFC, *A Framework for Adaptive Management of Horseshoe Crab Harvest in the Delaware Bay Constrained by Red Knot Conservation* 40 (Sept. 2009).

[18] *See* ASMFC Horseshoe Crab Delaware Bay Ecosystem Technical Committee, Meeting Summary 3 (Sept. 24, 2013) (recommending to discontinue egg surveys as a requirement); ASMFC, Proceedings of the Atlantic States Marine Fisheries Commission Horseshoe Crab Management Board 16-17 (Oct. 31, 2013) (formally removing egg surveys as a requirement).

[19] ASMFC Response 12.

[20] *Id.*

[21] *Id.*

delay ASMFC from including egg surveys in the ARM model. Indeed, the connection between female horseshoe crab abundance and red knot survival must logically include egg availability as an intermediate step. Thus, by modeling red knot survival as a linear function of horseshoe crab abundance, ASMFC implicitly assumes that horseshoe crab abundance strongly corresponds to egg availability. Although there are mechanistic questions about that link, ASMFC has nonetheless attempted to model the connection between horseshoe crab abundance and red knot survival. Given the availability of long-term egg survey data, the case is at least as strong for explicitly modeling the connection between red knot survival, egg density, and female horseshoe crab abundance.

IV. Dr. Shoemaker has reaffirmed his analysis of uncertainty in the ARM model and updated his assessment of trends in female horseshoe crab abundance.

As detailed in his attached report, Dr. Shoemaker has considered ASMFC's response regarding technical flaws in the horseshoe crab catch multiple survey analysis ("CMSA") model. Two aspects of that report bear noting here.

First, ASMFC acknowledged that Dr. Shoemaker's critique of how the CMSA model propagates uncertainty has merit and should be considered in future ARM revisions.[22] That acknowledgment is welcome, although it is inappropriate for ASMFC to continue using the inferior method pending some future revision. The Commission seeks to downplay this issue by noting that Dr. Shoemaker's suggested method of propagating uncertainty produces a similar equilibrium number of mature female horseshoe crabs as the CMSA model's method. But an essential question when propagating uncertainty is whether the model appropriately recognizes the degree of uncertainty (e.g., 95% confidence interval) associated with various harvest scenarios. Dr. Shoemaker has demonstrated that the CMSA model does not.

The critique about propagating uncertainty stems from the CMSA's treatment of the recruitment rate for mature female horseshoe crabs. The recruitment rate is an important parameter in the CMSA model, and this type of uncertainty is called "parameter uncertainty." The model's flawed treatment of parameter uncertainty is separate from—and additional to—the ARM model's flawed treatment of structural uncertainty, which Dr. Shoemaker explains in his first opening statement in the attached report. ASMFC characterizes the model as adaptive resource management, but such management entails testing various hypotheses. The relative weight given to each hypothesis changes as new information is learned about the ability of each hypothesis to represent the system. In contrast, the ARM model incorporates only one hypothesis and excludes consideration of any alternatives. While the model may be updated every few years to reflect new data, iterative updates do not amount to adaptive resource management. Under this flawed approach, the model never has to earn the 100% confidence value it is given, and ASMFC's asserted commitment to adaptive resource management is illusory. It is critical for the model to recognize all types of uncertainty when representing the ecosystem.

Second, Dr. Shoemaker reviewed ASMFC's response regarding trends in mature female horseshoe crab abundance. In his prior analyses, he observed that there had not been a

---

[22] *Id.* at 23.

statistically significant increase in such crabs since 2000, the first year when state-based harvest quotas became effective. ASMFC argued that the trend should be measured from 2010 to reflect the roughly ten years needed for female horseshoe crabs to reach maturity.[23] Dr. Shoemaker agreed that 2010 is a non-arbitrary threshold and re-ran his analysis from that year. He confirmed that the data from each of the three trawl surveys that inform the CMSA show apparent increases in adult female horseshoe crab abundance since 2010. Nevertheless, this finding is based only on the data reported from the trawl surveys and does not resolve concerns about the reliability of those surveys or the methodology for incorporating the data into the CMSA and the larger ARM model.

V. Unsupported estimates of newly mature female horseshoe crabs further compromise the ARM model's harvest recommendations.

The ARM model's recommendations are further undermined by the modelers' reliance on speculative estimates of a key data point needed to make the model work at all. Since 2019, the estimated abundance of newly mature female horseshoe crabs in the Delaware Bay population has been alarmingly low—even as low as zero. ASMFC has explained that the CMSA cannot operate with such low recruitment numbers and has devised various methods to "gap-fill" that data input.[24] Last year, the ARM Subcommittee and Delaware Bay Ecosystem Technical Committee ("DBETC") hypothesized that the low newly mature female horseshoe crab numbers did not reflect a true recruitment failure but rather a classification error, and responded by "re-proportion[ing]" 19.9% of the mature female estimate to the newly mature age class.[25] Their management recommendation to the Board, which included a substantial female bait harvest, was based on that recommendation.

However, at its meeting on September 13, 2024, the ARM Subcommittee and DBETC concluded that the previous hypothesis was incorrect. Accordingly, they now hypothesize that surveyors had misclassified newly mature female horseshoe crabs as immature, not fully mature, and they propose to gap-fill the newly mature female estimate by reallocating a quantum of immature female horseshoe crabs equal to a designated percentage of the newly mature *male* abundance estimate.

As of this writing, ASMFC has not released a written explanation of the new method, but it appears to lead to troubling results. Because the method will be applied retroactively, it will significantly increase the gap-filled estimates used since 2019, resulting in much larger population estimates. While the committees may have reason to believe that the newly mature females were misclassified as immature rather than fully mature, that does not mean that the estimate should be gap-filled based on surveys of newly mature males. How newly mature females were possibly misclassified is a separate question from how their abundance should be estimated.

---

[23] *Id.* at 14.

[24] Memorandum from Delaware Bay Ecosystem Technical Committee and Adaptive Resource Management Subcommittee to Horseshoe Crab management Board re: "Delaware Bay Horseshoe Crab Harvest Recommendation for 2024" 2 (Oct. 2, 2023).

[25] *Id.*

More troublingly, the whiplash over newly mature female estimates demonstrates the peril of patching up the ARM model with speculative hypotheses even as it is being used to generate management recommendations. For the second consecutive year, ASMFC cannot credibly claim that it is running a peer-reviewed model because a significant function of the model has been assembled on the fly. What ASMFC believed to be the best hypothesis last year was immediately utilized for management recommendations and turned out to be erroneous. Now a new hypothesis is being substituted without peer review or any meaningful public scrutiny. And because it is already known that the newly mature female estimate will be zero again next year, the pattern of utilizing unproven methods to generate harvest recommendations will continue.

This new development compounds a long history of ASMFC using unfounded estimates of newly mature female horseshoe crabs in its modeling analyses. In his 2022 analysis, Dr. Shoemaker observed that the ARM model's estimate of horseshoe crab recruitment was strongly influenced by nonsensical estimates that ASMFC plugged in for the years 2013-2016, when the survey that measures newly mature females was not performed. The average annual estimated recruitment for 2003-2012 was 1.2 million newly mature females, and the average annual estimate for 2017-2019 was 1.9 million. But for 2013-2016, lacking the empirical measurement, ASMFC plugged in extraordinary estimates averaging *4.2 million*—nearly *2 million more than the highest empirical estimate ever recorded*.[26] That average masks even more absurd estimates for individual years, including *9.6 million* in 2013.[27]

Understandably, ASMFC's peer reviewers for the ARM model specifically emphasized that estimates of newly mature females needed to become more reliable over time by utilizing empirical counts.[28] And in its April 2024 response to Dr. Shoemaker, ASMFC acknowledged that the CMSA's volatile recruitment estimates for 2013-2016 were "nonsensical."[29] But the use of nonsensical, unempirical estimates has persisted well beyond anything that the peer reviewers contemplated. Under ASMFC's latest method for gap-filling the missing recruitment data, the estimates return to nonsensical territory, swinging from 8.2 million in 2020 to 1.3 million in 2021 and back up to 6.5 million in 2022.[30] ASMFC offers little reason to believe that these wildly diverging estimates reflect actual biological reality in the Delaware Bay ecosystem.

As a result of all the foregoing gap-filling efforts, the model now significantly deviates from the version that was peer-reviewed, both by the absence of actual newly mature female data and by the increased weight being placed on the newly mature male estimate. While all data inputs are imperfect, ideally the use of multiple inputs will balance out those imperfections.

---

[26] ASMFC, *Supplemental Report to the 2021 Revision to the Adaptive Resource Management Framework* 16 tbl. 3 (2022).

[27] *Id.*

[28] *See* ASMFC, *Horseshoe Crab Adaptive Resource Management Revision Peer Review Report*, at 270 of PDF ("[T]he primiparous estimates for [the missing survey] years are not reliable, potentially introducing large uncertainties (and biases) in the projection model and ARM. The Panel agrees that such uncertainty will be reduced when more years of survey catch data become available in future.").

[29] ASMFC Response 23.

[30] Again, these data were presented at the September 13, 2024, ARM Subcommittee and DBETC meeting. No written explanation or additional context has been released to the public.

But if one input (newly mature females) is based entirely on another (newly mature males), any errors in the latter input will be replicated in the former and compromise more of the model. The ARM model's peer reviewers did not approve placing so much weight on—or taking that risk with—the estimate of newly mature males. The appropriate course now is to exercise caution and not recommend a female bait harvest based on an untested, unreviewed model. And regardless of any recommendations that the model may produce, ASMFC certainly should not reauthorize a female harvest.

As noted above, the problem with the newly mature female horseshoe crab estimate has been recognized and acknowledged for many years. When ASMFC accepted comments on the ARM revision in 2022, commenters were already raising concerns about low estimates dating to 2019. Now ASMFC suggests, based on a discussion with surveyors, that newly mature females were simply not being counted. It is striking that ASMFC spent five years devising hypotheses if the explanation was so straightforward, and it seems emblematic of a serious disconnect between ASMFC's complex computer model and conditions on the ground.

VI. Conclusion

The ARM model contains fundamental flaws rendering it unfit for managing the harvest of Delaware Bay-origin horseshoe crabs. Now that ASMFC has responded to Dr. Shoemaker's analysis, it remains evident that the model does not accurately represent the ecosystem, and its outputs are not a defensible basis for imposing additional risk on red knots and horseshoe crabs in Delaware Bay. Most importantly, the model cannot justify—and ASMFC must continue to prohibit—the bait harvest of female horseshoe crabs. Longer term, ASMFC should discontinue using the ARM model or make fundamental improvements through a transparent public process.

Respectfully submitted,

Benjamin Levitan
Senior Attorney
Earthjustice Biodiversity Defense Program
(202) 797-4317
blevitan@earthjustice.org

EXHIBIT A

Report of Dr. Kevin Shoemaker

Review of the Atlantic States Marine Fisheries Commission's (ASMFC) Adaptive Resource Management (ARM) framework for regulating Horseshoe Crab bait harvest in Delaware Bay

Kevin T. Shoemaker, Ph.D.

Associate Professor, University of Nevada, Reno

September 2024

This document is submitted in reference to the Atlantic States Marine Fisheries Commission's (ASMFC) response to two peer review reports examining the 2021 revised ASMFC Adaptive Resource Management (ARM) framework – which has been approved for use in managing the Horseshoe Crab fishery in Delaware Bay. This document, and the 2022 and 2023 peer review reports referred to in the ASMFC response, were written by Kevin Shoemaker, Ph.D.

Dr. Shoemaker holds an M.S. and Ph.D. in Conservation Biology from SUNY-ESF in Syracuse, NY, and a B.S. degree in Biology from Haverford College. He was a Postdoctoral Fellow in the Department of Ecology and Evolution at Stony Brook University and has served as Senior Scientist at Applied Biomathematics, an ecological research and development company located in Setauket, NY. Dr. Shoemaker is currently an Associate Professor at the University of Nevada, Reno, where he uses quantitative models to inform wildlife conservation and management. He has over 15 years of experience as a wildlife ecologist and conservation modeler and has authored over 50 peer-reviewed scientific articles and book chapters on topics in ecology and conservation. He has expertise in Bayesian inference, population ecology, population viability analysis (PVA) and ecological modeling.

Overview

In 2022 and 2023 I was asked by Earthjustice (a not-for-profit public interest legal organization) to provide an independent peer review of the quantitative models used by the Atlantic States Marine Fisheries Commission (ASMFC) in their '*Revision to the Framework for Adaptive Management of Horseshoe Crab Harvest in the Delaware Bay Inclusive of Red Knot Conservation*' (ASMFC 2021; hereafter, 'revised ARM'). In both peer review reports, I identified several lines of evidence that suggest the revised ARM framework, as approved by ASMFC in 2022, underestimates the risk of commercial harvest of female horseshoe crabs on the Federally Threatened *rufa* Red Knot (*Calidris canutus* rufa) and other shorebirds that rely on the Delaware Bay staging area. Earlier this year (2024) ASMFC's ARM subcommittee released a statement in response to my review of the revised ARM (hereafter, "ASMFC response"). In this document I explain why my analysis generally holds up to the critiques raised in the ASMFC response, and highlight a couple of areas where the ASMFC response led me to reconsider my original conclusions. My overall conclusion remains the same: the revised ARM fails to recognize evidence that commercial harvest of female horseshoe crabs could harm the red knot and other migratory shorebird populations, nor does it acknowledge the extent to which recovery of the red knot population may be tied to the growth of the Delaware Bay horseshoe crab population. Therefore, the revised ARM framework does not live up to its stated mission to "*Manage harvest of horseshoe crabs in the Delaware Bay to maximize harvest but also to maintain ecosystem integrity, provide adequate stopover habitat for migrating shorebirds, and ensure that the abundance of horseshoe crabs is not limiting the red knot stopover population or slowing recovery*" (ASMFC 2021).

I begin with a short essay (opening statement #1) arguing that the revised ARM framework failed to implement a key component of the Adaptive Resource Management (ARM) paradigm: multiple competing hypotheses. Not only does the revised ARM framework assume the relationship between red knots and horseshoe crabs is extremely weak, there are several compelling lines of evidence (including the re-analyses I presented in my 2023 peer review report) that this ecological relationship may in fact be much stronger than the "weak relationship" hypothesis that is currently formalized in the revised ARM. By assigning 100% of model weight to a "weak relationship" model -- whereby harvest of female horseshoe crabs is expected to have negligible impact on red knots -- the revised ARM misrepresents the risk of horseshoe crab harvest to red knots in contradiction with well-established science. Therefore, I argue that the ARM framework should be revised to incorporate at least one additional model that acknowledges the possibility of a strong and meaningful ecological relationship between red knots and horseshoe crabs.

Following this initial opening statement, I provide a revised analysis (opening statement #2, responding to ASMFC's comments on an analysis presented in my 2022 peer review report) that reinforces an important assertion from my reviews of the revised ARM framework -- specifically, that the fitted relationship between horseshoe crab abundance and red knot vital rates (survival and fecundity) is of insufficient magnitude to induce a decline in projected red knot population growth even under a major collapse of the horseshoe crab population. This point is central to my critique of the current ARM framework, as it clearly demonstrates that (1) the model is incapable of predicting the observed decline of red knots in the early 21st century, which is widely attributed to over-harvest of horseshoe crabs, and (2) the modeled relationship between red knots and horseshoe crabs is too weak to meaningfully constrain harvest recommendations of female horseshoe crabs. Finally, I provide a point-by-point response to ASMFC's comments.

Before I respond to the specific critiques raised by ASMFC, I emphasize that my peer review was motivated by the same stated principles that guide the ARM subcommittee: a commitment to science-based decision making in natural resources conservation and management. I reject the implication that my perspectives were infused with advocacy, or that my peer review reflected a "reluctance to learn within an adaptive management framework and a desire to cling to previous beliefs in spite of scientific advances". To the contrary, in the interest of encouraging productive scientific dialog, I reached out to the ASMFC reviewers soon after they began their review with an offer to share code and information and address any questions or concerns directly -- and although they did not respond, I would be happy to engage with the ARM subcommittee to discuss any of these issues in more depth. While I was compensated for my time by Earthjustice, no one attempted to exert any influence over my scientific conclusions, and my comments should be received in the spirit they were offered: as an independent scientific evaluation of the revised ARM framework. As a quantitative ecologist and conservation biologist, I promote the use of data and simulation models in support of conservation decisions, and I believe in the value of adaptive management for making decisions in the face of uncertainty.

Opening statement #1: *the revised ARM framework fails to account for structural uncertainty by incorporating multiple alternative hypotheses*

Under the adaptive resource management (ARM) paradigm, regular monitoring of the managed system enables decision makers to (1) react to new information (e.g., reducing or eliminating harvest quotas after observing population declines) and (2) update their assumptions and understanding of the managed system, learning from mistakes and reinforcing successes to continually develop improved management recommendations (Nichols et al. 2007; Williams 2011; Runge 2011). Furthermore, the objectives and other key premises of the system (data sources, monitoring protocols, allowable management actions, etc.) are revisited periodically: a process commonly known as the "double loop" (Williams et al. 2011; ASMFC 2021). Adaptive management, when properly applied, is central to science-based management of natural systems. However, I argue that the revised ARM (and ASMFC's response to my peer review reports) fails to embrace a core feature of the adaptive resource management (ARM) paradigm: the incorporation of multiple alternative hypotheses (Williams 2011). That failure results in a misrepresentation of the risk of horseshoe crab harvest to red knots and a missed opportunity to learn about the system.

In any ARM problem there is an inherent trade-off between maximizing the rate of learning and minimizing the risk of harming or destabilizing the system (Runge 2011). For example, we might be able to learn more about the resilience of the horseshoe crab population and the ecological dependency of red knots on horseshoe crabs by harvesting as many female horseshoe crabs as possible and then closely monitoring the population response of both species to this disturbance. In contrast, placing a moratorium on commercial harvest of female horseshoe crabs may reduce the learning rate but it also minimizes the risk of imperiling or impeding the recovery of a threatened species. It seems clear that the risk calculus must shift to some extent when a threatened or endangered species (TES) is part of the equation (Runge 2011), as is the case for the horseshoe crab harvest in Delaware Bay (involving a federally listed shorebird). A fully precautionary approach might lead to paralysis (possibly precluding beneficial conservation actions), while an opposing strategy that prioritizes action in the face of substantial risk to TES would risk irrevocable consequences. By formally embracing multiple alternative hypotheses, the ARM paradigm offers a compelling middle ground (Runge 2011).

In a multi-hypothesis ARM framework, each alternative model formalizes a plausible alternative hypothesis about how the focal system works (Williams 2011; Runge 2011). This enables ARM frameworks to accommodate structural uncertainty: one of the key sources of uncertainty that must be considered in natural resources management (Williams 2011). Together, the ensemble of models represents the current state of scientific knowledge (including a range of plausible hypotheses and assumptions) and captures the uncertainty and risks inherent to a managed natural system. Each alternative model is assigned a weight, or confidence value, that reflects its current standing relative to the other models included in the ARM framework. The weights assigned to each model at each successive decision point reflect each model's current degree of empirical support (the degree to which it effectively predicts current and historical system states and the observed response to prior management actions) and the degree to which the model captures the prior beliefs and risk tolerances of the stakeholder community (Williams 2011; Runge 2011).

By contrast, in the revised Delaware Bay ARM framework, a single hypothesis is effectively assigned a confidence value of 100%. Under this hypothesis, the relationship between horseshoe crab abundance and red knot demographic rates is so weak that it has little to no practical relevance to the dynamics of this system, as documented in this report (below) and in my 2022 peer review report. I will refer to this as the "weak relationship" hypothesis. My reanalysis, in which I detected a strong link between horseshoe crab egg densities and red knot survival (documented in my 2023 peer review report), along with numerous other published studies and government reports (e.g., Niles et al. 2009; USFWS 2014), provide evidence that the biotic interaction between horseshoe crabs and red knots may be substantially stronger and more ecologically meaningful than the ASMFC's model suggests. I will refer to this as the "strong relationship" hypothesis. The "strong relationship" hypothesis (unlike the "weak relationship" hypothesis) is capable of explaining the observed decline of the *rufa* red knot in the early years of the 21st century, for which the unregulated exploitation of horseshoe crabs in Delaware Bay is widely believed to be a primary cause (Niles et al. 2009; USFWS 2014). To accommodate structural uncertainty under the multi-hypothesis ARM paradigm (Williams 2011), it seems clear that a "strong relationship" model should be incorporated as a plausible hypothesis, and assigned some degree of credibility.

Furthermore, given the overwhelmingly negative public response to the prospect of harvesting female horseshoe crabs, it appears that the risk tolerance of the revised ARM may not be well aligned with that of the broader stakeholder community. By adding a plausible "strong relationship" model to the ARM framework, and by assigning an initial weight to this model that reflects diverse stakeholder perspectives, the ARM subcommittee could retain a robust, science-based management framework while also satisfying the many shorebird advocates within the stakeholder community that their perspectives are being formally considered and appropriately weighted. If the "weak relationship" model offered by ASMFC proves a more robust predictor of the future dynamics of this managed system relative to the "strong relationship" model and any other plausible alternative models, then this "weak relationship" model (the dominant hypothesis under the current ARM framework) will accrue a high credibility value over time and will come to dominate future recommendations for horseshoe crab harvest.

Regardless of the problematic issues with the original ARM framework that motivated the development of the revised ARM framework (documented in ASMFC 2021), the original ARM framework incorporated several alternative plausible hypotheses, including a weak, moderate, and strong biotic linkage between horseshoe crabs and red knots, respectively (McGowan et al. 2015). Furthermore, the original ARM framework used a formal stakeholder elicitation process to assign initial model weights to these models, ultimately leading ASMFC to assign substantial model weights to the moderate and strong interaction models, despite the fact that their empirical analyses suggested a much weaker relationship (McGowan et al. 2015). For reasons I do not fully understand, ASMFC abandoned a multi-hypothesis approach in developing their revised ARM framework. ASMFC supplied several reasons why they believe the revised ARM framework was an improvement over the original; for example, they point out some inadequacies and technology limitations with the previous framework, and highlight the fact that the revised ARM framework makes extensive use of empirical data from Delaware Bay (ASMFC 2021). However, none of these factors precludes the use of a multi-hypothesis ARM framework: for example, formally incorporating one ore more hypotheses that mechanistically link horseshoe crab surface egg densities (for which long-term data are available) with red knot demography.

In the conclusion of their response, ASMFC criticized my peer-review reports for failing to include concrete suggestions for improvement. While offering specific solutions was not a primary objective of my peer review reports, I will offer one suggestion: I encourage the ARM subcommittee to work with other independent researchers and the stakeholder community to develop an ARM framework that formally incorporates alternative plausible hypotheses about the strength of this two-species interaction. There is a well-developed literature that provides concrete recommendations for implementing the multi-hypothesis ARM paradigm. Although there are several data gaps and challenges to address, the explicit mechanism linking horseshoe crabs to red knots must be formally recognized: red knots depend on horseshoe crab eggs available near the ground surface, which requires perturbation of egg masses deposited by sufficient numbers of spawners prior to or concurrent with the arrival of red knot migrants. These mechanistic linkages will greatly benefit from the incorporation of available data on horseshoe crab surface egg densities as well as spawning counts and egg mass counts if available. Although some of the linkages in this system remain uncertain, the spirit of ARM encourages modelers and stakeholders to confront uncertainty by developing a comprehensive program for iterative learning through constructive and well-conceived actions. Following the above discussion, the multi-hypothesis ARM paradigm offers a compelling solution for making well-considered decisions in the face of uncertainty, while continually gaining new insights about how the system works. The ingredients are in place for a well-designed, multi-hypothesis ARM framework for this system and I hope ASMFC rises to this challenge.

Opening statement #2: *the relationship between horseshoe crabs and red knots in the revised ARM framework is exceptionally weak*

In my peer review reports I have paid particularly close attention to the strength of the relationship between red knot demography and horseshoe crabs, as this relationship is in many ways the crux of the matter: if the relationship is weak, then harvesting female horseshoe crabs is not a major issue for red knots, and if the relationship is strong, then red knot populations may suffer or their recovery may be stifled. My decision to focus on the strength of this relationship was not because of some preconceived bias or “clinging to belief” (as ASMFC claims in the conclusion of their response) but because this relationship is so important that it deserves special scrutiny. One of the most important issues I raised in my 2022 peer review report was that the relationship between red knot demographic rates (survival and recruitment) and horseshoe crab abundance (later published in Tucker et al. 2023) was so weak that changes in the horseshoe crab population would (under this model) have a negligible effect on the viability of the red knot population. Consequently, the revised ARM framework appeared unsuitable as a tool for making projections and contributing to policy decisions concerning management of this two-species system.

In their response, ASMFC criticized the back-of-the-envelope calculations in my 2022 report, noting that my calculations (performed before I gained access to the data and code for the red knot IPM) failed to recognize that ASMFC had log-transformed the horseshoe crab abundance values prior to incorporating these values in their integrated population model (IPM) (for my original report I used the raw values instead of the log-transformed values). This argument by ASMFC has more to do with mathematical technicalities than with ecology, and their objection is ultimately immaterial. When I run the same calculation with the log-transformed relationship, the conclusion remains the same: under the revised ARM framework, it would take a massive collapse of the horseshoe crab population (well under 0.5 million female horseshoe crabs across Delaware Bay) to cause a decline of the red knot population using mean parameter values from the red knot IPM (see below). Therefore, I do not concede that I was “wrong” (or “prejudicial”) on this issue in any of my analyses, as ASMFC claimed in their response under the “Criticism 8” header (below).

Prompted by the ASMFC review, I revised my original calculations to reflect the log-transformation used in the red knot demographic model – specifically, modeling red knot demographic rates as a function of log-transformed horseshoe crab abundance (Tucker et al. 2023). I generated figures illustrating these demographic effects to validate that they matched the relationships displayed in ASMFC 2021 and Tucker et al. (2023). After verifying a match (Fig. 1, left panels), I used this model to extrapolate the expected red knot survival and recruitment rates at very low horseshoe crab abundances (approaching zero) (Fig. 1, right panels). I then computed the expected population growth rate (lambda) for horseshoe crab abundances ranging from near-complete collapse (e.g., 1000 female horseshoe crabs) to recovery (around 20 million females), where values of lambda greater than or equal to one indicate a sustainable or growing red knot population.

These tests demonstrate that, under the revised ARM framework, red knot populations are expected to exhibit strong and sustained growth (lambda > 1) across all but the most extreme scenarios of horseshoe crab collapse (Fig. 2): red knot population growth would only be expected to exhibit mean net declines (lambda < 1) if the number of female horseshoe crabs in the Delaware Bay region fell below around 300,000 (the lowest recorded estimate from the last two decades places the number of females at around 4 million). Accordingly, the substance of my critique remains valid: the relationship between red knots and horseshoe crabs that was formalized within the revised ARM framework is exceptionally weak. Furthermore, this exercise demonstrates that the ARM model would not have predicted the decline of red knots due to horseshoe crab overharvest in the 1990s (which remains the dominant hypothesis for this observed population decline), which calls into question its usefulness in making projections and contributing to policy decisions that could help both species recover.

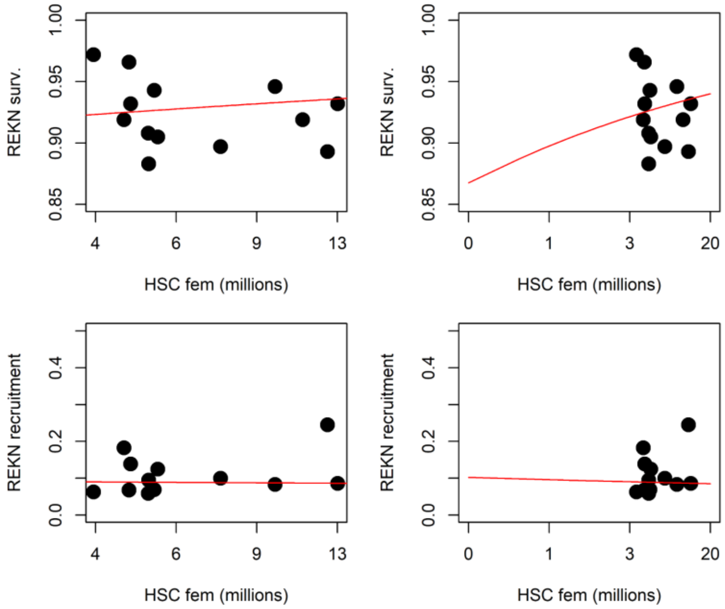


***Figure 1**. Visualizations of Red Knot (REKN) survival (y axis, top panels) and recruitment (y axis, lower panels) as a function of horseshoe crab abundance (x axis, all panels), derived from ASMFC's ARM model, later published as Tucker et al. (2023). Left-hand panels replicate Figure 4 from Tucker et al. (2023), whereas right-hand panels extend the x-axis to visualize these relationships at levels of horseshoe crab abundance ranging from well under 1 million (near-complete collapse of the stock) up to 20 million (an approximation of full recovery).*

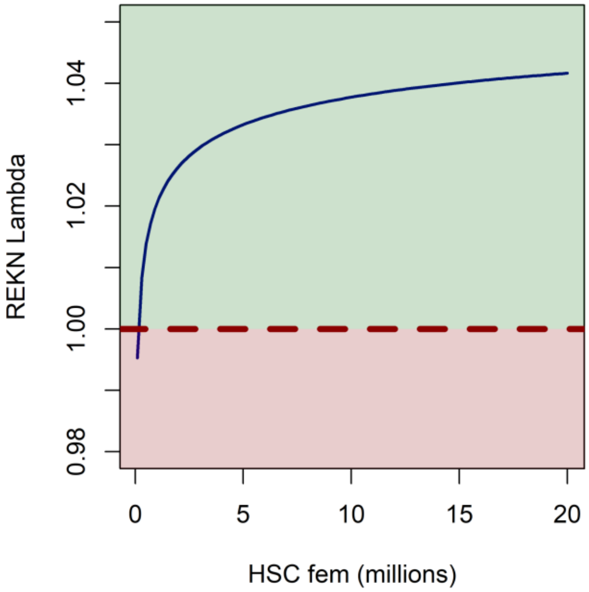


***Figure 2**. "Back of the envelope" illustration of the relationship between Red Knot (REKN) population growth, Lambda (y axis) and female horseshoe crab abundance (x axis, in millions), derived from ASMFC's ARM model, and published in Tucker et al. (2023). The*

range of the x-axis is intended to visualize the expected growth of the REKN population for horseshoe crab abundances ranging from well under 1 million (near-complete collapse of the stock) up to 20 million (an approximation of full recovery). Lamba ≥ 1 (green shaded region) represents a growing population whereas Lambda < 1 (red shaded region) represents a declining population. Under the revised ASMFC model, declines of the REKN population would only be expected under a near-complete collapse of the horseshoe crab population (total population less than 300,000 individuals across Delaware Bay).

Point-by-point response

NOTE: for the remainder of this document, all original text from the ASMFC response is in gray font, while my responses are indented and in dark green font. For clarity, I have removed some text from the original ASMFC response (for instance, historical summaries or overview statements) that I did not feel required a response. Also, I removed all figures from the ASMFC response- to view these figures, please refer to the original ASMFC response.

EXECUTIVE SUMMARY

While the ARM Revision represents significant advances in modeling and data use, the conversation around the revised ARM Framework quickly focused on the allowance of female horseshoe crab harvest when horseshoe crab population estimates are sufficiently high as to not limit red knot populations. The original ARM Framework had a technical flaw where it recommended 0 female horseshoe crab harvest when the adult female population was estimated to be less than 11.2 million, as it did from 2013-2022, or maximum female harvest (210,000 female horseshoe crabs) when the population was estimated to be greater than 11.2 million females, as it did in 2023. Rarely were the intermediate harvest levels selected by the model, as was shown through a simulation study. To correct this, the ARM Revision allowed a *gradual* increase of female harvest from 0-210,000 females as population estimates of female horseshoe crabs increased. The nuance of this change was lost in the discourse as stakeholders greatly opposed female harvest at any level, despite the original ARM Framework also recommending female harvest in recent years.

> Based on my revised tests, which are discussed in opening statement #2, “sufficiently high as to not limit red knot populations” in the context of the revised ARM framework means all levels of female horseshoe crab abundance except for extreme collapse of the horseshoe crab fishery (<=300,000 females; Fig. 2). I acknowledge that the revised ARM framework incorporated some improvements over the original ARM, but I was not tasked with reviewing the original ARM framework: since the revised ARM was formally approved in 2022, the revised ARM is now the legitimate subject of scrutiny.

Briefly, the ARM Subcommittee maintains that the red knot and horseshoe crab population models used in the ARM Framework currently represent the best use of the available data. Red knot survival rates and horseshoe crab population trends from the ARM Revision are consistent with other published values or data sources in the Delaware Bay region. This includes horseshoe crab egg density data, which were not provided to the ARM Subcommittee, but were subsequently published in the literature and show a similar trend to the horseshoe crab relative abundance indices.

> While the red knot demographic rates used in the revised ARM are consistent with some prior estimates, there are also many examples of lower survival rates in the published literature. While the previous literature is not conclusive on this point, in my re-analysis the Delaware Bay banding and resighting data support a survival rate of approximately 80%, much lower than the estimate of 93% that was used in the revised ARM framework. These two estimates have vastly different implications for the population ecology of this species, including the expected resilience of this population to horseshoe crab harvest and other threats (for example, a population with lower survival rates would likely be less resilient to a series of years with low resource availability) and the levels of recruitment that would be required to ensure population viability.
>
> I remain convinced that ASMFC's estimates of red knot survival are biased high due to the presence of misread errors in the resighting database. Perhaps the most convincing evidence for this is that survival estimates become substantially lower when "singlet" observations (resighting observations by a single observer at a single occasion, which are likely contaminated with flag misread errors) are dropped from the analysis. This method of subsetting the data has been suggested as a simple and effective technique for correcting potential biases in estimates of survival and survival trends due to misread errors (Tucker et al. 2019). If misread errors were not an issue, mean survival estimates should be similar whether or not these "singlet" observations are removed from the data set.

The ARM Subcommittee reiterates that an important benefit of the adaptive management process is the ability to make decisions even with imperfect knowledge of an ecological system. The overall goal of the ARM was to produce a decision-making framework informed by science and stakeholder values, given the available knowledge about the Delaware Bay ecosystem and horseshoe and red knot populations. At the time of the original ARM Framework, this knowledge was limited. However, the re-evaluation of the data, values, and knowledge on a regular basis is essential to the adaptive management process and is built into the ARM Framework. The 2022 ARM Revision represented a learning event where population models were re-designed to accommodate the advancement of data and knowledge since 2009. The peer reviews from Earthjustice fail to provide any real recommendations for improvement to the ARM Framework or provide other means for helping managers make an informed harvest decision beyond a mandate for zero female harvest at any population level. If the values of all stakeholders have changed (i.e., no female harvest under any circumstances), that change could be considered in a new approach in the future by the ARM Subcommittee. As it stands, the current ARM Framework represents the objectives previously established through stakeholder engagement: to manage harvest of horseshoe crabs in the Delaware Bay to maximize harvest but also to maintain ecosystem integrity, provide adequate stopover habitat for migrating shorebirds, and ensure that the abundance of horseshoe crabs is not limiting the red knot stopover population or slowing recovery.

First, I agree with the ARM subcommittee on the value of adaptive management for enabling informed decision making in the face of uncertainty and learning about the system via ongoing management and monitoring. However, navigating potential risks to a threatened or endangered species (TES) adds some complexity to the problem (as I discuss in opening statement #1 of this response). As I documented in my peer review reports, there are multiple lines of evidence suggesting that the revised ARM does not effectively account for the very real ecological risks of re-opening a commercial harvest on female horseshoe crabs. In failing to acknowledge the risks to red knots and the potential to jeopardize the recovery of this and other migratory shorebirds, the revised ARM appears to be mis-aligned with its own core objectives (from ASMFC 2021: "*Manage harvest of horseshoe crabs in the Delaware Bay to maximize harvest but also to maintain ecosystem integrity, provide adequate stopover habitat for migrating shorebirds, and ensure that the abundance of horseshoe crabs is not limiting the red knot stopover population or slowing recovery*".

As for the critique that my peer review failed to "provide any real recommendations for improvement to the ARM Framework or provide other means for helping managers make an informed harvest decision", my task as a peer reviewer was to evaluate the revised ARM on its merits rather than to develop an improved alternative. Nevertheless, in opening statement #1 above, I outline how a multi-hypothesis adaptive management framework could effectively incorporate alternative quantitative descriptions of the relationship between red knots and horseshoe crabs in Delaware Bay and reflect stakeholder perspectives, thereby representing a more legitimate approach to adaptive resource management than is reflected in the current version of the ARM framework.

Finally, I have read this review carefully and I thank ASMFC for their feedback. I continue to stand by the main conclusions from my 2022 and 2023 peer reviews -- with one notable exception: upon further consideration, ASMFC raised legitimate points regarding my use of linear regression to analyze the long-term trawl capture records for female horseshoe crabs. Nevertheless, the thrust of my critique of the revised ARM model remains valid. My re-analysis was offered as an independent evaluation of the data and was intended to contribute to a scientific dialog. In this spirit, I hope my peer-review reports and re-analyses contribute to ASMFC's ongoing efforts to understand and manage this system.

Criticism 1: Estimates of red knot survival used in the ARM appear to be artificially inflated, resulting in falsely optimistic estimates of population resilience.

- High survival and long lifespans are common for red knots and other shorebirds of similar size and life histories.
- Survival rates used in the ARM are calculated from the tagging data for red knots in the Delaware Bay region and are comparable with other published survival values.
- The tagging data were critically analyzed by the ARM Subcommittee to represent the

> best available data and caveats to the survival estimates were provided in the ARM Revision. The analysis of the tagging data and its use in the modeling was commended by the peer review panel.

Technical Response: Dr. Shoemaker asserts that red knot annual survival probability is more likely closer to 0.8 than the 0.9 used in the revised ARM Framework, corresponding to an expected lifespan of about 5 years. There is not strong evidence for this lower annual survival probability for *rufa* red knot. In fact, previous studies of *rufa* red knot in Delaware Bay (McGowan et al. 2011) and Florida (Schwarzer et al. 2012) also estimated annual survival probability at approximately 0.9. In a separate published analysis, only using data collected by the state of Delaware, Tucker et al. (2022) estimated red knot annual survival probability at 0.89, and at 0.91 for ruddy turnstones, a species with similar body size and a similar annual life cycle.

> The evidence I provided in my 2023 peer review report strongly suggests that the ASMFC's estimates of red knot survival are biased high, and average survival is closer to 80% in this system (versus 93% per the revised ARM framework). The primary evidence for this is that red knot survival estimates become much lower after "singlet" observations of flag codes (unconfirmed sightings that are likely to be contaminated with misread errors) are removed from the analysis, suggesting that flag misread errors are likely biasing the ASMFC's survival estimates high. A strong secondary line of evidence is that when the banding data are used as the sole source of information (these observations involve direct capture and are therefore much less likely to include misread errors), the mean survival estimate is again around 80%.
>
> Finally, I think it is important to note that the adult red knot survival estimate used by ASMFC averages 93%, not 90% as stated in the comment above. While this may seem like a trivial point, the difference between 93% survival (corresponding to median expected lifespan of around 9.5 years and mean lifespan of nearly 15 years) and 90% survival (corresponding to median lifespan of around 6.5 years and mean lifespan of nearly 10 years) can make the difference between growth and decline for many real-world populations.

Additionally, observations of birds more than 5 years old are common in the mark- recapture data set (approximately 20% of birds), with a maximum of 17 years between physical recaptures. These observations are a conservative minimum estimate of lifespan.

> This comment appears to confuse the concept of maximum lifespan with average lifespan. Even if median or expected lifespan is low, some fraction of individuals would be expected to reach more advanced ages; an expected lifespan of 5 years old does not preclude some fraction of individuals from reaching age 20 or beyond. At a constant survival rate of 80% (median lifespan of 3.1 years, mean lifespan approaching 5 years), we would expect more than 30% of individuals to live to age 5 and beyond (close to the "20% of birds" referred to in the above comment) and about 25% of individuals to live to age 6 and beyond (ignoring for simplicity that survival is likely to be lower in the first year of life). At a constant 80% survival rate, a little over 2% of birds would be expected to live past 17 years of age and

> around 1% would live as long as 20 years. With thousands of unique individuals in the database, we would expect to observe many cases of high longevity in the database even if the mean annual survival rate was approximately 80%.
>
> In contrast, the expected distribution of ages under a constant annual survival rate of 93% appears inconsistent with the Delaware Bay capture-recapture database: under this scenario, nearly 70% of individuals would be expected to survive to 5 years of age and beyond (far greater than the 20% cited in the above response by ASMFC), around 30% would reach 17 years of age and beyond, and around 3% would live to 50 years of age and beyond. Therefore, the information ASMFC cited above (i.e., that approximately 20% of birds in the database are more than 5 years old, with a maximum of 17 years verified age) is much more consistent with an average survival rate of 80% per year (as my reanalysis suggests) than with an average survival rate of 93% (as in the revised ARM).

Further, it is worth noting that almost all vertebrate species with delayed maturation life cycles, like red knots, that do not recruit to the breeding population until their third year, exhibit high adult survival rates. This is especially true when annual reproductive output is low, as it is with red knots, which lay only four eggs in a single nest per year.

> This comment is hard to interpret, given that “high” and “low” are not defined. The red knot recruitment estimates used in the revised ARM are indeed very low (around 0.1 adult females recruited per female per year) and would require a very high survival (greater than approximately 90%) to result in a stable or growing population. However, red knot recruitment rates (in the revised ARM framework, a compound of reproductive output and survival to the first breeding migration) are poorly understood, and further research is needed to better understand this critical demographic process.

Outside of the Delaware Bay system, high survival and long lifespans are also reported for red knots and other shorebirds of similar size and annual cycle. For example, Piersma et al. (2016) report that annual apparent survival for red knots in Western Australia were well above 90% in most years of their study. In another example, Boyd and Piersma (2001) reported that they recaptured 155 birds in their sample >14 years after initial capture and 2 over 24 years after initial capture. There are published studies that report survival rates at 80% or lower, but to assert that the estimated survival rates used in the ARM based on the mark-recapture data are outliers or excessively high is erroneous.

> While the previous literature is not definitive on this topic, I was primarily basing my conclusions on a reanalysis of the raw banding and resighting data from Delaware Bay rather than on prior studies. Regardless, the question of mean survival rates (and the role of staging areas in regulating survival and trends in survival) for red knots and other similar shorebirds is important for conservation and management and I hope this discussion continues in the form of peer-reviewed publications and other constructive scientific dialog.

In his report, Dr. Shoemaker claims that the survival estimates in the ARM are biased by individual misidentification, or flag misreads. Before analyzing the data, the ARM Subcommittee conducted a thorough QA/QC, including filtering records to only lime and dark green flags that were first deployed by New Jersey or Delaware, removing records of 5 duplicate flags (n = 36), flags apparently resighted before they were deployed (n = 711), and flags that were never deployed (n = 1). Removal of these records represents only 0.35% of the total resightings.

Members of the ARM Subcommittee have worked extensively on the issue of flag misreads, including conducting a thorough simulation study investigating the situations in which misreads might bias survival estimates and the implications of that bias (Tucker et al. 2019). The key points from that work are: 1) misreads disproportionately affect survival estimates from the first years of the study, causing apparent negative trends in survival over time, and 2) there is an important tradeoff to consider between potential bias due to misreads and loss of precision if data filtering is applied. In that paper, the authors suggest a data filtering step of removing all observations of flags that were only seen once in a year as a way to potentially mitigate misidentification errors. However, there are nuances to consider when determining whether this is necessary, because this data filtering will inevitably remove some number of valid observations, and the authors identify thresholds that depend on study length and error rate. For a 10-year study, removing single observations becomes beneficial if the error rate is >5%; below that rate the bias is minimal relative to the detrimental effects of removing valid observations. In the Delaware Bay mark-recapture dataset, the misread error rate is between 0.38% (712 impossible observations/187,587 total) and 4.5% (8,448 single observations).

Additionally, the characteristic apparent negative trend in survival over time that would indicate bias due to misreads is not observed. To examine this further, the distribution of the number of resightings in a year for every flag (Figure 1) was plotted, with and without removing single observations. The shape of the resulting histogram indicates that removing these records results in fewer flags being seen once in a year than would be expected, i.e., that the data filtering removes a large number of valid records (> 3,000).

> First, I acknowledge the important work done by members of the ARM subcommittee related to the issue of flag misread errors in shorebird resighting surveys. Notably, I relied heavily on Tucker et al. (2019) in my reanalysis of the resighting data, and used the method they suggested (removing 'singlet' observations from the analysis) to correct the potential bias in survival rates due to misread errors. Given the prior work on this issue by members of the ARM subcommittee, I was surprised that they did not attempt to correct for this possible source of bias when estimating red knot vital rates for the revised ARM. While they claim that the characteristic negative trend in survival across time (an artifact of this type of flag misread errors) is not observed, I am not convinced on this point. Upon visual inspection, there does appear to be a negative trend in survival across time in Fig. 3a from Tucker et al. (2023; also Fig. 44 from the revised ARM; ASMFC 2021), and this pattern also appeared in my analyses of the same data using data that included the "singlet" observations (which were potentially contaminated by misreads).
>
> While I understand that ASMFC performed quality checks and removed obvious misread errors (e.g., flag codes observed before they were deployed) there is simply no way to detect an errant flag code if that code had been previously deployed in Delaware Bay

(although one recent paper suggested a model-based approach for estimating the misread process; Rakhimberdiev et al. 2023). The longer the period of flag deployment and the more birds that are tagged, the more likely an errantly recorded flag code may match with a previously deployed code. Importantly, if the errant match is to a bird that died many years prior, the capture-recapture analysis will adjust the estimated survival rate upward to reflect the "survival" of the long-dead bird (therefore, the longer the time series, the stronger the potential bias due to this class of flag misread errors). Finally, I note here (as I did in my 2023 peer review report) that all or nearly all of the available flag codes have been deployed at Delaware Bay (at least for the lime green flags). If so, any misread errors are likely to match with previously deployed flag codes and thereby inflate survival estimates.

To be safe, it makes sense to remove "singlet" observations, retaining only those flag codes that were confirmed via multiple observations to be present in Delaware Bay each year. This ensures that survival estimates are not biased from the potential misreads. Clearly, many of those "singlet" observations are true observations, and discarding these records necessarily involves omitting a substantial amount of valid data from downstream analyses. As an ecologist I understand the drive to use all available data. But in this case, even a small number of misread errors can induce an unacceptable bias in survival estimates. Furthermore, the dataset is so information-rich that we can afford to filter out a relatively small fraction of the data ("singlets" comprised approximately 9% of total resighting observations and around 35% of unique individual-year occurrences) to address an important source of potential bias in survival estimates.

Finally, I reiterate that the primary evidence that ASMFC's estimates of red knot survival are biased high is that there was a marked decline in the mean survival estimate after the singlet observations were removed (per Tucker et al. 2019). I do not know of a reasonable alternative interpretation of this result except as strong evidence for the influence of misread errors in the resightings database. Moreover, analyzing the capture/banding data (where misread errors are likely negligible) as the sole source of information also yielded a mean survival rate of around 80% after accounting for potential transients. Overall, I remain convinced that the red knot survival estimates used by ASMFC were biased high due to the presence of misread errors. I recommend that the ARM subcommittee correct for this source of bias, either by eliminating "singlet" observations or by explicitly modeling the flag misread process (e.g., Rakhimberdiev et al. 2023).

The integrated population model uses the mark-recapture data to estimate survival as well as parameters related to stopover site use within each year. There were concerns that removing single observations would bias estimation of within-year parameters, and because the error was below the thresholds identified by Tucker et al. (2019) and the characteristic negative trend in survival was not observed, single observations were kept in the data set for the analysis.

In this statement, the ARM subcommittee indicates that their decision not to account for potential misread errors was due largely to the perception that the "singlet" observations were necessary for fitting additional parameters in their open robust design (ORD) model

> (the component of the red knot integrated population model that is primarily responsible for survival estimation). The ORD model uses the mark-resight data to fit multiple parameters related to within-year stopover use and availability for capture (e.g., timing of entry and exit to the staging area), in addition to among-year processes -- most notably, survival. The ORD model is impressively complex, and appears to perform well at parameter estimation when the data do not violate key assumptions (see Tucker et al. 2023 and my 2023 peer review report). However, like all statistical models, biases can arise due to violation of model assumptions. One of the key assumptions of the ORD model (like most capture-recapture analyses) is that the unique identification marks assigned to each individual (in this case, flag IDs) are neither lost nor mis-identified. Violation of this assumption can result in biased parameter estimates (especially survival).
>
> It appears the red knot modelers were concerned that removing “singlet” observations could bias the estimates for some of the within-year parameters estimated in the ORD model, such as the dates of entry and exit each year. I can understand why the authors of the revised ARM wanted to fit a complex model that incorporated within-year processes. But there are always trade-offs when building ecological models. In this case, there is an apparent tradeoff between potential biases in survival estimates and potential biases in estimating within-year parameters like entry/exit dates. The modelers could have chosen to use a simpler capture-recapture model that did not explicitly incorporate detailed within-year processes (such as the Cormack Jolly Seber models I used in my 2023 peer review report) -- in this case, there would have been little downside to removing the “singlet” observations. However, the ARM subcommittee ultimately chose to use the more complex ORD framework.
>
> Although I am sympathetic to the modelers in this case, I ultimately disagree that the benefits of adopting the more complex model should outweigh the potential biases in survival estimation due to misread errors. From a conservation and management perspective, survival represents one of the key processes of population ecology (survival and reproduction rates are typically referred to as “vital” rates in wildlife demography). Biased survival estimates can easily tip the balance between a growing and declining population. Biases in the estimated date of entry into the staging area (for example) would tend to be much less consequential for the revised ARM than biases in adult survival rates. Nonetheless, simulation trials would be necessary to quantify the degree to which removing “singlet” observations could bias the within-year parameter estimates and whether biases in within-year parameters could have an effect on survival estimates in the ORD model.

The ARM Revision (ASMFC 2022) contains a thorough discussion of this topic on pages 63-64, in which several hypotheses for the disagreement in annual survival probability estimates from the older studies was described. Dr. Shoemaker points to lower estimates of survival from studies from the early 2000s, when red knot annual survival probability was estimated to be close to 0.8. It is likely that older estimates were negatively biased to some extent due to short study periods, low detection probably, and unmodeled temporary emigration from the system. It is also possible that during that time, when horseshoe crab populations were lower, red knot survival probability was truly lower. Alternatively, because permanent emigration from the system cannot be distinguished from mortality in older mark-recapture studies, a higher rate of permanent emigration (i.e., birds abandoning Delaware Bay for other spring stopover sites) would appear as lower survival probability. It is possible that there is a threshold of horseshoe crab abundance below which red knot survival probability might be expected to drop dramatically. If such a threshold exists, it was not observed over the time series included in the model (2005-2018). It has also been proposed that southern-wintering birds (with longer migrations) have lower annual survival probabilities than northern-wintering birds. Declines in the number of red knots overwintering in Argentina (Niles et al. 2009) suggest a decline in the southern-wintering subpopulation and therefore it is possible that in more recent years a greater proportion of the Delaware Bay stopover population are northern-wintering birds. As discussed in the report, this is a key area for future research.

> I appreciate this discussion and I understand there are many nuances that must be considered when comparing survival estimates across multiple populations or time periods. However, none of this information contradicts my reanalyses.
>
> In the above statement ("It is possible that there is a threshold of horseshoe crab abundance below which red knot survival probability might be expected to drop dramatically"), ASMFC acknowledges that the relationship between horseshoe crabs and red knots may in fact be stronger than the weak relationship they detected using the 2005-2018 time series. If a stronger relationship is plausible and consistent with the observed red knot decline (which has been attributed to unregulated commercial harvest of horseshoe crabs), it seems prudent to include this hypothesis within an ARM framework for this system. ASMFC maintains that the revised ARM represents a major advance because it uses data from the Delaware Bay system. However, in this case I think the ARM subcommittee may have prioritized mathematical elegance (ability to fit a single integrated model using only data collected from the target population) over comprehensiveness (e.g., including knowledge about the system prior to the deployment of leg flags). The more comprehensive approach (incorporating data from additional populations and time periods, including multiple alternative models) may be messier, but will better reflect relevant knowledge and more effectively guide critical decisions about this system. Furthermore, by fitting and comparing multiple models and data sources we can learn more rapidly about this two-species system and better understand where potential biases lie.

Criticism 2: Trawl-based indices of horseshoe crab abundance are inadequate for modeling the biotic interaction between red knots and horseshoe crabs.

- The inclusion of trawl surveys as indices of horseshoe crab abundance may be imperfect but it is the best available science and its use has been approved by several independent peer reviews.
- Most of the criticisms and caveats relevant to trawl surveys would also apply to egg

density and red knot abundance estimates.

- There is consensus among the trawl surveys for an increasing trend in horseshoe crab abundance since 2010.
- Trawl surveys are the standard for bottom dwelling organisms and for evaluating the abundance of many species.

Technical Response: Dr. Shoemaker argues that the trawl surveys used to monitor horseshoe crab abundance and serve as the basis of the catch multiple survey analysis (CMSA) are "…imperfect snapshots of the abundance of horseshoe crabs occupying Delaware Bay, obscured by differing survey methodologies and poorly understood aspects of horseshoe crab ecology, including seasonal and daily activities, habitat preferences, and degree of clustering on the seafloor." The ARM Subcommittee agrees that the trawl surveys are imperfect; catchability differs in each survey and possibly differs both within and between years. Such is the nature of fishery-independent surveys, and these same arguments also apply to indices of abundance for red knots and horseshoe crab egg density estimates. However, the use of the trawl surveys to index horseshoe crab abundance has gone through multiple peer reviews (e.g., ASMFC 2009b, ASMFC 2019, ASMFC 2022, Anstead et al. 2023) and found to be a scientifically sound measure of horseshoe crab abundance.

> I agree that there is substantial uncertainty in all of the data sets related to horseshoe crab abundance in Delaware Bay, including the trawl surveys, spawning surveys and surface egg density estimates. Since ASMFC primarily used trawl-based indices of abundance (in addition to harvest, bycatch estimates, etc.), I focused my peer review reports on the uncertainty inherent to the trawl-based surveys. The presence of substantial uncertainty in this system underscores the critical importance of treating uncertainty appropriately--from acknowledging measurement uncertainty (uncertainty in the raw measurements), parameter uncertainty (uncertainty about the true value of a particular parameter) to formally incorporating structural uncertainty (multiple alternative hypotheses for how the system works). Furthermore, given that a Federally Threatened species is involved, I argue that plausible "worst-case" scenarios or hypotheses should be assigned substantial weight until they can be effectively ruled out. In this case, given the extreme uncertainty about horseshoe crab demography, behavior and abundance, I think it is prudent to acknowledge a non-negligible possibility that this population is not currently experiencing a strong recovery. While the ARM subcommittee claims that the revised ARM accounts for uncertainty, their accounting is incomplete. Most importantly, the revised ARM fails to acknowledge structural uncertainty; in effect, they are assigning a 100% credibility score to their chosen model structures (e.g., the CMSA model) and data sources (e.g., assigning substantial weight to the trawl-based surveys while ignoring the horseshoe crab egg density data). The horseshoe crab population may indeed be recovering (and as discussed below, there is some evidence for a recent population increase) but multi-model inference (using model weights to express the uncertainty among alternative models) is needed if we want to more realistically express our overall belief in this hypothesis.

Dr. Shoemaker faults the trawl-based indices of abundance used by the ARM Subcommittee for not considering environmental covariates that could influence the catch of horseshoe crabs, and he obtained the raw data to recalculate the indices using generalized linear models (GLM) and generalized additive models (GAM). The ARM Subcommittee does not disagree with this approach to standardizing abundance indices based on environmental covariates, and this sort of analysis was conducted as part of the 2019 stock assessment (ASMFC 2019) but it did not improve the indices of abundance (e.g., decrease errors, reduce large annual fluctuations). The peer review panel for the ARM Revision (2022 ASMFC) recommended using a model-based index for the Delaware Trawl Survey because it is a fixed station survey; consequently, the ARM Subcommittee applied this approach prior to using this survey in the CMSA. The Virginia Tech Trawl Survey has a well-designed sampling scheme that stratifies sampling based on habitat; thus, habitat features that could influence catchability are already incorporated into the abundance estimates from this survey. Finally, and as stated earlier, a GLM did not improve the precision of the New Jersey Ocean Trawl Survey (ASMFC 2019) and the ARM Subcommittee continued using a simpler calculation of the abundance estimate (the delta-mean catch-per-unit-effort).

> I agree that both model-based and design-based approaches can be useful in this context. In this case, ASMFC chose to use a fully design-based approach for generating abundance indices from the three trawl surveys. While the approach used by the ARM subcommittee was a fairly standard approach for analyzing trawl survey data, I was surprised that they did not use model-based standardization to further control for environmental and seasonal factors known to influence horseshoe crab capture rates. The rationale for performing model-based standardization is particularly strong considering that (1) horseshoe crab captures are known to be strongly influenced by factors like temperature, depth and season, and (2) these key environmental drivers are measured as part of all three trawl surveys used in the revised ARM. The decision to ignore the available covariate data places a degree of trust in the design-based controls that does not seem warranted in this case. Importantly, ignoring the covariate data implicitly assumes that these data have zero effect on the trawl survey results -- a strong assumption that is likely to be false in this case. I maintain that ASMFC should use all available covariates to help standardize observations across surveys and across years, although I would welcome continued dialogue about the nuances of this analysis.

Like trawl surveys for any aquatic species, there is considerable variation in the catches of horseshoe crabs among individual trawl samples resulting in high inter-annual variation in abundance indices. Dr. Shoemaker concludes there is a lack of statistically significant correlation coefficients among the trawl surveys, and there is a fatal flaw in using those data to infer abundance. The ARM Subcommittee disagrees with this analysis and can demonstrate that there is in fact a significant correlation between trawl surveys and with the CMSA estimates of abundance (see response to Criticism 3). There is observation error associated with each survey (e.g., being in the right place at the right time) and it is not uncommon for a relatively high catch in one survey to correspond with a relatively low catch in another for the same survey year, so it is not surprising that there could be some “non-significant” correlations or correlation coefficients that one may consider low. However, each trawl survey could very well show a statistically significant trend. It is the consensus among surveys about the trend that is important, not how closely individual observations from the respective surveys track one another. The ARM Subcommittee acknowledges that each survey does not perfectly track the population, which is why the CMSA uses multiple surveys. In addition, it is very possible, from a statistical sense, that two time series of abundance data could not show a statistically significant correlation, but could still both show a statistically significant trend (Figure 2).

> Here I agree that more data is better than less data, and more independent datasets are better than fewer. Correlation tests and scatterplots remain a valuable exploratory analysis for detecting the degree to which different datasets share information. However, as ASMFC points out above, uncorrelated datasets can yield emergent patterns when their information is combined. In fact, after reviewing the ASMFC response to my peer review reports, and after running some confirmatory analysis, I see evidence for a recent increase in the Delaware Bay horseshoe crab population. Taken together, I agree that the three trawl-based surveys provide some evidence for a recent increase in the horseshoe crab population since around 2010.
>
> However, the evidence for a recent increase in the Delaware Bay horseshoe crab population based on the trawl-based surveys is predicated on several important assumptions, including: (1) all three trawl-based surveys are equally valid (and therefore should be assigned equal weight in the analysis), (2) each survey is equally informative with respect to the key state variable of interest (e.g., the abundance of female horseshoe crabs), and (3) that each survey is an independent sample from the population of interest. Potential violations of each of these assumptions should be carefully considered; it would be prudent to perform additional sensitivity tests to evaluate the effects of plausible violations -- and possibly to formally incorporate alternative models in which one or more of these assumptions is relaxed.

Dr. Shoemaker also conducted his own capture-recapture analysis to determine the relationship between trawl-based indices of horseshoe crab abundance, horseshoe crab egg density, and red knot survival. Contrary to the results of the ARM Subcommittee, Dr. Shoemaker did not find any positive relationships between horseshoe crab abundance and red knot survival. Although additional analysis of these data is welcome, the ARM Subcommittee questions the value of such a comparison due to the many differences in how the data were analyzed. Dr. Shoemaker's analysis only used information about whether a bird was seen at least once in a year in a standalone Cormack-Jolly-Seber model, whereas the ARM Revision uses both within-year and among-year observations in an open robust design model that is embedded within an integrated population model. These differences in modeling approaches make it difficult to draw meaningful conclusions regarding differences in results. The analysis done by the ARM Subcommittee did find a positive relationship between horseshoe crab abundance and red knot survival, providing the demographic link between population models used in the ARM Framework.

> Here I do not find ASMFC's response convincing. In my reanalysis of the banding and

resighting data, I used a Cormack-Jolly-Seber (CJS) framework to estimate annual survival rates. The CJS method has for many decades been the gold standard for estimating survival on the basis of capture-recapture data. In fact, the open robust design (ORD) model used in ASMFC's integrated population model for red knots uses a modified CJS framework to estimate survival and other inter-annual population processes (Tucker et al. 2023). Regardless, estimates of apparent survival from different analytical methods are comparable, as they represent the same fundamental ecological process. Of course, this statement requires that both approaches are statistically valid-- but ASMFC does not appear to be questioning the validity of my methods.

Given that it is meaningful to compare my results with ASMFC's capture-recapture results, the fact that the CJS approach failed to detect a statistical signal linking red knot survival to trawl-based horseshoe crab population estimates is notable. This was true whether I used the CMSA estimates of horseshoe crab abundance (following ASMFC's approach) or any of the trawl-based surveys (NJ, DE, VT) separately (whether or not these indices were adjusted to control for seasonality and environmental conditions). Although I do not have a ready explanation for why my results differed from ASMFC's integrated population model, I think it would be prudent and instructive to run additional tests to try to understand the underlying reasons for these differing results - especially given the fundamental importance of this relationship to this two-species ARM framework.

Finally, I reiterate that, although ASMFC detected a positive relationship between red knot demographic rates (specifically, adult survival) and horseshoe crab abundance, this relationship was not ecologically meaningful (see my response under section titled "Criticism 8", below). Therefore, in one sense the results of our two independent analyses yield the same conclusion: that red knot demographic rates are not directly or meaningfully correlated with trawl-based indices of horseshoe crab abundance during the time period for which data are available. In contrast, using the same CJS modeling framework, I found that red knot survival was meaningfully and positively related to an alternative horseshoe crab population index -- surface egg densities.

Criticism 3: Red knot survival is strongly sensitive to horseshoe crab egg density, indicating that persistent degradation of the horseshoe crab egg resource could have dire consequences for the red knot population.

- During the development of the ARM Revision, horseshoe crab egg density data were requested, but were not provided to the modeling team. Therefore, these data could not be considered as an input to the models.
- Trends in horseshoe crab egg density (extracted from Smith et al. 2022 following the publication of the ARM Revision) are correlated with other data inputs for the years included in the ARM models and thus the inclusion of egg density data in the models is unlikely to result in any meaningful difference from the current ARM Framework in terms of harvest recommendations.

- Smith et al. (2022) showed a general increasing trend in horseshoe crab egg density in recent years similar to that of horseshoe crab abundance, consistent with findings from the ARM Revision.

Technical Response: The debate over the inclusion or exclusion of egg density data has been ongoing since the ARM Framework was initiated in 2007. The ARM Subcommittee does not deny that eggs are the true link between horseshoe crabs and red knots. However, the reasons for excluding egg density data from the ARM model, which range from sampling design to data availability, have been extensively discussed since the inception of the original ARM Framework, in both published versions of the ARM Framework (ASMFC 2009a, 2022) and in response to a minority report on the ARM Revision (ASMFC 2022). Ultimately, egg density data could not be considered in the ARM Revision because they were not provided to the ARM Subcommittee when requested. When egg density data were published (Smith et al. 2022), the trends appeared to be increasing during the years modeled, consistent with trends of the trawl- based indices used in the model.

> I am not able to comment on data availability issues. Nevertheless, reading the minority reports on the revised ARM prompted the idea of running capture-recapture analyses using surface egg density data as an alternative metric to represent year-to-year variation in the horseshoe crab resource at the Delaware Bay staging area. As discussed above, this analysis demonstrated that red knot survival was meaningfully and positively related to surface egg densities.

Egg density data are highly variable, both spatially and temporally within a spawning season, and discrepancies in egg density results have been noted depending on who processed samples and how they were processed.

> I agree that the surface egg density data is variable from sample to sample, but the sample size is large each year (hundreds to thousands of samples), and covers a large area within 16 beach segments that span most of the New Jersey side of Delaware Bay. Therefore, the average egg density observed each year still seems likely to contain useful information about annual mean densities via the law of large numbers. Furthermore, I did my best to use model-based controls to account for differences in effort and differing sampling methods.

To incorporate egg density data into the ARM would require development of two linked models, in which the relationship between horseshoe crab abundance and observed egg density is quantified in one, and the relationship between egg density and red knot survival/recruitment is quantified in the other. Such analysis and data exploration were not conducted during the ARM Revision primarily because the egg density data were not provided. The ARM Subcommittee is not opposed to using the egg density data as another index of horseshoe crab abundance once a reliably quantifiable relationship can be established. However, the first time the ARM Subcommittee saw the recent egg density results was in 2021 in the form of a draft manuscript (later published as Smith et al. 2022) as part of a minority report by Dr. Larry Niles. If the owners of the egg density data had been willing to provide the raw data, those data would have been considered in the revision of the ARM Framework. Instead, the ARM Subcommittee accounted for egg availability to shorebirds by including the timing of horseshoe crab spawning in the red knot integrated population model and made a research recommendation to examine the relationship between egg density estimates and horseshoe crab abundance estimates.

> I am pleased to hear that the ARM subcommittee is amenable to using the surface egg density data in the ARM. However, it does not seem appropriate to treat the egg density data as just "another index of horseshoe crab abundance" for use in the CMSA model. The CMSA model (which I have some additional concerns about; see below) is meant to provide an index of horseshoe crab abundance in and around Delaware Bay. The egg density data, on the other hand, is a measure of the usable food resource available to red knots. My re-analysis of the Delaware Bay red knot data strongly suggests that the egg density data provide a crucial empirical link between the red knot population and the horseshoe crab population. I suggest that a more useful and ecologically informed approach would be to use the surface egg density data to specify a mechanistic link between female horseshoe crab abundance (as described in the next paragraph) and the red knot population (possibly even mediated by a spawning process model). Simply incorporating the egg density data into the CMSA model would inappropriately combine fundamentally different data and ecological processes (and would raise difficult questions about how to weight these data relative to the trawl surveys), and would dilute key information about the functional link between these two species.
>
> I think ASMFC should consider incorporating the egg density data even if a precise functional relationship between horseshoe crab abundance estimates and surface egg densities cannot be immediately established. Logic dictates that a relationship must exist, although there are several intermediary mechanisms linking these system states (female abundance linking to egg masses deposited prior to red knot arrival, linking to surface egg availability via beach disturbance processes; perhaps as part of a structural equation model; Grace et al. 2010) that will add ecological realism to the overall process model. While there is uncertainty about the exact functional form of the relationship between surface egg density and horseshoe crab abundance (as is the case for many ecological relationships), it is a known causal linkage and so even a linear model could provide a simple and logically defensible quantitative description of this relationship. Furthermore, the adaptive resource management paradigm enables researchers to incorporate uncertainty into policy decisions. In this spirit, the ARM could incorporate several alternative plausible functional forms to describe the relationship between horseshoe crab abundance and surface egg densities just as the original version of this ARM framework incorporated several alternative functions relating red knot mass (itself a function of horseshoe crab population) to red knot survival (McGowan et al. 2015).

In Dr. Shoemaker's report, he finds that surface egg densities are uncorrelated or negatively correlated with the CMSA results and other indices of abundance used in the ARM Framework. In this analysis, he uses data from 1990-2022 although the CMSA and ARM Framework use data beginning in 2003. The CMSA model starts in the early 2000s to coincide with the start of many of required data sets used in the analysis (e.g., Virginia Tech Trawl, biomedical harvest, estimated dead discards from other fisheries). If the correlation analysis is abbreviated to include only the years used in CMSA modeling, all time series are positively correlated (Figure 3) for female horseshoe crabs (Dr. Shoemaker's analysis does not specify if his correlation analysis is for males, females, or both). In fact, the egg density time series from Smith et al. (2022) is positively and significantly correlated with the CMSA estimates of female horseshoe crabs. Therefore, it is likely that if the egg density time series were included in the ARM Framework as another index of horseshoe crab abundance, the CMSA results would not be much different from the current results.

> I reached out to the ARM subcommittee on Oct 21 2023, soon after ASMFC announced that they would issue a formal response to my peer review report, to inform them that I would be happy to address any questions that came up during their review of my work. If they had a question about how I analyzed or subsetted the trawl data (all of my analyses of the trawl data were for females only), then they could have asked me directly. They did not do so.
>
> In response to the above comments, I re-ran the correlation tests with a subset of the data that only included years from 2003 onward. The results were no different from my original analysis- there were weak (statistically inconclusive) negative correlations between the trawl-based abundance indices and the surface egg density index. However, the sign of the correlation flipped when I used the raw (without model-based standardization) trawl-based indices and the unadjusted egg density index. Nonetheless, correlation coefficients for the raw indices remained very weak (0.2 to 0.3) and were statistically inconclusive at alpha = 0.05.
>
> However, this discussion is of limited importance in comparison with the key point -- surface egg densities (whether raw or adjusted) strongly influenced red knot survival in my reanalysis of the capture-recapture data. In contrast, abundance indices from the trawl-based surveys showed no conclusive relationship with red knot survival. These facts provide strong support for incorporating the surface egg density data in the revised ARM (and not simply as another index of horseshoe crab abundance for use in the CMSA model- see above). I conclude that the trawl-based abundance estimates are not an adequate substitute for the information contained in the surface egg density data.

Additionally, Dr. Shoemaker analyzed the egg density data from Smith et al. (2022) and accounted for differences in survey methodology through time. The results of his reanalysis showed no trend in egg density although Smith et al. (2022) showed a general increasing trend in recent years similar to that of horseshoe crab abundance from the CMSA (Figure 4).

> In my re-analysis of the long-term egg density data, I added an offset term to account for differences in survey methodologies through time and thereby enable more robust comparisons among these different time periods. I have discussed this issue with the lead author of Smith et al. (2022), who agrees that the methods I used to re-analyze the trend in the long-term surface egg density data improved upon the methods used for trend estimation in Smith et al. (2022); which did not account for differences in survey effort in

> different segments of the time series (J.A.M. Smith, *pers. comm.*).

Dr. Shoemaker also conducted an analysis that shows the effect of egg density on red knot survival. However, this survival analysis is not documented in great detail and only includes data from the New Jersey side of the Delaware Bay. Thus, it is questionable whether this analysis is representative of the red knot population as a whole.

> It is unfortunate that similar egg density data were not available for the Delaware side, but that fact does not invalidate my analysis; in ecological modeling we do the best we can with the available data in spite of known limitations. Furthermore, I fail to see why this relationship would not hold on one side of the bay if it holds for the other. Nonetheless, my results strongly suggest that it will be important to continue collecting surface egg density data. Fortunately, it appears that standardized horseshoe crab egg density surveys will be available on both sides of the bay going forward.
>
> While my peer review report admittedly did not contain the level of analytical detail that would be expected of a scientific paper, I offered to share the code for running these analyses with ASMFC and to address any questions or concerns about my reanalyses. This offer still stands.

If these analyses by Dr. Shoemaker are correct, it still begs the question of how to incorporate this into the ARM Framework. In Dr. Shoemaker's report, red knot survival is positively correlated with egg density but egg density has not changed over time; however, female horseshoe crab abundance has increased. Therefore, while egg density and female horseshoe crab abundance must ultimately be linked, this relationship is not evident in the data. The lack of an empirical relationship ultimately complicates any effort to quantify a model linking horseshoe crab abundance to red knot survival through egg density. Dr. Shoemaker falls short of proposing a way to do this.

> If my analysis is correct, there is reason to believe the relationship between red knots and horseshoe crabs is much stronger than the current ARM framework suggests, and that surface egg densities provide a critical link for understanding and describing this relationship. Regardless of the nuances and complications that might be involved in incorporating these data in the revised ARM, the rationale for incorporating surface egg density data into this ARM framework is very clear.
>
> In my peer review of the revised ARM, I was only tasked with evaluating its scientific merits; offering suggestions for improvement was not a primary objective of my previous reports. However, I would be happy to work with ASMFC to discuss incorporating horseshoe crab surface egg density data in the next iteration of this ARM framework.

Regardless, for the time series of the CMSA model, egg density is positively correlated with the other time series of horseshoe crab abundance used. Because egg density data are not readily available to the ARM Subcommittee (either for the model development in 2021 or possibly on an annual basis that would be required for their inclusion), the data only cover New Jersey beaches, and their use and sampling design have been questioned over the years, the trawl surveys remain the best available data for horseshoe crab abundance in the ARM Framework.

> The surface egg density data are now available. Further, it seems likely that the results of future surface egg density data would be furnished to ASMFC on a regular basis. If these data are important for linking red knot demographic rates to horseshoe crab abundance, and if they are indeed available, then ignoring these data seems to contradict the spirit of the term "best available data".

Criticism 4: The ARM exaggerates the evidence for an increasing trend in the number of female horseshoe crabs in the Delaware Bay.

- The analysis provided in Dr. Shoemaker's report contains errors, including the use of incorrect data subsetting for the indices and application of an analysis that was inappropriate for the data.
- The trawl-based indices were thoroughly considered by the ARM modelers and represent the best available data for tracking horseshoe crab abundance.
- The goal of the ARM modelers was not to find an increasing trend, but to develop the data in the most statistically sound way possible regardless of the answer.

Technical Response: Dr. Shoemaker suggests the ARM Subcommittee exaggerates the evidence for an increasing trend in horseshoe crab abundance through time. A long time to maturity for horseshoe crabs (9-10 years) suggests that recovery from overfishing would take some time to become evident in fishery-independent surveys. With reductions in harvest in the Delaware Bay region in the early 2000s, it makes sense that any increase in abundance would not be seen until approximately 10 years later (~2010). This is what was observed in the three trawl surveys used to index abundance. When a simple linear regression model is fit to each one of the trawl surveys beginning in 2010, all of them show statistically significant increasing trends (Figure 5). Dr. Shoemaker argues that "…trawl-based indices of horseshoe crab abundance are a noisy and unreliable indicator of annual fluctuations in the horseshoe crab population, and are likely an inadequate metric for quantifying the biotic interactions between red knots and horseshoe crabs in the Delaware Bay." The ARM Subcommittee emphatically disagrees with this statement given the life history of horseshoe crabs, the amount of time since bait harvest has been curtailed, and the agreement of the three trawl surveys for an increasing trend in abundance. Harvest management appears to have worked to increase abundance. A rebuttal to this point is also given in Criticism 2.

> First, I agree that we would expect to observe a delay between the initiation of regulation and the initiation of an observable population recovery due to the delayed maturation of female horseshoe crabs. I also agree that a segmented regression (or even a spline or polynomial fit), rather than an ordinary linear regression, is an appropriate approach for analyzing trends in the long-term trawl data (see below). Therefore, I agree that linear regression was too simplistic to be used for this purpose (Fig. 12 from my 2023 peer review report).
>
> I ran additional tests to confirm the ARM subcommittee's statement that "when a simple linear regression model is fit to each one of the trawl surveys beginning in 2010, all of

them show statistically significant increasing trends". Using my adjusted catch per unit effort (CPUE) indices that controlled for several potentially confounding factors, my analyses confirmed the apparent increases in horseshoe crab CPUE since 2010 (note that, as of this writing, I do not have access to the trawl survey results after 2022; Fig. 3). It is interesting to note that none of these relationships were statistically significant at alpha = 0.05 when trawl data from 2000 and onward were included in these regression analyses. However, since the 2010 threshold was not arbitrary, but was based on the expected delay in an observed population rebound (see above), there is nevertheless evidence for a recent increase in the Delaware Bay horseshoe crab population.

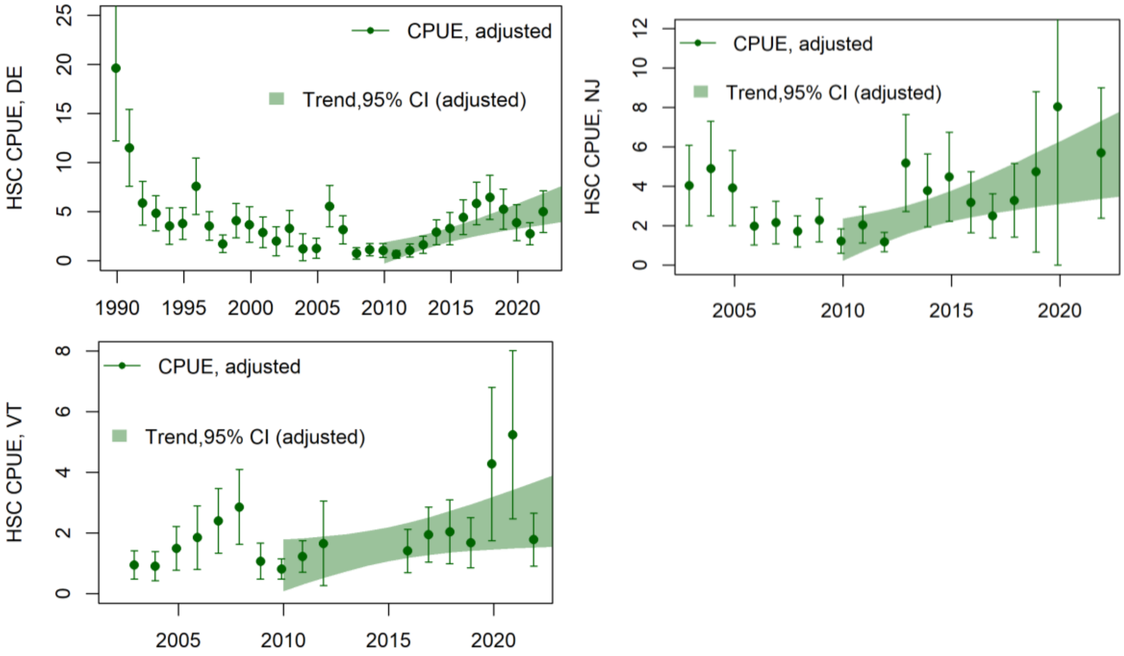


Figure 3. *Analyses indicating that there is an apparent positive linear relationship between three trawl-based horseshoe crab abundance indices (from NJ, DE, and Virgina Tech data) since the year 2010. Each figure displays catch per unit effort (CPUE) estimates adjusted for the effects of seasonality, water temperature, depth, and dissolved oxygen. Error bars represent 95% credible intervals for a linear regression of CPUE over time since. The green polygons represent 95% confidence intervals for the linear regression of the adjusted CPUE against time in years since 2010.*

Dr. Shoemaker again faults the indices of abundance used by the ARM Subcommittee for not being standardized according to environmental covariates in a GLM approach, and he specifically demonstrates his standardization on the New Jersey Ocean Trawl data. However, during an initial review of his report by New Jersey and Delaware staff, it was recognized that he subset the data incorrectly, using the wrong time periods including sample periods when the crabs are not fully available to the survey, resulting in data and an index of abundance that are not used the by ARM Subcommittee. Dr. Shoemaker included the January samples, when the overwintering crabs may remain farther offshore than the survey's sample area, accounting for the significantly decreased catches during this period. He also included the June samples, when most of the adult crabs have migrated into bays and estuaries to spawn, again making them unavailable to the survey. The inclusion of these two sampling periods has an inappropriately dampening effect on the resulting indices which cannot be corrected through a GLM standardization and will not provide an accurate index of relative abundance. Again, a GLM standardization was attempted with the New Jersey Ocean Trawl data during the 2019 benchmark stock assessment (ASMFC 2019), but it was found to not provide any improvement over a simple delta-mean index. Standardization of the trawl survey catches by a GLM or GAM is still something worth exploring in future assessments as additional years of data may provide the necessary information to better evaluate the true effects of covariates on catches.

> I stand by my reanalysis of the New Jersey Ocean Trawl survey data. For these trawl data, as with the red knot data, I made an effort to analyze the data independently, using my training and experience rather than relying on ASMFC's analytical methods. In their response, ASMFC claims that my analysis of the New Jersey trawl data (which included survey data collected from all months of the year) was incorrect, stating that "the inclusion of these two sampling periods has an inappropriately dampening effect on the resulting indices which cannot be corrected through a GLM standardization and will not provide an accurate index of relative abundance". However, ASMFC did not provide further evidence or rationale for this statement, and I maintain that my methods were appropriate.
>
> In my re-analysis of the NJ trawl data (and the other two trawl surveys; more detail can be found in my 2023 peer review report), I relied on a model-based approach to control for potentially confounding factors such as water temperature, trawl depth and seasonal effects (ordinal date). Specifically, I modeled horseshoe crab captures as a complex, non-linear function of survey effort, environmental factors, and season. By using spline fits within a Generalized Additive Model (GAM) framework I was able to account for complex relationships between catch-per-tow and factors such as ordinal date (controlling for seasonality and allowing for strong fluctuations across different times of year; see Fig. 10 from my 2023 peer review report). Therefore, I was able to use the full NJ trawl dataset while accounting for times of year during which crabs were not fully available for capture. These models passed tests of model adequacy (using quantile residuals, implemented in the 'DHARMa' package in R) and appeared to perform admirably in accounting for these complex, potentially confounding factors.
>
> In contrast, the ASMFC experts relied on sampling design and data sub-setting to control for any potentially confounding factors. I argue that there are very good reasons to use model-based controls to enable standardized comparisons across surveys and years. Sampling design and data sub-setting cannot control for all the factors known to affect horseshoe crab detection rates. Furthermore, data sub-setting effectively discards data that could potentially help to shed light on key questions of interest; in contrast, model-based controls enable us to use all available data. Horseshoe crab capture rates are known to be strongly influenced by multiple factors, including temperature and seasonality. Since information on environmental factors is collected as part of each trawl survey used in the

> revised ARM, failure to use these data is a notable oversight of ASMFC's approach. By failing to use model-based standardization, ASMFC is implicitly assuming that these factors have zero effect on horseshoe crab captures -- which is a strong and likely false assumption.
>
> After a research scientist from the New Jersey DEP contacted me with their concerns, I re-ran my analyses with only the April and August samples. Finding no substantive difference in my results (and after running additional tests to confirm that the GAM standardization analysis was adequately accounting for the effects of seasonality), I proceeded with my original analyses in my 2023 peer review report. Of course, it is possible that there are legitimate reasons for a different analytical choice, and I would be happy to have a further discussion on the merits of sub-setting this dataset.
>
> Overall, I maintain that there are strong reasons to use model-based standardization methods (e.g., GLM, GAM, or machine learning approaches like random forest) to control for factors that could confound the inter-annual variation in catch-per-tow, and I am glad to hear the ARM subcommittee is open to using model-based standardization methods in future assessments.

Beyond the issue of the erroneous data standardization of the New Jersey Ocean Trawl Survey data by Dr. Shoemaker, he made a questionable analytical choice leading to the conclusion that female horseshoe crab abundance has not increased. Dr. Shoemaker used both the "raw" and "adjusted" catch-per-tow data from the entire time series of the three trawl surveys in a linear regression analysis to determine if there was a trend in abundance through time (Figure 6). The Delaware Bay crab population is known to have declined to a minimum level by the early 2000s (prompting harvest restrictions), thus, a linear model fit through the entire time series (1990 to present) of all surveys is nonsensical. The near zero slope of the linear model is driven by the high CPUE from the Delaware Trawl Survey at the very beginning of the time series (1990 – 1992). That horseshoe crabs declined in the 1990s and early 2000s is undisputed. All surveys show a low point around 2010, with an increase afterwards. The pattern of the combined surveys looks like a "U" – decreasing and then increasing. A linear model fit to such a pattern will show a non-significant slope (i.e., trend) over the entire time period. It is unclear whether Dr. Shoemaker investigated the resulting residual pattern, as that would have confirmed the inappropriateness of using a simple linear trend model. Perhaps this analysis is indicative of Dr. Shoemaker's unfamiliarity with the changes in horseshoe crab harvest management through time, but it nevertheless perpetuates the unfounded belief that the horseshoe crab population has not responded positively to harvest restrictions. As previously stated in the rebuttal to Criticism 2, all surveys have shown an increasing trend since 2010 (Figure 5). Alternatively, a segmented regression model could be fit to the time series of data to demonstrate how abundance trends have changed through time. When this is done, both the Delaware and New Jersey Ocean Trawl Surveys show declining abundance followed by an increase after 2010 (Figure 7). Given the lengthy time to maturity of horseshoe crab, it has long been understood that it would take about a decade to begin seeing an increase in abundance following the initiation of harvest restrictions.

After further consideration, I agree that there is a detectable statistical signal of a recent population increase in the trawl data. I also agree that horseshoe crabs are a long-lived species, and one would expect to observe a substantial delay between the implementation of harvest regulation (in 2000) and the recovery of the population (a large cohort born in 2005 would only mature and contribute to population growth in 2015 or later). Therefore, (1) time periods prior to the initiation of harvest regulations should not be included in this analysis (Fig. 12 from my 2023 peer review report), as few would claim that the horseshoe crab population was increasing in the 1990s (in fact, excessive commercial harvest of horseshoe crabs in Delaware Bay during this decade is widely believed to have caused a major decline in both horseshoe crab and shorebird populations; Niles et al. 2009) and (2) a segmented or nonlinear regression model makes sense for this analysis, as it can accommodate an initial period of decline or depletion followed by a more recent period of increase (e.g., decline in capture rates during the early 2000s followed by a recent recovery since around 2015).

For the sake of completeness, I am including a revised version of Figure 12 from my 2023 report that has been updated to use a GAM and GLM (with a quadratic relationship with time) to allow for a non-linear relationship with catch per unit effort over time (Fig. 4). Both methods yield the same result: an increase in the abundance of female horseshoe crabs since around 2010, indicating that the trawl surveys (considered together) contain evidence for a recent increase in female horseshoe crab abundance.

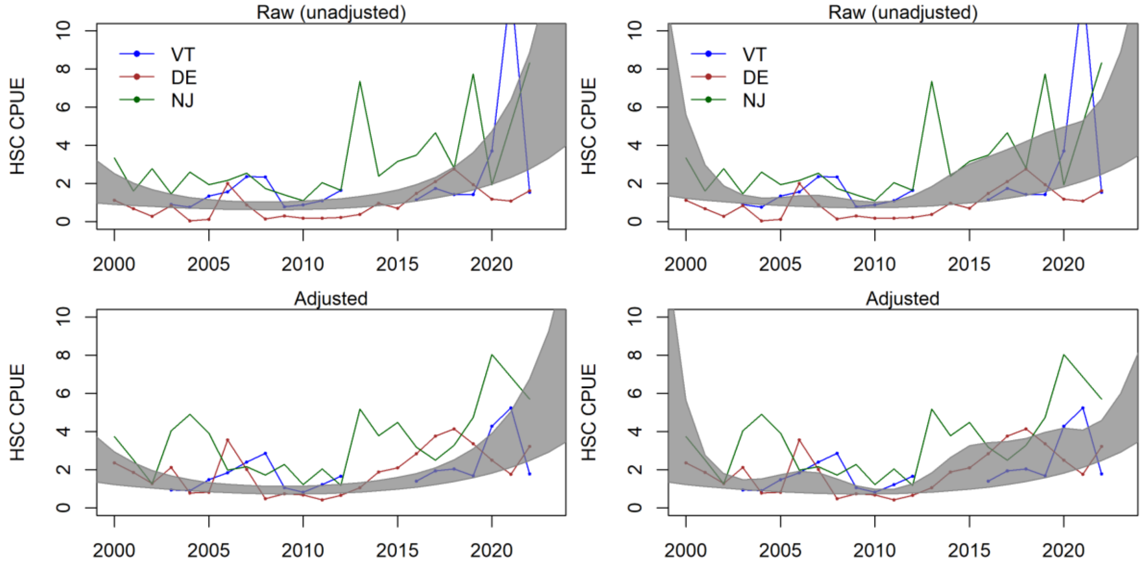


Fig. 2. *Updated version of Fig. 12 from my 2023 report, modified to add a (left) quadratic and (right) spline (GAM) trend of horseshoe crab catch-per-unit-effort (CPUE) over time. Both methods suggest a positive trend in female horseshoe crab abundance beginning around 2010, regardless of whether the raw or adjusted CPUE estimates are used.*

Dr. Shoemaker also reanalyzed egg density data from New Jersey to further argue that horseshoe crab abundance has not increased. These data were published by Smith et al. (2022) and showed a variable but increasing trend in egg densities over the last two decades (Figure 4). However, upon reanalysis, Dr. Shoemaker contradicts Smith et al.'s (2022) conclusion for an increasing trend, suggesting that it was an artifact of differing sampling methodologies through time. There is not much the ARM Subcommittee can say concerning trends in egg density data beyond what is published by Smith et al. (2022) because those data were not supplied to the ARM Subcommittee when requested during the ARM Revision. The acknowledgement by Dr. Shoemaker of the changing methodology in egg density data does corroborate one of the reasons the ARM Subcommittee has been reluctant to make use of egg density data since the development of the original ARM Framework in 2007. If the owners of the egg density data would follow the established ASMFC data acquisition processes by sharing the data when requested at the beginning of a stock assessment, the ARM Subcommittee would certainly evaluate the utility and inclusion of such data in the ARM modeling process just like any other data source.

> Notably, the strong positive relationship between horseshoe crab egg density and red knot survival did not depend on whether or not I used the results from Smith et al. (2022) or my adjusted numbers. I am glad that the ARM subcommittee is open to using these data in the ARM framework.

Criticism 5: The integrated population model used for estimating red knot population parameters is overparameterized and likely to yield spurious results.

- Dr. Shoemaker's criticism of the red knot model is unsubstantiated and misrepresents the models used in the ARM Framework.
- Much like the trawl surveys, the red knot data are imperfect but represent the best available data.
- Dr. Shoemaker assumes that too many parameters will produce incorrect results, when the relationship between overparameterization and biased models is more nuanced.

<u>Technical Response:</u> The critique of the state-space model ignores the fact that this model is not analyzed independently, but as a sub-model within an integrated analysis. This viewpoint is apparent in several places in Dr. Shoemaker's critique, as he writes about using the two data sources (i.e., red knot count data and mark-recapture data) to "train" the two sub-model components as if they were separate endeavors where information from one has no influence on the model parameters in the other. Integrated population models combine the likelihoods of two or more sub-models, allowing researchers to estimate demographic parameters from multiple models and data sources simultaneously (Schaub and Abadi 2011). In the ARM Framework, the admittedly limited count data are integrated with 100,000s of mark-resight observations from Delaware Bay. A third component, a Markov population model, provides a strong structural prior that links estimates from multiple sub-models based on an understanding of the life history of the species. One key benefit of this approach is the ability to estimate parameters that would not be estimable with any one model or data source alone. In the case of

the ARM Framework, the estimation of the red knot recruitment rate is informed by both the analysis of the count data (state-space sub-model) and the mark-recapture data (open robust design sub-model).

First, I point out that integrated population models (IPMs) such as the red knot model used in the revised ARM framework are complex, and their statistical properties are not fully understood by practitioners or statisticians (Schaub and Kery 2021). Second, I do not dispute the value of integrated models for conservation and management, and I agree that the red knot IPM is an integrated model whose components borrow information and inherit constraints from one another. However, it is also true that (1) the red knot IPM consists of two primary submodels (state-space submodel and open robust design [ORD] submodel), (2) the available data sources do not contribute equally to informing each submodel, and (3) one of the available data sources is very information-rich (the banding and resighting data, with tens of thousands of observations each year) and the other is very information-poor (the peak count data, with a single observation per year). Therefore it is instructive to deconstruct this IPM into a set of separate component models for heuristic reasons even as we acknowledge this is not strictly the case.

Ultimately, the red knot population simulation model (used for optimizing the harvest functions and fit within the red knot IPM) is a Markov population process described by (1) initial adult abundance, (2) adult survival (including an effect of horseshoe crab abundance in addition to arctic snow cover and spawn timing), and (3) recruitment (a compound parameter incorporating reproduction and first-year survival, also including an effect of horseshoe crab abundance). The information-rich data source (the mark-resight data) primarily informs the open robust design (ORD) submodel, resulting in well-informed estimates of annual survival (although likely biased high due to misread errors; see earlier discussion). Importantly, the information-rich mark-resight data are virtually non-informative with respect to two of the three demographic processes: initial abundance and recruitment. The reason for this is that the ORD likelihood (like all Cormack-Jolly-Seber variants) is conditioned on the initial capture event and is therefore only informed by the history of subsequent recaptures (i.e., it isolates the survival and state-transition processes from other demographic processes such as abundance and recruitment). Aside from survival, the ORD submodel (informed by the mark-resight data) is also used to estimate the fraction of the flyway population using the Delaware Bay staging area each year -- a process that appears to be mis-specified in ASMFC's red knot IPM (see discussion of 'pi' parameter below).

With the information-rich mark-resight data contributing little to the critical initial abundance and recruitment processes, the information-poor source of data (the peak count data) necessarily does the heavy lifting when it comes to estimating these parameters (but contributes very little to the survival estimates). Some components of the state-space submodel are informed by the information-rich dataset- notably, the fraction of the stopover population available to be observed during each 3-day interval is derived largely from the ORD submodel but forms an important part of the state-space likelihood.

> However, this “cross-over” between the two likelihoods does little to mitigate the central issue that the information-poor peak count data is the primary source of information for estimating red knot recruitment and initial abundance.
>
> The ARM subcommittee seems to be making the claim that the recruitment parameters are estimated jointly from the mark-resight data and the peak-count data. While true in a strict mathematical sense (in any integrated model there will likely be at least some information leakage among the joint model components), this is not the case in any important practical sense. As I mentioned above, the way these data enter the likelihood function, as specified in the L1 component of the open robust design (ORD) model, ensures that this data can only directly inform the survival process (along with temporary emigration and some within-year processes like the timing of stopover entry and exit). Effectively, the information-poor peak-count data are used to estimate initial abundance as well as the changes in abundance from year to year (annual lambda, or population growth). The model then solves for the unknown recruitment rates, conditional on the estimated survival rates (from the mark-resight data) and the annual population growth rates (from the peak-count data). The ORD model by itself is largely uninformative with respect to recruitment- it is the addition of the peak-count data that makes it possible to estimate recruitment. Therefore, it is disingenuous to claim that the mark-resight data contribute to the estimation of recruitment in any real sense.
>
> Finally, a claim like “the admittedly limited count data are integrated with 100,000s of mark-resight observations from Delaware Bay” ignores the fact that the 100,000s of mark-resight observations contribute virtually no information for fitting two of the three key demographic processes estimated by the IPM: abundance and recruitment. The implication that the red knot IPM is rescued from standard statistical concerns (such as over-fitting to the data) because it borrows information from the information-rich band-resight observations to supplement deficiencies in the information-poor peak-count data is misleading and dangerous. It can become all too easy to claim “empirical” support for poorly specified or unsupported model components by making facile but rhetorically appealing claims about integrated likelihoods. For this reason, it is very important to break down these complex models (for heuristic reasons) into their subcomponents and discuss which data sources are doing the heavy lifting for fitting all key parameters-- at least until the statistical properties of integrated population models are more fully understood and documented.

By ignoring the structural linkage that shares information between model sub-components, Dr. Shoemaker set up a misleading basis to make unsubstantiated claims about model overparameterization and to falsely demonstrate spurious results produced by the ARM model. Regarding overparameterization, he referred to the familiar rule-of-thumb of 30 data points per model parameter as sample size guidance for robust estimation. While this guidance is useful in traditional applications where data are used to inform the parameters of a single model, its relevance for integrated modeling – where information is shared across multiple model components – is unclear. His assessment that 18-28 parameters were estimated from 14 data points is a serious mischaracterization of the model and requires overlooking the fact that information from mark-resight data also informs the state-space model. In the ARM Framework, the number of parameters estimated from the count data alone is three: one initial population size and two counting errors. The recruitment parameters (three parameters: mean, variance, and effect of horseshoe crab abundance) are estimated jointly using information from all three components of the integrated population model. The availability parameters are specified with highly informative priors, which were developed externally to the model. In the ARM Subcommittee's view, the availability parameters should be more appropriately thought of as data informing the model, not estimates on which inference was based.

> I do not think it is misleading, unsubstantiated or false to claim that the peak count data are the primary source of data for estimating recruitment and initial abundance. In counting up the number of parameters estimated primarily using the 14 peak-count data points I acknowledged that some of these parameters (such as the 'availability' parameters) were assigned strong priors, and that some represented individual random effects (for which the calculus for estimating degrees of freedom is unclear). I dispute that any of the parameters in Table 2 (including the recruitment parameters) are estimable on the basis of the information-rich mark-resight data. Therefore, there are at least 8 to 10 free parameters (and probably more) estimated primarily from the information-poor peak count data- which approaches or even exceeds the available sample size (*n* = 14). As IPMs have poorly understood statistical properties, I referenced a common rule of thumb in statistics that is generally relevant to non-informative statistical models (those without good prior information).

Dr. Shoemaker used a simulation exercise to purportedly demonstrate production of spurious results by the model. By replacing the peak counts with white noise in the simulation runs, he anticipated that the simulated abundance at the end of the time series should match the initial abundance on average. Instead, he was surprised to discover negative trends in simulated abundance and that final abundances produced by the model were most often lower than initial abundance. He did not know the cause of this outcome, and he speculated on a variety of reasons having to do with simulation methods, starting values, etc. The cause is simple to explain, but it requires acknowledgement that the information sources are linked to each other through the Markov population model. By providing a stream of pattern-less peak count data to the model, Dr. Shoemaker effectively contaminated information about recruitment, leaving survival rate as the only reliably informed parameter. Therefore, a population simulated with no recruitment and survival probability <1 will most often decline. Though he failed to understand the cause of the observed simulation behavior, and he cautioned against using his results to infer a systemic bias in the model, he nevertheless concluded that the model is unstable and has a strong tendency to produce spurious results.

> IPMs are a relatively new - and particularly complex - class of ecological models, and the statistical properties and biases inherent to these models are poorly understood by statisticians. It is possible that the simple tests I ran using "white noise" (random numbers from a normal distribution with mean, variance and sample size that matched the peak

count data and with no temporal trend) to substitute for the peak count data (which was meant to assess the tendency for spurious estimates of growth or decline) may not have been sufficiently informative. It is also possible that the constraints introduced by the Markov population model had the effect of inducing a negative bias in these tests. Nevertheless, the rationale provided by the ARM subcommittee seems overly simplistic. While it is true that a population will necessarily decline with zero recruitment (and survival <1) the explanation for this issue is certainly more nuanced; in my tests, the IPM estimated recruitment as a free parameter- and recruitment was constrained to be greater than zero. Nonetheless I had limited time to run tests, and given the results of my simulations (well over half of the tests resulted in an estimated population decline) an unintended source of bias may have affected my test results.

A better (but more time consuming) validation test would be to develop a complete simulation of the *rufa* red knot population, including a demographic process model (including survival, fecundity, abundance) and an observation model capable of generating data similar to the real-world system (including mark-resight and peak-count data) under a wide range of demographic scenarios (e.g. differing levels of survival and recruitment), and a wide range of observation error scenarios. With simulated data from such a model, researchers could test how often the IPM was able to successfully recover the true parameter values, including recruitment, variation in recruitment, and covariate effects on recruitment (including HSC abundance). The open robust design submodel has been extensively tested using similar tests with simulated data (Tucker et al. 2022), but I did not find any evidence that the full IPM was subjected to similar validation tests. If they did run simulation-based trials using data generated under known assumptions and parameter values, they did not report the results in the ARM report or in Tucker et al. (2023)(or in the code release for the IPM). Such tests require a good deal of time and thought to develop and run. However, investing such time and thought in such testing is necessary and important given the central role of the IPM in informing important ASMFC policy decisions affecting a threatened species.

Integrated population models are complex and largely untested, and there are unintended biases that can occur (Riecke et al. 2019), so it is important to test these models extensively, especially when used in the context of decisions that can detrimentally affect threatened and endangered species. Therefore, the ARM subcommittee should run a battery of validation tests before concluding the model is stable and that it reliably is able to recover key demographic information about the system -- including temporal variability and covariate effects. We cannot assume that complex models like the red knot IPM are free from serious biases and other statistical issues. Because they are relatively new and untested, IPMs should be presumed flawed until they have been adequately validated (such as running the simulation tests described above) -- this is especially true for an IPM that is used for making important decisions that could impact a threatened or endangered species. In this case, the burden is on ASMFC to demonstrate that the red knot IPM is capable of serving its intended role in the revised ARM.

The critique of the state-space sub-model also contains an assertion that overparameterized models are necessarily biased. While overparameterization can result in poor generalization to new datasets, it does not guarantee biased results. In fact, bias could also arise if models are under-parameterized and fail to capture system complexity. The relationship between bias and overparameterization is not as straightforward as is portrayed in Dr. Shoemaker's report.

> Indeed, over-parameterized models are not guaranteed to be biased. Instead, over-parameterized models tend to overpredict the training data (predicting the data used for training the model with high precision) but perform very poorly when confronted with independent data not used to train the model (out-of-sample data). The fact that the red knot IPM is being used to predict the population response to harvest management in the future means that over-parameterization could be a serious issue for the revised ARM.
>
> The above point about under-parameterization is important and relevant to this discussion. The trade-off between under-parameterization and over-parameterization is often known (somewhat confusingly) as the "bias-variance" trade-off. In this case, the term bias refers to under-parameterized models, which can provide biased estimates even for the data used for training. The term "variance" refers to the property of an over-parameterized model making inaccurate and often wildly off-base predictions when challenged with new data (the model treats the noise in the training data as if it were a useful signal, and therefore models fitted to different samples from the same statistical populations will make very different [variable] predictions despite the fact that the data samples reflect identical underlying processes). In general, over-parameterization can be assessed by withholding some data from the training set and testing to see how well the model is able to predict the left-out data. This is an important part of the model validation process -- and one that could add substantial credibility to the red knot IPM if applied to the Delaware Bay system.

The ARM Subcommittee readily acknowledges that the red knot count data are a much weaker data set than the mark-recapture data, but they were the only count data collected consistently over the all of the years of the monitoring program, so the ARM Subcommittee made the best use of them to better understand the system. As described in ASMFC 2022 (page 80), this model could be greatly improved by including auxiliary information such as survey-specific covariates (e.g., observer ID, tide state, weather conditions), integration of simultaneous ground count data, or future implementation of digital photography or double-observer methods. One of the challenges of working with historical monitoring data is the inability to influence study design or data collection processes. There were no auxiliary data that were consistently collected (or, at least, made available to the ARM Subcommittee) for aerial surveys that would allow counting error to be better estimated. Similarly, the ARM Subcommittee knows that concurrent ground counts were conducted in at least some years, but those data were not provided. The ARM Subcommittee made the best use of the available data, and conducted these analyses within the management decision context. Sometimes in decision support roles, scientists have to develop the best analysis to support decisions even when data are imperfect (McGowan et al. 2020). All modeling exercises require assumptions and constraints, and those included in this model represent the best understanding of the system at this time; the ARM Subcommittee hopes and intends for this model to be updated as more information and more data become available. It should be noted that all previous attempts to model red knot populations in this system and assess the linkages between knots and horseshoe crabs in this management context required significant assumptions, and the ARM Subcommittee believes that their approach in the ARM Revision alleviates or improves many of those assumptions. Previously, all attempts to model productivity and recruitment in this population relied upon estimates from Europe and basic assumptions about life history (i.e., setting juvenile survival as a percentage of adult survival, see McGowan et al. 2011) and this approach uses data from this flyway in a complex but much improved model to estimate those parameters.

> I appreciate the thoughtful discussion on the low information content of the count data and ways in which this critical information source for the IPM model could be improved in the future. Overall, I maintain that the peak count data are asked to do some heavy lifting in the red knot IPM for which they are ill-suited.
>
> Stating that this is a ‘much improved model’ does not make it so. Complex models like the red knot IPM must be subjected to rigorous testing, and it appears the IPM (unlike the open robust design subcomponent) has not been adequately tested (see above). Also, I do not really understand why the use of data from other populations (e.g., European red knots, which have a similar life history) and time periods (e.g., the period of recent population declines in the early 21st century) is so heavily devalued by the ARM subcommittee. If there is useful information on the recruitment process that can be gleaned from other populations, why not use this information? I am not sure it is an improvement to use only data from the western Atlantic flyway if the best available information for this population comes in the form of 14 low-precision data points.

Criticism 6: The integrated population model exhibits poor fit to the available data.

- Dr. Shoemaker provides conflicting arguments for the use of the goodness of fit test for the red knot model.
- Goodness of fit tests applied to the red knot model indicated poor fit in one model component, but the portion of the model including the survival probability of red knots did not fail the test.

Technical Response: There are no unified goodness of fit tests for integrated population models, so the commonly-accepted approach is to assess model fit independently for each sub-model. Posterior predictive checks (PPCs) are the standard type of goodness of fit tests for Bayesian models. The PPC for the state space model indicated adequate fit (*P* = 0.44 where *P* = 0.5 indicates no evidence of either over- or under-dispersion, and *P* near 0 or 1 suggests poor model fit), but the PPC for some components of the open robust design model indicated lack of fit to the data.

> I also made this point in my 2023 report, but I agree there are no unified goodness of fit

> tests for IPMs, and that PPCs (in spite of some known flaws) are currently the preferred method for checking model adequacy. Nevertheless, I was not able to confirm adequate fit for any of the three subcomponents of the open robust design submodel, including the likelihood component responsible for estimating adult survival. I was able to confirm that the PPC for the state-space model indicated adequate fit, but the most authoritative available manual for IPMs (Schaub and Kery 2021) notes that this test has been shown to indicate model validity even in cases in which the model is demonstrably not valid. Therefore, following Schaub and Kery (2021), I do not consider the PPC results for the state-space model to constitute convincing evidence for adequate model fit (as I stated in my 2023 peer review report).

This critique contains shaky logic. First, Dr. Shoemaker asserts that PPCs are a good method for checking model fit and criticizes the lack of fit of the open robust design model. Indeed, Dr. Shoemaker used a PPC in his analysis of banding data to conclude that his model had "reasonable fit." Next, he states that PPCs are not a reliable indicator of goodness of fit to cast doubt on the ARM Subcommittee's statement that the state space model "passed" the test. By Dr. Shoemaker's logic, PPCs are only to be trusted when they indicate lack of fit. Dr. Shoemaker's inconsistent logic with respect to checking goodness of fit casts doubt on the integrity of the analysis. Putting that aside, the apparent lack of fit for the open robust design model will be discussed. The open robust design model consists of three likelihoods, and PPCs indicated lack of fit for likelihood L3 (P = 0.9), which describes the process of reencountering individuals within years. This lack of fit could arise due to unmodeled heterogeneity in true arrival and persistence probabilities as a result of pooling encounters into three-day sampling periods. If aggregations occur over a time period that is short relative to the expected length of stay, the expected bias is minimal (Lindberg and Rexstad 2002; O'Brien et al. 2005). Average stopover duration for red knot at this site has been estimated to be 12 days (Gillings et al. 2009); 3 days should be a short enough window to avoid biased estimates of arrival and persistence but could introduce heterogeneity and overdispersion. The likelihood that contains the apparent annual survival probability is likelihood L1, which describes the process of encountering marked birds across years. PPCs for this likelihood did not indicate lack of fit (P = 0.31).

> The ARM subcommittee misunderstood my argument in my 2023 report (see above). I did not state or imply broadly that PPCs are not useful in the context of IPMs. The only PPC test I raised questions about was the PPC test specifically for the state-space model; the PPC test (Bayesian *p*-value) in this particular case has been shown to indicate adequate fit even in cases where the model is known to be incorrectly specified (Schaub and Kery 2021). I did not broadly question the value of PPCs, nor did I unfairly imply that I only trust PPCs when they indicate lack of fit. Indeed, I used PPCs to assess goodness-of-fit for my survival models, and I used any indications of lack of fit as motivation to improve these models. In my tests with the red knot IPM, the open robust design subcomponents all exhibited poor fit to the data, whereas the state-space component exhibited adequate fit (as stated above).

> In addition to the points raised by Schaub and Kery (2021), questioning the value of the PPC results in the context of the state-space component, it is important to note that "passing" posterior predictive checks is much more challenging for rich data sets like the mark-resight data and much less challenging for smaller datasets like the peak-count data (the primary data source for fitting the state space model). Therefore, "passing" PPC-based tests for very small datasets like the peak-count data can be a pretty low bar that does not generally validate model adequacy.

CONCLUSIONS

Continuous scientific review and critique is welcome as that is how science advances. There will always be room for improvement in any modeling effort in the management of natural resources. This is part of the double-loop learning in an adaptive management effort whereby model design and management are periodically reevaluated (Fabricius and Cundill 2014; Williams and Brown 2018). In this specific case, however, advocacy is infused into the scientific debate. The 2022 ARM Revision represented some great advancements in the understanding of the population dynamics of horseshoe crabs and red knots, and their interactions during the double-loop of the adaptive management process.

> I agree about the value of scientific critique and debate, and I hope this exchange is useful for advancing scientific understanding of this system. I have taken my role as an independent scientific reviewer seriously, and my critiques are meant to ensure rigorous use of the best available science in this important decision-making context. I think it is unfair to claim that I infused advocacy into the debate or undermined the scientific process in any way.

It is curious that these advancements have stirred so much controversy because the technical criticisms of the ARM Revision could have equally applied to the original ARM Framework. In fact, the original framework merited specific criticism because it relied on life history parameters informed by literature values taken from outside the Delaware Bay or based on expert opinion. The ARM Subcommittee questions if the true problem is not with the process or technical modeling, but rather with the final result and harvest recommendation.

> It seems clear that if a model recommends action that could potentially harm a threatened or endangered species (or impede their recovery), it is only prudent that the model is subjected to increased scrutiny.

An important benefit of the adaptive management process is the ability to make decisions even under imperfect knowledge of an ecological system (Williams et al. 2002). The overall goal of the ARM Framework was to produce a decision tool informed by science and stakeholder values, given the available knowledge about the Delaware Bay ecosystem and horseshoe and red knot population dynamics. In the original ARM Framework, knowledge about some system components, for instance red knot population dynamics, was quite limited. The ARM Revision represented a double-loop learning event, in adaptive management terms, and population models were re-designed to accommodate 1) the large volumes of high-quality data collected on both species since the original ARM's inception, and 2) changes to both populations over that period. In the view of the ARM Subcommittee, the effect of a change to an ecological model must be judged according to its effect on both the properties of the overall decision framework, and the ability of the ARM Framework to incorporate new monitoring data to improve understanding of the system. One important goal in the development of the ARM Revision was to design population models for horseshoe and red knot that would allow for rapid and efficient learning given the monitoring efforts in place for each species (Williams 2011). This critical feature of the ARM Framework—the ability to learn from monitoring—is not addressed by Dr. Shoemaker or Earthjustice; and yet it was a major consideration by the ARM Subcommittee. The design of ecological models for use with adaptive management should also be guided by the decision objectives (Fuller et. al. 2020), a point not addressed by Earthjustice.

> I generally agree that adaptive management has great value for managing systems in the face of uncertainty. However, I think a multi-hypothesis approach to adaptive management is essential for capturing the spirit of adaptive management (see opening statement #1). By accommodating a range of plausible models of the system, including at least one model that formalizes a strong and ecologically meaningful link between red knots and horseshoe crabs, a multi-hypothesis approach to adaptive management will better encapsulate the scientific literature on this system (in which a strong relationship between these two species is indeed plausible). In addition, from a purely scientific perspective, a multiple hypothesis approach can yield more effective inference than a single model approach (Platt 1964). Finally, this approach is better able to accommodate the full spectrum of values within the stakeholder community.

Much of the 2022 and 2023 criticism by Dr. Shoemaker (as well as the comments by Earthjustice) stem from the belief that there must be a strong relationship between horseshoe crab abundance, horseshoe crab egg density on the beaches, and red knot survival. They claim that because the ARM Subcommittee did not find this “strong” relationship when examining the empirical data from the Delaware Bay region, the ARM Revision must therefore be fraught with error. It is apparent that Dr. Shoemaker reviewed the ARM Subcommittee’s work with an unwillingness to entertain the idea of anything but a “strong” relationship. A specific example of this is his statement in his 2022 report where he postulated that the collection of additional data may show that the relationship between horseshoe crab abundance and red knots survival could disappear or become negative. He states, “This outcome would pose an existential problem for the ARM Framework, decoupling the two-species Framework and rendering the red knot model unusable in the context of management.” Of course, the “no relationship” outcome would be expected if horseshoe crabs become sufficiently abundant to not limit red knot survival, but that knowledge does not challenge the scientific validity and usefulness of an adaptive management framework for decision making. Such comments demonstrate a reluctance to learn within an adaptive management framework and a desire to cling to previous beliefs in spite of scientific advances.

> I think I was clear: the only point of including a red knot population simulation model within this ARM framework is because of the potential risk to this population posed by

> horseshoe crab harvest. If the model showed no response of the red knot population to horseshoe crab harvest (even under scenarios involving an extreme collapse of the horseshoe crab stock) then there would be no point in including a red knot simulation model as part of the ARM framework in the first place. Please refer to opening statement #2 for more discussion about the rationale for focusing on the strength of the relationship between red knots and horseshoe crabs.

There is no doubt that Dr. Shoemaker is a very knowledgeable quantitative ecologist. However, his critiques are unhelpful in advancing a two-species adaptive management effort. His criticisms focus on specific components of the overall ARM Framework, and why each may be wrong, but nowhere does he provide any recommendations for how to assemble the pieces into a unifying framework to make management decisions. For example, he makes strong arguments for using egg density to predict red knot survival but provides no recommendations for how to link egg density to female horseshoe crab abundance, which is directly affected by harvest management. He also makes a large issue about uncertainty in the horseshoe crab population projections but fails to recognize how uncertainty is handled in the optimization (approximate dynamic programming) or make any recommendations on alternative methods to conduct an optimization given the uncertainty.

> As an independent peer reviewer, my primary goal was to review the existing ARM framework on its merits and not to provide a vision for how this system could be improved. Nevertheless, I suggest that a multi-hypothesis approach could offer important benefits in this case, and I would be very happy to engage in further discussions with the ARM subcommittee.

The ARM Framework is designed to continuously improve the underlying models through double-loop learning, and the ARM Subcommittee welcomes constructive input on how to do so. Unfortunately, the critiques by Dr. Shoemaker (and Earthjustice) fail to make any real recommendations for improvement or provide any other means for helping managers make an informed harvest decision beyond consideration of the values of a single stakeholder group. If the values of all stakeholders have changed (i.e., no female harvest under any circumstances), that change could be considered in a new approach in the future by the ARM Subcommittee. As it stands, the current ARM Framework represents the values previously established through stakeholder engagement: to manage harvest of horseshoe crabs in the Delaware Bay to maximize harvest but also to maintain ecosystem integrity, provide adequate stopover habitat for migrating shorebirds, and ensure that the abundance of horseshoe crabs is not limiting the red knot stopover population or slowing recovery.

> While it was not my role to suggest recommendations for improvement, I hope ASMFC considers adopting a multi-hypothesis ARM framework. I certainly do not advocate for a framework that only considers the values of a single stakeholder group, and I hope ASMFC can find a way forward that uses science to bring stakeholders together rather than driving them further apart.

Criticism 7: The estimate of mean horseshoe crab recruitment and propagation of error

within the horseshoe crab population dynamics model is inappropriate.

- The estimate of mean horseshoe crab recruitment used by the ARM Subcommittee is the most biologically realistic. If mean recruitment were lower, as Dr. Shoemaker suggests, the current population estimate of horseshoe crabs would be well above a predicted “carrying capacity” of the Delaware Bay region.
- Dr. Shoemaker’s proposed method of error propagation is worth considering in a future revision of the ARM model, but comparison of his population projections to those by the ARM Subcommittee are nearly identical.

Technical Response: The revised ARM Framework uses the same mathematical model to estimate the abundance of horseshoe crabs (the CMSA) and to project the horseshoe crab population into the future while accounting for annual removals of individuals due to bait harvest, dead discards from other fisheries, and mortality associated with biomedical facilities. In his 2022 critique, Dr. Shoemaker expresses his opinion that uncertainty in model parameters was not propagated through time in an appropriate manner. This criticism does have some merit and his proposed methodology is worth the ARM Subcommittee considering in future revisions of the ARM Framework. Dr. Shoemaker contends the current horseshoe crab projection model greatly underestimates uncertainty and its effects on predicted future abundance. Although Dr. Shoemaker’s proposed methodology may be more appropriate, the ARM Subcommittee believes these concerns are overstated as there is still much uncertainty in the projected population – female horseshoe crab abundance can range between 5 – 15 million under a no harvest scenario.

> I agree that the proper treatment of uncertainty is critical for decision making and I am glad to hear that ASMFC is considering incorporating some of the changes I suggested within future iterations of this ARM framework.

Another parameter Dr. Shoemaker criticized was the estimate of mean horseshoe crab recruitment because of the gap in the Virginia Tech data from 2013 - 2016. The ARM Subcommittee agrees that CMSA estimates of recruitment during these years are poor; therefore, the average of them was used when calculating the overall mean recruitment level. One could argue that recruitment estimates during the Virginia Tech gap years should simply be thrown out. However, doing so ignores the obvious above-average recruitment during those years that must have occurred to increase the multiparous population to the degree that was observed in the following years. The treatment of the missing years of recruitment data balanced the nonsensical estimates of the CMSA with the biological reality that recruitment during these years had to have been relatively high. All other things being equal, changing the mean female horseshoe crab recruitment from 1.67 to 1.26 million, as suggested by Dr. Shoemaker, would result in an unexploited population size at equilibrium of 6.4 million (95% CI: 3.4 – 14.5 million) compared to 8.5 million (95% CI: 4.5 – 19.2 million) in the current parameterization of mean recruitment. If Dr. Shoemaker were correct in his estimate of mean recruitment, the latest population estimates from the Virginia Tech Trawl Survey swept area estimate and CMSA are well above this equilibrium level and the population will likely decline even in the absence of any harvest. It is also interesting to note that Smith et al. (2006) estimated the female population size via a mark-recapture study at 6.25 million in 2003, shortly after the period of high horseshoe crab harvest. This is another line of evidence that the mean recruitment parameter used in the ARM Framework (1.67 million) is more appropriate than the one proposed by Dr. Shoemaker (1.26 million) given the observed increases in female abundance since the population was estimated by Smith et al. (2006).

First, it is important to point out (as I did in my 2022 report) that the mean recruitment rate parameter is as critical to this ARM framework as any other parameter, since the recruitment process determines the degree to which the horseshoe crab population is resilient to harvest. Therefore, the methods used by ASMFC to estimate horseshoe crab recruitment deserve special scrutiny.

While I understand the rationale of the ARM subcommittee for using the average recruitment estimate from the CMSA model from the Virginia Tech (VT) gap years when computing the mean recruitment rate parameter, I do not find this rationale convincing. If the CMSA results for these years were nonsensical (which we all agree upon), it does not necessarily follow that the arithmetic mean of those nonsensical results will be meaningful. In general, when a model produces nonsensical results, it should provide a signal to the modelers that there is something fundamentally wrong with the model. Furthermore, although the mean recruitment rate during the VT gap years is more sensible than the wildly non-credible estimates for the individual years, the mean value across these years (for which no data was available) was still greater than any single year for which data were available. In this sense, the mean value for the VT gap years also seems inconsistent with the data; such a discrepancy should prompt a re-evaluation of the underlying assumptions, and (ideally) modifications to the model that bring the model more in line with real-world observations of the system.

The ARM subcommittee argues that recent estimates of multiparous abundance from the Virginia Tech trawl are most consistent with the CMSA model results. Specifically, they argue that mean recruitment (under the CMSA model) would need to be higher than the estimate I suggested in my 2022 peer review report (which was based only on the years for which data are available) in order to produce an equilibrium abundance consistent with recent abundance estimates. This argument requires two assumptions: (1) the current horseshoe crab population is at an equilibrium state, and (2) most importantly, that their simulation model is an adequate representation of the horseshoe crab population. However, the nonsensical results from the VT gap years casts serious doubt on the adequacy of the model in the first place (see above).

I do not follow the argument regarding the Smith et al. (2006) study so I will not comment further on that point. Overall, the use of a “worst-case” scenario is commonly used in cases where a risk-averse approach is warranted (for example, when, as here, an action has a risk of harming a threatened or endangered species). In this case, the worst-case scenario (recruitment of 1.26 million) is also supported by the only available data source directly relevant for estimating recruitment rates for this population: the VT trawl surveys.

Therefore, I maintain that there is a strong case for including this as a plausible value to represent mean recruitment in this poorly understood population.

Dr. Shoemaker shows his female horseshoe crab population projection from his reformulated Bayesian CMSA model that includes his parameterization for recruitment and method for propagating uncertainty. It is interesting that given all his criticism of the ARM model, his model produces nearly identical results with respect to an equilibrium number of primiparous and multiparous females (Figure 8) and associated uncertainty. If anything, his equilibrium population size may be slightly higher than what the revised ARM Framework predicts and the uncertainty on each seems equivalent.

Simulation results from my Bayesian CMSA model were similar to the results from the ASMFC simulations under baseline conditions. However, a more important test would be to see if these two models produce similar results under a more extreme harvest scenario: that is, whether the ASMFC framework properly represents the stability or instability of the system under plausible future harvest regimes. The simple tests I included in my 2022 peer review report indicated that the way the ASMFC model propagated uncertainty may have overstated the stability of this system and its resilience to harvest (Fig. 3 of my 2022 report, middle and lower panels). Additional tests would be required to confirm this hypothesis. Regardless, I think there is a strong case for ASMFC to revise the horseshoe crab simulation model to ensure proper treatment of uncertainty.

Dr. Shoemaker did not comment on the harvest policy functions, which are the mathematical equations that actually tell the ARM Subcommittee how many horseshoe crabs to harvest given the abundance of horseshoe crabs and red knots. He also did not comment on the Approximate Dynamic Programming (ADP) process by which the harvest policy functions were derived. When solving for the optimal harvest policy functions, ADP incorporated the full range of uncertainty in population projections for both horseshoe crabs and red knots, and within the ADP process, the optimal harvest policy functions would be more conservative with greater uncertainty. Thus, any recommendation of harvest coming from the revised ARM Framework explicitly incorporates uncertainty in population projections.

During my peer review of the revised ARM framework, I focused my attention on reviewing the demographic models, which was appropriate because this is my primary area of expertise.

Criticism 8: That the ARM model would not predict a decline in red knots under a total collapse of the horseshoe crab population is evidence that the model is fatally flawed.

- Dr. Shoemaker is incorrect that the ARM model would not predict a decline in red knots if the horseshoe crab population collapsed. The assertion that red knots would continue to increase in the absence of horseshoe crabs is mathematically impossible in the model.

Technical Response: In his 2022 critique, Dr. Shoemaker states, "…the apparent inability of the ARM model to predict a decline in red knot abundance under a total horseshoe crab population collapse…undermines the apparent purpose of the model." This judgment can be seen echoed throughout the materials submitted by Earthjustice in 2022 and 2023, where the narrative is peppered with claims of predicted red knot population increases even at complete depletion of horseshoe crabs from Delaware Bay. The critics' implication is this: if the model is unreliable at the population level of zero horseshoe crabs, how can it be trusted for harvest management at any population level of crab? This is an unfortunate and prejudicial coloring of the model because Dr. Shoemaker was wrong in his 2022 judgment. He not only failed to correct the false assertion in his analysis, but he also amplified it (p. 22) in his later critique.

In Dr. Shoemaker's 2022 critique, he acknowledged that he relied on a "back of the envelope" calculation to arrive at his conclusion because he lacked access to the model data and code at the time. Were he to obtain access to the materials, he fairly asked, "[w]hat would happen to the red knot population projections if female horseshoe crab abundance were set to zero?" For his 2023 evaluation, Dr. Shoemaker was provided access to the data and code, yet he failed to address his own question. He would have observed that the data used to establish the relationship between female horseshoe crab abundance and red knot survival was the logarithm of female horseshoe crab abundance (ASMFC 2022) and not female abundance as it comes straight from the CMSA estimates. Consequently, the model predicts that red knot survival declines to 0 as female horseshoe crab abundance decreases, and a population increase in red knots under this condition is mathematically impossible.

> This argument by the ARM subcommittee has more to do with mathematical technicalities than with ecology. Please see opening statement #2 for a detailed response to this comment.

Misunderstanding and mischaracterization of the model aside, prediction by any model for a scenario well outside of the data bounds of model development is a dangerous exercise. A complete loss of horseshoe crabs through harvest is an extreme and unlikely hypothetical scenario that was not considered by the ARM Subcommittee. Such a collapse would require a harvest level greatly exceeding any previously observed harvest level, let alone any harvest level that is within the range of possible values given the current fishery management plan stipulations. The critics should give the ARM Subcommittee and Board some benefit of the doubt: if the horseshoe crab population should fall below any historically observed levels, and outside the bounds of model development, the ARM Subcommittee is sure all would agree that horseshoe crab harvest should be drastically reduced or ceased. This demonstrates an attempt to sensationalize an extremely rare possibility and paint scientific management of the species as reckless.

> First of all, there is great heuristic value in understanding how the red knot population model, as implemented in the revised ARM, would fare under a collapse of the horseshoe crab stock. Importantly, this exercise illustrates that the ASMFC model, as currently specified, could not predict the observed decline of red knots in the late 1990s and early 2000s, which has been attributed largely to the decline of horseshoe crabs due to unregulated harvest in the 1990s (Niles et al. 2009). Notably, the original ARM used by ASMFC included candidate models with a stronger relationship between red knot demography and horseshoe crabs, and the modelers took care to demonstrate that these

models were capable of explaining the observed declines in the red knot (McGowan et al. 2011), thereby recognizing the value of performing this scenario test and of including a “strong interaction” model within the candidate model set.

Further, ASMFC argues that statistical extrapolation (making predictions outside the bounds of the data) can be dangerous and misleading. While there is some merit to this argument in a general sense, it ignores the fact that the model’s primary utility was to make predictions across a broad range of future scenarios. In the context of the ARM optimization routine, simulation results from scenarios spanning a wide range of horseshoe crab abundance and harvest rates are used to generate optimal harvest functions for use in setting harvest quotas. This exercise requires extrapolation- the red knot simulation model must be able to predict what would happen under scenarios of reduced horseshoe crabs (and/or increased and recovering red knot populations, which in aggregate may require a higher total abundance of eggs) to be useful for making informed decisions across a wide range of plausible future system states. Finally, if ASMFC argues that the model is valid only within a particular range of horseshoe crab abundance, they should identify that range and explain why such limitation doesn't raise broader concerns about the revised ARM framework.

I don't think anyone seriously believes (or has claimed) that ASMFC would continue recommending commercial harvest of horseshoe crabs in the face of an observed and ongoing collapse of the horseshoe crab population. But that is not the point of my analysis. The point is that the decision-making value of this framework requires that the underlying models are able to make reasonable predictions across a wide range of scenarios- including a major decline (or increase) in one or both species. The revised ARM proved unable to do so.

Finally, in reviewing the methods used by the ARM subcommittee to prepare the horseshoe crab abundance estimates for use in the red knot IPM, I noticed that they log-transformed the CMSA estimate (in units of millions) and used this log-transformed covariate directly in their analyses. In Bayesian modeling (and GLMs more generally) it is common practice to center and scale all covariates, which typically involves subtracting raw measurement by the sample mean (zero-centering), often followed by dividing the resulting quantity by the sample standard deviation. This practice is useful for enabling regression coefficients to be directly comparable, but even more importantly, zero-centering aids in model convergence by reducing collinearity among the free parameters being estimated (for example, it reduces collinearity between intercept terms and regression coefficients). In the red knot IPM, all covariates were centered and scaled prior to analysis, with the exception of horseshoe crab abundance (which was log-transformed but not centered and scaled). I point this out because it is a surprising choice by the modelers, and it may have added to the instability of model convergence and potentially influenced the model results. For this reason, and as an appropriately cautionary approach, I would recommend running some tests to ensure that this decision did not unintentionally influence key model outputs.

Criticism 10: There is an incorrect specification of “pi” parameter in the red knot integrated population model.

- This is a criticism that does warrant further consideration by the ARM Subcommittee.

Technical Response: Dr. Shoemaker asserts that there is a missing parameter that should be included in the derivation of π_{jj} (the probability of being present in Delaware Bay in occasion *t* of year *j*) to represent the fraction of the population using Delaware Bay in the previous year. This seems to be a valid criticism, but requires further scrutiny to understand whether this parameter is derived incorrectly and, if so, what the implications might be. The ARM Subcommittee is exploring solutions.

> I am glad to hear the ARM subcommittee is looking into this issue. I agree that the implications of this issue for the results of this analysis are unclear- and not necessarily minor.

Criticism 11: There is an over-representation of Mispillion Harbor in red knot resighting data.

- Use of data from Mispillion Harbor does not result in biased inferences.

Technical Response: More resighting data is collected in Mispillion Harbor than any other site in Delaware Bay. However, red knots move around the Bay during the stopover period and are often resighted in more than one location within a year. The open robust design sub-model makes use of those repeated observations instead of collapsing all information about each bird into a single 0 or 1, as Dr. Shoemaker did to fit his Cormack-Jolly-Seber model. Given this, it is unclear how Dr. Shoemaker decided that a given bird belonged to the “Mispillion” or “Not Mispillion” group, given that many birds are seen both within and outside of Mispillion Harbor in a given year. The proportion of birds seen only in Mispillion ranges from 0.12 to 0.54 (0). The proportion of birds never seen in Mispillion ranges from 0.17 to 0.69. Given this variation and lack of systematic bias towards birds only being resighted in Mispillion Harbor, we do not believe there is reason to think that the large number of observations from this site result in biased inference.

> I do not think this is a major area of concern (which is why I included it as a supplement). My tests did not indicate a strong bias that was induced by the over-representation of this site in the resighting dataset. I do think it is worth noting, though, that the resighting data are so heavily dominated by this one site.
>
> The method I used to separate “Mispillion” birds from “non-Mispillion” birds was simply to filter the red knot resightings data frame to include or exclude all observations from this site. I performed this sub-setting operation before I collapsed within-year observations into zeros and ones- therefore, some birds were included in both analyses. I made it clear from the outset that I was happy to address any questions the ARM subcommittee had, but on this issue, as on others, no one from the ARM subcommittee reached out to ask such questions directly.

Literature cited

Atlantic States Marine Fisheries Commission (ASMFC). 2021. Revision to the Framework for Adaptive Management of Horseshoe Crab Harvest in the Delaware Bay Inclusive of Red Knot Conservation and Peer Review Report. Arlington, VA. 302 pp. http://www.asmfc.org/species/horseshoe-crab

Grace, J.B., Anderson, T.M., Olff, H. and Scheiner, S.M., 2010. On the specification of structural equation models for ecological systems. Ecological monographs, 80(1), pp.67-87.

McGowan, C.P., Smith, D.R., Sweka, J.A., Martin, J., Nichols, J.D., Wong, R., Lyons, J.E., Niles, L.J., Kalasz, K., Brust, J. and Klopfer, M., 2011. Multispecies modeling for adaptive management of horseshoe crabs and red knots in the Delaware Bay. Natural Resource Modeling, 24(1), pp.117-156.

McGowan, C.P., Smith, D.R., Nichols, J.D., Lyons, J.E., Sweka, J., Kalasz, K., Niles, L.J., Wong, R., Brust, J., Davis, M. and Spear, B., 2015. Implementation of a framework for multi-species, multi-objective adaptive management in Delaware Bay. Biological Conservation, 191, pp.759-769.

Nichols, J.D., Runge, M.C., Johnson, F.A. and Williams, B.K., 2007. Adaptive harvest management of North American waterfowl populations: a brief history and future prospects. Journal of Ornithology, 148, pp.343-349.

Niles, L.J., J. Bart, H.P. Sitters, A.D. Dey, K.E. Clark, P.W. Atkinson, A.J. Baker, K.A. Bennet, K. Kalasz, N.A. Clark, J. Clark, S. Gillings, A.S. Gates, P.M. González, D.E. Hernandez, C.D.T. Minton, R.I.G. Morrison, R.R. Porter, R.K. Ross, and C.R. Veitch. 2009. Effect of horseshoe crab harvest in Delaware Bay on red knots: are harvest restrictions working? BioScience 59:153-164.

O’Brien, S., B. Robert, and H. Tiandry. 2005. Consequences of violating the recapture duration assumption of mark-recapture models: A test using simulated and empirical data from an endangered tortoise population. Journal of Applied Ecology 42:1096–1104.

Platt, J.R., 1964. Strong Inference: Certain systematic methods of scientific thinking may produce much more rapid progress than others. science, 146(3642), pp.347-353.

Rakhimberdiev, E., Karagicheva, J., Saveliev, A., Loonstra, A. J., Verhoeven, M. A., Hooijmeijer, J. C., ... & Piersma, T. (2022). Misidentification errors in reencounters result in biased estimates of survival probability from CJS models: Evidence and a solution using the robust design. Methods in Ecology and Evolution, 13(5), 1106-1118.

Riecke, T.V., Williams, P.J., Behnke, T.L., Gibson, D., Leach, A.G., Sedinger, B.S., Street, P.A. and Sedinger, J.S., 2019. Integrated population models: model assumptions and inference. Methods in Ecology and Evolution, 10(7), pp.1072-1082.

Runge, M.C., 2011. An introduction to adaptive management for threatened and endangered species. Journal of Fish and Wildlife Management, 2(2), pp.220-233.

Schaub, M. and Kéry, M., 2021. Integrated population models: Theory and ecological applications with R and JAGS. Academic Press.

Smith, J.A.M., A. Dey, K. Williams, T. Diehl, S. Feigin, and L.J. Niles. 2022. Horseshoe crab egg availability for shorebirds in the Delaware Bay: dramatic reduction after unregulated horseshoe crab harvest and limited recovery after 20 years of management. Aquatic Conservation: Marine and Freshwater Ecosystems. 2022:1-13. DOI: 10.1002/aqc.3887.

Tucker, A.M., C.P. McGowan, R.A. Robinson, J.A. Clark, J.E. Lyons, A. Derose-Wilson, R. Feu, G.E. Austin, P.W. Atkinson, and N.A. Clark. 2019. Effects of individual misidentification on estimates of survival in long-term mark-resight studies. The Condor: Ornithological Applications 121:1–13.

Tucker, A.M., McGowan, C.P., Lyons, J.E., DeRose-Wilson, A. and Clark, N.A., 2022. Species-specific demographic and behavioral responses to food availability during migratory stopover. Population Ecology, 64(1), pp.19-34.

Tucker, A.M., McGowan, C.P., Nuse, B.L., Lyons, J.E., Moore, C.T., Smith, D.R., Sweka, J.A., Anstead, K.A., DeRose-Wilson, A. and Clark, N.A., 2023. Estimating recruitment rate and population dynamics at a migratory stopover site using an integrated population model. Ecosphere, 14(2), p.e4439.

U.S. Fish and Wildlife Service (USFWS), 2014. *Rufa* red knots: Background Information and Threat Assessment. Supplement to Endangered and Threatened Wildlife and Plants; Final Threatened Status for the Rufa red knots (*Calidris canutus* Rufa). U.S. Fish and Wildlife Service, Pleasantville, New Jersey, USA (2014) [Docket No. FWS-R5-ES-2013-0097; RIN AY17]

Williams, B.K., 2011. Adaptive management of natural resources—framework and issues. Journal of environmental management, 92(5), pp.1346-1353.

EXHIBIT B

Newstead, D.J. et al. 2024
Survival of red knots in the northern Gulf of Mexico
Front. Ecol. Evol. 12:1375412
doi: 10.3389/fevo.2024.1375412

Frontiers in Ecology and Evolution

TYPE Original Research
PUBLISHED 09 April 2024
DOI 10.3389/fevo.2024.1375412

Survival of red knots in the northern Gulf of Mexico

David J. Newstead[1]*, Stephen J. Dinsmore[2], Bart M. Ballard[3], Lawrence J. Niles[4] and Joanna Burger[5]

[1]Coastal Bend Bays & Estuaries Program, Corpus Christi, TX, United States, [2]Department of Natural Resource Ecology and Management, Iowa State University, Ames, IA, United States, [3]Caesar Kleberg Wildlife Research Institute, Texas A&M University-Kingsville, Kingsville, TX, United States, [4]Wildlife Restoration Partnerships, Greenwich, NJ, United States, [5]Department of Ecology, Evolution, and Natural Resources, Rutgers University, New Brunswick, NJ, United States

OPEN ACCESS

EDITED BY
Piotr Nowicki,
Jagiellonian University, Poland

REVIEWED BY
Joe Eisaguirre,
United States Geological Survey,
United States
Daniel Robert Ruthrauff,
U.S. Geological Survey, United States

*CORRESPONDENCE
David J. Newstead
✉ dnewstead@cbbep.org

RECEIVED 23 January 2024
ACCEPTED 18 March 2024
PUBLISHED 09 April 2024

CITATION
Newstead DJ, Dinsmore SJ, Ballard BM, Niles LJ and Burger J (2024) Survival of red knots in the northern Gulf of Mexico. *Front. Ecol. Evol.* 12:1375412. doi: 10.3389/fevo.2024.1375412

Highly migratory shorebirds are among the fastest declining avian guilds, so determining causes of mortality is critically important for their conservation. Most of these species depend on a specific geographic arrangement of suitable sites that reliably provide resources needed to fuel physiologically demanding life histories. Long-term mark-resight projects allow researchers to investigate specific potential sources of variation in demographic rates between populations. Red Knots (*Calidris canutus*) occur in three relatively distinct regions across the northern Gulf of Mexico, and two of these areas have been experiencing episodic harmful algal blooms (red tide) with increased frequency in recent decades. Since knots are mostly molluscivorous during the nonbreeding season in the Gulf, they are potentially exposed to red tide toxins at high concentrations via their filter-feeding prey. We used long-term mark-resight data from Texas, Louisiana, and Florida (USA) to estimate apparent survival, and to assess the effects of red tides on survival of Red Knots. We also assessed effects of tracking devices deployed in conjunction with the projects over the years. While overall apparent annual survival rates were similar across the three locations (0.768 – 0.819), several red tide events were associated with catastrophically low seasonal (fall) survival in Florida (as low as 0.492) and Texas (as low as 0.510). Leg-mounted geolocators, but not temporary glued-on VHF tags, were associated with a reduction in apparent survival (~8%/year). Movement of knots between the three areas was rare and site fidelity is known to be high. Harmful algal blooms are predicted to increase in frequency and severity with climate change and increased anthropogenic degradation of coastal habitats, which may further endanger these as well as other shorebird populations around the world.

KEYWORDS

survival, shorebird, harmful algal bloom, red tide, molluscivore, red knot, Gulf of Mexico, mark-resight

1 Introduction

Understanding demographic parameters is fundamental to monitoring and managing wildlife populations, but the highly migratory nature of many shorebird species makes estimation of these parameters distinctly challenging (Faaborg et al., 2010). Species may have broad or disjunct breeding ranges, geographically distinct nonbreeding populations, and rely differentially on migratory stopovers between the two. Being able to isolate parameters (and factors that may affect them) to specific populations requires knowledge of connectivity (Webster et al., 2002; Rushing et al., 2017), since consequences of factors affecting one part of the annual cycle can have carry-over effects on subsequent ones (Goss-Custard et al., 1995; Norris, 2005; Duijns et al., 2017). Survival rates of adults and post-fledged juveniles have been demonstrated to be the most consequential to population growth rates of several migratory shorebirds (Hitchcock and Gratto-Trevor, 1997; Calvert et al., 2006). For migratory shorebirds that use different geographic areas for discrete parts of their annual cycle, changes in habitat quality in any part of the cycle can have a strong effect on survival (Johnson et al., 2006; Duriez et al., 2012).

Coastal habitats worldwide have been degraded by human activities such as shoreline development, pollution, and freshwater diversions (Kennish, 2002), decreasing their capacity to support populations of migratory shorebirds (Fernández and Lank, 2006). Beyond direct losses, anthropogenic disturbance can be functionally equivalent to habitat loss or degradation by rendering sites unusable (Gill and Sutherland, 2000). Norris and Marra (2007) demonstrated that differences in habitat quality in one part of the annual cycle can have interseasonal effects on population dynamics depending on the strength of migratory connectivity. When connectivity is strong, further habitat loss from projected sea level rise is likely to result in bottlenecks with potential consequences to populations proportionately larger than the habitat loss itself (Iwamura et al., 2013).

Harmful algal blooms (HABs) occur in aquatic environments and can be considered extreme biological events resulting in major disruption to coastal ecosystems through complex food web dynamics (Landsberg et al., 2009). HABs have occurred in the Gulf of Mexico far back into recorded history (Magaña et al., 2003). They have increased in frequency and now occur commonly on the coasts of Texas/Mexico and western Florida (Hallegraeff, 1993, van Dolah 2000, Walsh et al., 2006; Brand and Compton, 2007; Tominack et al., 2020). Blooms in the Gulf of Mexico resulting in fish kills associated with the dinoflagellate *Karenia brevis* are typically known as "red tides." The organism produces brevetoxin, a very potent neurotoxin that kills fish through absorption across gill membranes (Abbott et al., 1975) or consumption of toxic biota (Tester et al., 2000). These toxins can accumulate and result in mortalities of higher vertebrates directly and indirectly through food web dynamics (Landsberg et al., 2009). Filter-feeding molluscs – especially bivalves – readily accumulate brevetoxins in high concentrations (Bricelj et al., 2012; Van Hemert et al., 2022) and occasionally experience direct lethal effects, as well as sublethal effects that result in subsequent recruitment failure (Summerson and Peterson, 1990). However, most mollusk species survive exposure to brevetoxins, accumulating high concentrations of toxins that can then be ingested by consumers (Landsberg, 2002). In addition to effects from direct consumption, brevetoxin from lysed cells can reach extremely high concentrations that can persist in waters and sediments for several weeks after the bloom organism has dissipated (Pierce and Henry, 2008; Castle et al., 2013), exposing probe-feeding shorebirds to additional dosages through passive uptake. Despite strong evidence correlating bird mortalities with HABs (Van Hemert et al., 2021, 2022), data from experimental studies or laboratory examination of tissue samples are relatively scarce (Shumway et al., 2003). Impacts are likely underestimated due to depredation and decomposition of carcasses, and removal of carcasses through tidal action (Sutherland et al., 2012). Further, a lack of long-term demographic monitoring of affected avian species has confounded determination of population level effects, though a recent study found a relationship between HAB occurrence and survival in Gulf-wintering Piping Plovers (Ellis et al., 2021).

The Red Knot (*Calidris canutus*) is a Holarctic breeding shorebird comprising six currently recognized subspecies. In the Western Hemisphere, the *C. c. rufa* subspecies spends nonbreeding seasons in the southern US and neighboring Mexico, especially the states bordering the Gulf of Mexico (henceforth, the "Gulf"), the Caribbean, and several regions in South America from northern Brazil to Tierra del Fuego (Niles et al., 2008). Additionally, some knots wintering on the Pacific coast of southern Mexico (Oaxaca) south to Chiloé Island, Chile occur in Texas and Louisiana during migration – primarily during spring – and consist of both *C. c. rufa* and *C. c. roselaari* (Newstead, unpubl. data). Though the total population of knots that do this is not known, it is suspected to be considerably less than those wintering in the Gulf. Knots in the Gulf are concentrated primarily in three general areas: southwestern Florida, the barrier islands of Louisiana, and the coast of south Texas and Tamaulipas. These Gulf states are among the highest latitude wintering sites (~24° – 29° N) of the *C.c. rufa* subspecies, used not only during the extensive nonbreeding season but also for pre-migratory and post-breeding stages. Observations of marked individuals (Tuma and Powell, 2021, Newstead, unpubl. data) confirm high site fidelity to each of these locations, consistent with studies on other subspecies (Harrington et al., 1998; Leyrer et al., 2006; Buchanan et al., 2012; Musmeci et al., 2022).

Geolocator studies (Newstead et al., 2013, Newstead, unpubl. data) show that the Texas and Louisiana populations migrate almost exclusively through the interior of the North American continent rather than using sites along the Atlantic coast. The decline of more than 75% of the Atlantic Flyway *rufa* population over the course of two decades (Niles et al., 2008) prompted its listing as Endangered in Canada in 2007 (COSEWIC, 2007) and as Threatened under the US Endangered Species Act in 2014 (USFWS, 2014a). Recognition and understanding of the Gulf populations – particularly the Texas and Louisiana populations – have been relatively recent discoveries, and there has been no previous estimation of survival parameters that can be compared across the three locations. The Red Knot is considered primarily a molluscivore during the non-breeding season (van Gils et al., 2006; Baker et al., 2013). The species' reliance on coquina clams

(*Donax* spp.) when using Gulf beaches makes it particularly vulnerable to HABs and they have been observed exhibiting symptoms of neurotoxic shellfish poisoning during red tide events (DN, personal observation). Carcasses of knots encountered freshly dead or dying were found to have exceptionally high levels of brevetoxin in all tissues tested, with the highest levels in the liver and gastrointestinal tract (Rafalski, 2012).

New tracking technologies continue to contribute major breakthroughs in our understanding of avian life histories (Bridge et al., 2010; Robinson et al., 2010; Wilmers et al., 2015). The use of archival light-level data loggers (geolocators), radiotransmitters, GPS and cellular technologies has drastically expanded our understanding of migratory strategies and revealed previously-unknown sites of essential importance (Stutchbury et al., 2009; Newstead et al., 2013; McKellar et al., 2015; Chan et al., 2019). While these discoveries have been critical in directing further research and conservation actions to places that can best benefit the species, the effects of tracking devices on the movements, activities, and, ultimately, survival of tracked animals remains a source of concern (Barron et al., 2010; Elliott et al., 2012; Scarpignato et al., 2016). Meta-analyses on device effects on birds (survival, behavior, reproductive success and others) have revealed some significant negative consequences varying by species, device type, attachment method, migration distance, and many other factors (Barron et al., 2010; Costantini and Møller, 2013). Specific to shorebirds, most studies have reported no significant impact of leg-mounted geolocators based on metrics from the year following deployment (Conklin and Battley, 2010; Pakanen et al., 2015; Mondain-Monval et al., 2020). Reductions in one-year return rates were detected for only two of 23 Arctic-breeding shorebird populations carrying geolocators relative to individuals carrying only a unique leg marker, with no detectable effect on the Great and Red knots included in the analysis (Weiser et al., 2016). However, Pakanen et al. (2020) found that when they extended their analysis of Dunlin (*C. alpina*) tracked over multiple years, apparent survival was lower for birds carrying geolocators compared to those without. These findings suggest that negative effects may accumulate over time or result in incremental increases in mortality risk. When possible, longer-term datasets should be analyzed to determine consequences that may not be evident based on one-year return rates alone. Small VHF transmitters attached to birds tracked using direct or automated radiotelemetry have also yielded important findings for many shorebirds, especially for local movements (Green et al., 2002; Warnock and Takekawa, 2003; Rogers et al., 2006; Duijns et al., 2019). Most VHF tag deployments on shorebirds have utilized an adhesive to affix the transmitter to the back, which subsequently falls off the bird with the next molt cycle or sooner, and these studies have generally reported no short-term survival consequences (Drake et al., 2001; Barron et al., 2010; Buchanan et al., 2019; Stantial et al., 2019).

Annual survival is a key underlying demographic parameter that can vary with environmental conditions, and strongly influences population trends. When data are sufficient, annual survival can be apportioned into partial (e.g. seasonal or semi-annual) components, providing greater insight into what particular locations or processes are contributing to demographic change (Gauthier et al., 2001; Leyrer et al., 2013; Piersma et al., 2016; van Irsel et al., 2022). We used mark-resight data from three projects involving captures of Red Knots in the three main Gulf of Mexico wintering areas to compare annual (and seasonal when possible) survival rates between populations, and to assess effects of an increasingly prevalent coastal ecosystem stressor (HABs) and the use of tracking devices on survival.

2 Methods

2.1 Study area

The northern Gulf of Mexico is bordered by a fairly contiguous extent of sandy beaches punctuated by passes connecting to bays and other receiving waters. Sediment grain size composition and origin (biogenic and terrigenous) vary widely across the Gulf, which affects the character of benthic infaunal communities and consequently the distribution of shorebirds that use them. Red Knots occur regularly in three primary areas across the Gulf – Florida, where they are most concentrated in the southwestern region between Clearwater and Marco Island; Louisiana, where they occur on the beaches of Grand Isle and the adjacent Caminada Headlands as well as the offshore barrier islands of the Breton Island National Wildlife Refuge; and Texas, where they are most common on the southern half of the coast from the Corpus Christi area to the border with Mexico, and likely well into contiguous parts of Tamaulipas where habitat is very similar. These three main areas are at least 600 km from one another and are considered as separate population units for the purpose of recovery planning (USFWS, 2021). These three geopolitical states are henceforth referred to as “locations” to avoid potential confusion with conditional states related to the analysis.

2.2 Field methods

For this project, captures of Red Knots occurred in Texas on Mustang and North Padre Islands between October 2009 – October 2019, in Louisiana on Grand Isle and the Caminada Headlands from the eastern end of Elmer’s Island west to Port Fourchon between April 2014 – April 2019; and in Florida from Longboat Key to Sanibel Island between October 2005 – March 2010.

All Red Knots were captured using a cannon-net (~ 9 m X 9 m, or ~10 m X 25 m) on beaches where birds were foraging or resting. Standard processing included a federal metal band on tarsus or tibia, a uniquely inscribed alphanumeric green flag on the opposite tibia, measures of bill and total head length (nearest 0.1mm), flattened wing chord length (mm), and mass (grams). A clip of the distal portion of the 6th primary covert was retained from most captured birds for isotopic analysis (carbon, nitrogen, hydrogen isotopes; for a project to assign migrants to wintering sites), and a blood sample was taken by brachial venipuncture on a smaller sample of birds for future genetic analysis.

In Florida, capture effort was concentrated between November – March (>95% of all captures) between years 2005-2010. Capture effort in Texas was mostly focused on fall and spring periods (>90% of all captures between September-November, or April-May) with smaller catches in other months, between 2009-2019. Louisiana captures were all in April, from 2014-2019. The distribution of resights by month was similar to that of the captures, except for Florida when many resights were recorded in months before and after the main winter months which constituted the bulk of the capture efforts.

Multiple tracking projects were conducted during the course of the projects. Archival light-level dataloggers (henceforth, "geolocators"; British Antarctic Survey [BAS] Model MK10 and MK12 or Migrate Technologies Intigeo W65) were mounted on leg flags and attached to the tibiotarsus as described in Niles et al. (2010). All assemblies weighed < 1.4 g. Radiotelemetry studies in Louisiana and Texas included deployment of small VHF transmitters (Lotek NTQB-4-2, 0.9 g) glued to the intrascapular region, as described in Newstead (2014).

2.3 Encounter histories and covariates

Encounter data were compiled from multiple resight projects and public domain records in bandedbirds.org; additional records were made available directly to the author. Encounter occasions began with the first capture effort in Florida in winter 2005/6 and ended in winter 2019/20 season.

Only records from Florida, Louisiana and Texas were used to build encounter histories. Birds were assigned to one of the three locations based on their original capture location. If an individual was encountered outside the location of initial capture (i.e., in one of the other two locations) and there were no subsequent records within the capture location it was removed from the dataset. This eliminated only a small number of birds from the dataset that may have switched wintering location or underwent atypical migrations.

Resightings were divided into three encounter occasions per year: the fall encounter (July 20 – October 31; 104 d; midpoint September 9), winter encounter (December 15 – January 31; 48 d; midpoint January 7), and spring encounter (April 1 – May 30; 60 d; midpoint May 1; Figure 1). Based on the midpoints of the encounter occasions, the year was thus divided into three intervals: (fall to winter – 120 d; winter to spring – 114 d; spring to fall – 131 d). These are referred to as the fall, winter, and summer intervals, respectively. The time range from the earliest captures to the most recent encounters spans 43 occasions (42 intervals).

As defined, the intervals generally reflect distinct and important phases in the annual cycle: during "fall" birds are returning from the Arctic and undergoing a body molt including flight feathers; during "winter" birds are managing a balance of predation risk, prey resource availability and maintaining sufficient fat reserves; during "summer" adult birds undertake a major migratory journey to Arctic breeding grounds, spend two to three months attempting to breed, and then return to nonbreeding areas. While juvenile birds nearly all remain on nonbreeding areas in their first full summer, they are exposed to factors such as extreme heat and increased human disturbance that adults mostly escape.

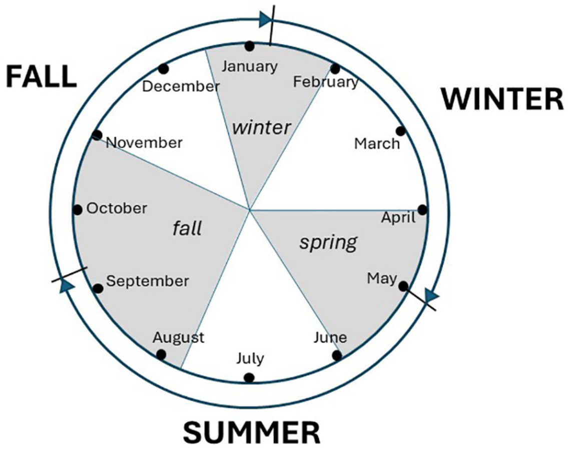


FIGURE 1
Encounter periods (grey), occasion midpoints (lines extending from outer circle), and season interval names for apparent survival analysis on Red Knots in the northern Gulf of Mexico.

Individuals were grouped into one of three age classes based on age at capture. Birds that were not aged upon capture were classified as "unknown" age. Birds classified as hatch-year prior to, or second-year during, a spring occasion were classified as juvenile. Birds aged as second-year or after-hatch-year following a spring occasion (i.e., they had survived the first full oversummer interval so were > 1 year old), and all birds aged as after-second-year were classified as adult. Juveniles and birds of unknown/unspecified age were assumed to recruit into the adult age class following the first summer interval.

Since occasions are assumed to be instantaneous, the initial occasion for birds captured during intervals was assigned to be the subsequent occasion, so that estimates would not be biased by partial interval effects.

We included covariates in the dataset to test whether negative effects of tracking devices resulted in lower apparent survival. Effects of leg-mounted geolocators and glue-on VHF transmitters were assessed using a set of time-varying binary covariates for each. Once deployed, an individual with a geolocator was assumed to retain the geolocator permanently unless it was removed. VHF transmitters glued to the intrascapular region typically fall off within a few months of deployment, so the covariate was applied for only the subsequent interval.

Since HABs (especially "red tides") have been observed to result in direct mortality to Red Knots, we hypothesized that exposure to toxins could result in lower apparent survival either through additional (undetected) direct mortality or sublethal effects. Effects of red tide were assessed using several approaches. Red tide sampling occurs in Florida (inshore and offshore) with good spatial and temporal coverage through the HABSOS system (NOAA National Centers for Environmental Information, 2014). The monthly bloom severity index (BSI) developed by Stumpf et al. (2022) was used to identify intervals when red tide blooms were affecting the southwest Florida coast. Red tide effects can occur at relatively low concentrations, but generally begin having pronounced effects resulting in fish kills at concentrations

>1,000,000 cells/L. The summed BSIs for months corresponding to intervals in this study were used to classify red tide as absent/minimal (summed BSI = 0, covariate = -1), moderate (summed BSI > 0 but < 5, covariate = 0), or severe (summed BSI > 5, covariate = 1). In Texas, red tide monitoring is conducted mostly in response to known or suspected occurrences. Since events vary greatly in their range and extent of impact to marine life, fish kill reports were used as a secondary source to confirm an event to a degree that would have resulted in high likelihood of the shoreline being exposed to the effects of the bloom. For Texas, red tide events were ascribed to seasons based on Tominack et al. (2020), and severity was assigned as appropriate to the geography utilized by knots. A covariate set was thus created for each location based on red tide being absent/minimal, moderate, or severe (-1, 0, and 1, respectively) during each interval. A covariate set including all red tide events was made for each location separately, and another that included all locations together (but the red tide covariates applicable to each location separately).

Because knots are highly mobile and likely vary in their degree of exposure to harmful algal blooms depending on various environmental factors, we also tested the effect of each individual red tide season against all others. Separate covariate sets were created for each red tide season occurrence in Florida and Texas to assess the effect of red tide events independently. The covariate value of 1 was assigned to intervals when red tide was present (either moderate or severe), and 0 for all others. Based on our criteria, there were a total of 17 and 6 red tide season events for Florida and Texas, respectively, applicable to the 42 intervals of the study, so a covariate set was created for each of these.

2.4 Statistical analyses

Models were evaluated using a Cormack-Jolly-Seber (CJS) framework in Program MARK (v. 9.0, White and Burnham, 1999) to estimate apparent survival (φ) and encounter (*p*) probabilities. Apparent survival is the probability that a knot alive at occasion *i* was alive and in the study area at occasion *i + 1*. Its inverse includes mortality and permanent emigration from the study area. Goodness-of-fit testing was run on the fully time-varying model and contingency tables were examined individually to assess whether patterns indicated lack of independence in the data. The median $\hat{c}$ approach was applied to account for overdispersion in all subsequent models. Model evaluation was based on quasi-Akaike's Information Criterion adjusted for sample size ($QAIC_c$) and model weights (ω_i). We built models in an ordered 3-step process described below.

2.4.1 Step 1: determining best underlying model structures

Preliminary evaluation of the dataset indicated major differences in the distribution of encounters between locations and seasons, so model fitting began with a series of models holding φ constant by location and allowing for variation in p by location, season, and age. Using the best parameter structure for p, models incorporating variability in φ by location, season and age (and combinations thereof) were then tested to determine the best fit for a base model. Models in which covariate parameters were poorly estimated (standard errors of effect coefficient very close to zero or greater than 2.0) were removed from the resulting model set. Models within 2 ΔQAICc of the top model were considered well-supported, and the top model was carried forward for testing of the time-varying covariate datasets.

2.4.2 Step 2: building a candidate model set with red tide index and tracking device effects

We then built a candidate set of models that included covariates added to the most competitive base model. We considered the effect of tracking devices (geolocators, VHF transmitters) independently as well as combined. Given differences in habitat distribution and the character and duration of red tide events between Texas and Florida, we considered the effect of red tide on each location modeled independently, as well as together. We then considered models that included both tracking device and red tide effects. Covariates were considered predictive if the 95% confidence intervals (C.I.) of effect coefficients did not include zero. Apparent survival and encounter probabilities were reported based on the top model that did not include a red tide effect. If all parameters were well estimated in a model including seasonal variation within a location, the model including those terms and the tracking device effects was used to estimate those season-specific parameters (i.e. to provide estimates unaffected by tracking devices). To facilitate comparison with other studies, apparent seasonal survival (φ_a) estimates and 95% C.I.s were converted to apparent annual estimates using the delta method (Powell, 2007), either as a product of the three separate seasonal estimates or exponentiation of the non-season specific estimates.

2.4.3 Step 3: evaluating survival in specific red tide seasons

To evaluate the effect of specific red tide events, we used the most competitive base model and independently added each red tide season to the model as applicable to each location. We considered a red tide event to be poorly estimated if its inclusion resulted in other parameters being poorly estimated. Red tide events (seasons) were considered significant if the 95% C.I.s of the effect coefficient did not overlap zero. For significant seasons, the magnitude of the effect on φ was calculated as the percentage difference between the mean estimate of the survival probability in that season relative to the survival probability of all other seasons for that location.

To estimate survival for each significant red tide season, we ran a *post-hoc* model treating each of those seasons individually and accounted for any significant tracking device effects. If any coefficient became non-significant in this model, that covariate was removed and the reduced model run until all terms were significant.

The strength of differences between locations was assessed by whether 95% C.I.s overlapped, and covariate effects were assessed by whether the 95% C.I. included zero. C.I.s are presented in brackets following the mean, unless otherwise noted.

3 Results

Encounter histories were constructed from 2,412 knots (Florida: 1,373 captured between 2005-2010; Louisiana: 255 captured between 2014-2019; Texas: 784 captured between 2009-2019), and 4,078 resights (Florida: 3,013; Louisiana: 188, Texas, 877; Supplementary Table 1). Geolocators were deployed on 68, 49, and 114 knots in Florida, Louisiana and Texas, respectively. VHF transmitters were deployed on 18 and 115 knots in Louisiana and Texas, respectively.

There were 17 red tide seasons in Florida (8 severe, 9 moderate) during the 42 intervals since marking began. Two were in summer (one moderate, one severe). Both summer events preceded severe fall events. Of nine fall events (three moderate, six severe), six persisted into the subsequent winter interval. There were no winter events that were not preceded by a fall red tide event. In Texas, there were 6 red tide seasons (3 severe, 3 moderate) during the 31 intervals since marking began. All Texas red tide seasons were in fall.

3.1 Best underlying model structures

The goodness-of-fit test indicated some overdispersion in the data but examination of contingency tables did not suggest any systematic source of bias. Differences in resighting effort (p) between years and locations were likely responsible for high model deviance. Subsequently, all models were adjusted using median $\hat{c}$ = 1.155. The best models for the encounter parameters included location and season. All models including age resulted in multiple parameters being poorly estimated, so these were removed from further consideration. All subsequent model runs utilized the $p_{(\text{location, season})}$ parameterization.

The top base model for explaining variation in Red Knot apparent survival included a constant seasonal survival term (φ_{c}) for each location. A competing model allowed for season-specific (φ_{f}, φ_{w}, φ_{s}) parameters for Florida, but not for Texas and Louisiana. A model with constant seasonal survival across locations received the lowest model weight of the three. The two most competitive models were carried forward for model development incorporating HAB and tracking device covariates.

3.2 Assessment of candidate models including red tide index and tracking device effects

All models testing tracking device and red tide effects on the base model that included seasonal variation in survival in Florida had uniformly higher QAICc than the corresponding models based on the constant seasonal survival base model. Since the inclusion of variation in seasonal survival in Florida did not improve model fit in any case, these models were removed from the candidate model set.

The best fitting model included effects of geolocators and red tide in Florida (Table 1). The four top models each had a likelihood >0.125 (indicating support; Burnham and Anderson, 2002), and all included the geolocator covariate. The geolocator effect was negative and significant in all models that included it. VHF transmitter and red tide covariates were also all negative but non-significant when included in the models. Multiple parameters were poorly estimated in all models that included red tide in Texas only. The effect of geolocator in the top-ranked model without a red tide effect ($\hat{\beta}$ = -0.445 [-0.655, -0.236]) equates to an estimated reduction in seasonal apparent survival of 4.1%, 3.2%, and 3.8% for Texas, Louisiana, and Florida, respectively.

The top-ranked model that did not include a red tide effect was used to estimate apparent survival for each location. With tracking devices accounted for separately in the model, mean apparent seasonal survival was highest for Louisiana, intermediate in Florida, and lowest in Texas, though C.I.s overlapped (Table 2). Resighting probabilities varied between seasons within each location.

TABLE 1 Model ranking including combinations of red tide and tracking device covariates applied to the best-fitting base model (Φ_{location}, $p_{\text{location, season}}$) for Red Knots from Texas, Louisiana, and Florida populations from 2005-2019.

Model	Red tide	Tracking device	Δ QAIC$_c$	ω_i	Likelihood	K	QDeviance
1	Florida	geo	0.00	0.37	1.00	14	21295.7
2[a]	–	geo	0.33	0.32	0.85	13	21298.1
3	–	geo, VHF	1.92	0.14	0.38	14	21297.6
4	All	geo	2.33	0.12	0.31	14	21298.1
5	All	geo, VHF	3.92	0.05	0.14	15	21297.6
6	Florida	–	13.47	0.00	0.00	13	21311.2
7[b]	–	–	14.28	0.00	0.00	12	21314.0
8	–	VHF	16.14	0.00	0.00	13	21313.9
9	All	–	16.19	0.00	0.00	13	21313.9
10	All	VHF	18.03	0.00	0.00	14	21313.8

[a] Top-ranked model not including a red tide effect, on which reported seasonal survival estimates and geolocator effects are based.

[b] Base model (no covariates) from Step 1 on which subsequent model development was based.

Estimation of distinct seasonal apparent survival probabilities was only possible for Florida. When seasonal variation for Florida was added to the top-ranked model, mean apparent survival was highest during winter (0.944 [0.915, 0.963], intermediate in fall (0.914 [0.834, 0.957] and lowest in summer (0.907 [0.821, 0.954]), though C.I.s were wide and overlapping.

3.3 Individual red tide season effects

Parameters were estimable for models including individual red tide seasons on the base model for one (of six) Texas seasons, and nine (of seventeen) Florida seasons (Table 3). The 2009 fall red tide season in Texas was significant ($\hat{\beta}$ = -2.515 [-3.291, -1.739]), as were four total seasons in Florida comprising two extended events in 2012 (fall: ($\hat{\beta}$ = -1.553 [-1.742, -0.764]; winter: ($\hat{\beta}$ = -1.470 [-1.930, -1.010]) and 2018 (fall: ($\hat{\beta}$ = -2.504 [-3.169, -1.840]; winter: ($\hat{\beta}$ = -1.831 [-2.817, -0.845]). Red tide seasons with non-significant terms had higher standard errors, indicating data was insufficient to estimate an effect.

The *post-hoc* model retaining all significant covariates included the geolocator effect and four of the five significant red tide seasons (Table 4). Point estimates of seasonal survival during red tide events in Florida ranged from 0.492 (fall 2018) to 0.884 (fall 2012). Seasonal survival during the Texas fall 2009 red tide was 0.510.

4 Discussion

Our results confirm episodes of sharply reduced survival of Red Knots during red tide events, and suggest this could be a significant driver of survival in Texas and Florida. While only a red tide effect in Florida was included in the top model of the candidate set, tests on individual seasons – when all parameters were estimable – were all either strong and significant, or were weak with relatively high standard errors. This is indicative of sparseness of data in some seasons (especially low winter resight probability in Texas) which likely resulted in a failure to find an effect when one may have occurred. Instead of chronically lower annual survival, knots in these locations may be experiencing relatively high survival punctuated by acute episodes of high mortality from red tide.

Several studies on knots have demonstrated often sharply contrasting survival estimates comparing different time series (Baker et al., 2004; González et al., 2006; Leyrer et al., 2013), population segments (Harrington et al., 1998) and body condition (McGowan et al., 2011), and age (Schwarzer et al., 2012). A robust model accounting for transience, temporary emigration, persistence and food availability at a stopover site illustrated that many different processes can affect estimates of apparent survival over short timeframes (Tucker et al., 2021). Further, the focal populations of these studies often preclude simple comparison of survival estimates across studies. For example, knots captured in Delaware Bay during spring migration are primarily breeding age individuals who have already survived nearly two full years during which mortality is expected to be highest (and thus unaccounted for in estimates), whereas estimates based on populations that included those younger cohorts (including ours) would be expected to be lower. Nevertheless, our estimates of apparent annual survival rates of Red Knots from the three Gulf of Mexico locations were within the ranges of those reported by most other studies on *rufa* Red Knots. Of the three Gulf locations, mean apparent annual survival was lowest in Texas and highest in Louisiana, though differences were not significant.

An effect of age on survival was not detectable in our models, but we note that the first occasion a knot becomes "available" to our study sites follows a critical and typically very high-mortality time interval following hatching in the Arctic, including surviving to fledging and the first southbound migration (~first 3 months of life). However, we are aware of no published survival estimates for this species which include that highly sensitive period. Accurate estimation of age-specific survival in the first- and second-year periods (prior to the first return to the Arctic as a breeder for most knots) was likely related to limitations in data for these age groups.

Our study estimated apparent survival, which is the complement of both mortality *and* permanent emigration. These are the first published survival estimates for knots in Texas and Louisiana, but a relatively recent study examined true survival in

TABLE 2 Mean estimates and standard errors (SE) for apparent seasonal and annual survival and encounter probabilities of Red Knots for each location from the $\varphi_{(location,\ geolocator)}\ p_{(location,\ season)}$ base model.

Location	Φ seasonal	Φ annual	Encounter (*p*)	
Texas	0.916 (0.005)	0.768 (0.012)	spring	0.180 (0.011)
			fall	0.264 (0.012)
			winter	0.009 (0.002)
Louisiana	0.936 (0.013)	0.819 (0.033)	spring	0.331 (0.036)
			fall	0.021 (0.006)
			winter	0.071 (0.013)
Florida	0.925 (0.002)	0.790 (0.006)	spring	0.118 (0.005)
			fall	0.271 (0.007)
			winter	0.194 (0.006)

TABLE 3 Effect coefficients ($\hat{\beta}$) and 95% confidence intervals for covariates tested individually on the $\Phi_{(location)}\ p_{(location,\ season)}$ base model for Red Knot apparent survival in the northern Gulf of Mexico.

Covariates		$\hat{\beta}$ [95% C.I.]
Tracking devices		
Geolocator		**-0.445 [-0.655, -0.236]**
VHF		-0.312 [-1.727, 1.103]
Red tide		
Red tide - all		-0.040 [-0.279, 0.199]
Red tide - Florida		-0.203 [-0.422, 0.015]
Individual red tide seasons		
Texas		
2009	**fall**	**-2.515 [-3.291, -1.739]**
2012	fall	0.113 [-1.678, 1.903]
Florida		
2006	fall	0.196 [-1.540, 1.933]
	winter	-0.361 [-0.774, 1.495]
2009	fall	-0.079 [-0.882, 0.723]
2012	**fall**	**-1.253 [-1.742, -0.764]**
	winter	**-1.470 [-1.930, -1.010]**
2015	fall	-0.246 [-1.659, 2.151]
	winter	-1.472 [-4.847, 7.792]
2016	fall	0.386 [-2.094, 2.866]
2018	**fall**	**-2.504 [-3.169, -1.840]**
	winter	**-1.831 [-2.817, -0.845]**

Significant covariates and terms are in bold. Effects could not be estimated for the covariate set "Red tide – Texas" and several individual red tide seasons (Florida – summer 2006, fall and winter 2011, fall and winter 2017, summer 2018; and Texas – fall 2011, fall 2015, fall 2016, fall 2018).

Florida. Between 2005-2010, true annual survival of Florida-wintering knots was estimated at 0.89 for adults and 0.95 for juveniles, using a Barker model (Schwarzer et al., 2012). The Barker model accounts for emigration and re-immigration based on encounters in a secondary encounter area (in this case, James Bay, Ontario, and the US Atlantic coast), resulting in annual survival estimates that separate the two processes by which an individual can leave the population (mortality or permanent emigration). Our dataset encompasses the same individuals and years of the Schwarzer et al. (2012) study, but because of the use of different modeling approaches and longer timespan of our study, we would not expect our estimates to be consistent. However, comparison may provide some insight into the potential population dynamics of the Florida winterers. We explore two potential explanations, which are not mutually exclusive: 1) during the course of the past decade the survival rate has in fact declined since the Schwarzer et al. (2012) study; and, 2) more knots formerly associated with Florida wintering areas are spending extended periods of time or the full nonbreeding period at sites along the southeast US coast, or into the Caribbean.

The significant reduction in survival associated with several red tide events in Florida provides some support for the hypothesis that mean survival rates truly have declined particularly in the past decade. It must be noted that because there were no new birds marked in Florida beyond 2010 in this analysis, it is possible that an age-related effect (i.e. senescence) could have depressed our apparent survival rates. However, the five-year timespan of the Schwarzer et al. (2012) study encompassed only four seasons (two events) that met our criteria as moderate or severe in terms of BSI. Three of these were the contiguous summer-fall-winter seasons during the bloom of 2006-7 (two of those were moderate severity), and the other was the brief and moderate bloom of fall 2009. By contrast, red tide occurred in thirteen seasons over the subsequent decade. Each bloom affected multiple consecutive seasons (including the one beginning in fall 2017 that lasted well over a year and a half), potentially compounding the effects. The years assessed in the Schwarzer et al. (2012) study (the same as the first five years of ours) represent a relative lull in red tide frequency and severity in Florida compared to the latter decade included in our study.

There is also evidence that our apparent survival estimates for Florida could be lower because of permanent shifts in wintering range outside of Florida. Lyons et al. (2018) estimated the wintering population of the southeast US (including Florida) at 10,400 individuals using data from the fall migration in 2011, while surveyors conducting the International Piping Plover Census (Elliott-Smith et al., 2015) counted 5,069 Red Knots during the 2006 count and approximately 3,900 in 2011. These numbers are not directly comparable, as they are based on different methodologies, but they reflect uncertainties as to *where* specifically Red Knots are wintering in the southeastern U.S. While there are not consistent repeated estimates from each location within this region over that time, resight data indicates

TABLE 4 Seasonal apparent survival estimates of Red Knots in each location based on the highest-supported *post-hoc* model incorporating five significant covariates – geolocators, and the four red tide events as applicable to the affected location.

Location	Intercept	Geolocator[a]	Red tide event			
			Fall 2009	Fall 2012	Winter 2012	Fall 2018
Texas	0.918	0.884	0.510	–	–	–
Louisiana	0.935	0.908	–	–	–	–
Florida	0.932	0.902	–	0.884	0.786	0.492

[a]The geolocator effect is assumed the same across locations. A model with a geolocator effect varying by location had less support.

that some birds have indeed shifted from the Florida wintering group to the Atlantic coasts of Georgia and South Carolina (USFWS 2014b, Pelton et al., 2022). The parameter estimates for fidelity and re-immigration based on the Barker model used by Schwarzer et al. (2012) indicate some support for this hypothesis. The apparent survival estimates for Florida in this study confound permanent emigration (such as a shift in wintering area from Florida to Georgia/South Carolina) with mortality, so it is possible that some portion of the decrease in apparent survival was attributable to emigration.

Apparent survival estimates for the Texas and Louisiana populations from this study could also be biased low (relative to true survival), if some proportion of those birds had also shifted to other wintering sites. However, there is currently no solid evidence to support this, and relatively minimal exchange of individuals even between the locations suggests it is unlikely.

The four significant red tide seasons in Florida were actually two prolonged events that lasted through the fall and winter intervals of the 2012 and 2018 nonbreeding season, compounding the effect on annual survival. In those years, estimated annual survival (assuming mean of non-red-tide survival for the unaffected season) would have been ~0.56 (in 2012) and ~0.33 (in 2018). While the 2009 red tide in Texas primarily affected one season (fall), it was severe enough that annual survival would have been ~0.43. These estimates indicate the loss of large proportions (~44 – 67%) of the entire population in a single year. Though there is no fixed quantitative threshold of a "catastrophe" in population dynamics, certainly the scale of these losses for a *K*-selected species are alarming. Simulation studies have demonstrated that population trends tend to be depressed when *variability* in survival is high, relative to a population where it is low, given the same arithmetic mean of survival (Boyce, 1977; Hitchcock and Gratto-Trevor, 1997). Indeed, catastrophic events, especially when combined with other environmental stressors, can drastically accelerate negative population growth rates towards extinction in closed populations (Simberloff, 1988). In this case, the effect of catastrophes on one wintering population may be tempered somewhat depending on the degree of migratory connectivity between breeding and wintering areas. As the processes by which young Red Knots recruit into a particular wintering population remain poorly understood, it is not clear that high recruitment could offset low survival years to stabilize a wintering population over the long term. Population declines documented in other wintering areas for *C. c. rufa* suggest a negative long-run population growth rate, and our results indicate red tides could be contributing to very high variability in Red Knot survival, at least in the Texas and Florida populations. Under these conditions, populations become more vulnerable to extinction especially when the frequency and magnitude of random catastrophes are increasing (Lande, 1993).

Sparse data (low encounter probability) for certain seasons in some locations likely resulted in the inability to fully estimate parameters for multiple red tide events, but is it possible that birds are able to avoid red tide effects in some years, but not in others? Knots could potentially reduce their exposure to toxins either through a shift in prey selection, or a shift in range.

There is evidence that some shorebirds avoid prey with high concentrations of algal toxins. Black oystercatchers (*Haematopus bachmani*) shifted diet to prey items that did not harbor algal toxins when those toxins were present in sea mussels – their preferred prey – and discarded mussel tissue with high toxin concentrations when they did capture it (Kvitek and Bretz, 2005), while other shorebird species tended to avoid areas where toxins were present. Red knots, however, consume bivalve prey whole and crush it in their gizzard rather than removing the flesh first (which would provide an opportunity to taste and reject), potentially making them more susceptible to accumulate high amounts of toxin. A prey selection mechanism to reduce exposure would only be viable if a suitable non-toxic alternate prey source were available. On the Gulf-facing beaches, *Donax* spp. is by far the dominant bivalve mollusk that is most likely to occur in ample densities to support knots, and it is known to concentrate HAB toxins at extremely high levels (Cummins et al., 1971). It is also possible that red tides could affect birds by negatively affecting recruitment of their bivalve prey (Summerson and Peterson, 1990; Rolton et al., 2016), which might have both immediate and long-term effects. A study comparing two red tide outbreaks (2006 and 2011) on beaches of south Texas found that one event resulted in a near complete die-off of the benthic macrofauna while that same faunal community was virtually unaffected in the other event, despite extensive fish-kills occurring in both (Lerma, 2013).

As discussed previously, permanent emigration of birds from the Florida wintering population to another site in the southeast US would be one way to avoid red tide effects. However, avoidance may not require permanent emigration. Since red tides most commonly occur during fall months, simply prolonging a southeast US stopover before moving on to Florida could reduce the degree of exposure. The abundance and duration of knots stopping at the Altamaha River delta (Georgia) varies between years and is likely influenced by availability of the dwarf surf clam (*Mulinia lateralis*; Lyons et al., 2018), so "good years" at this site might reduce the proportion of birds arriving in southwest Florida to toxic conditions, at a time when they are already under high physiological stress due to the demands of molt which is coupled with decreased immunological function (Buehler et al., 2008). If knots stay in the southeast US long enough to complete their molt, they would also likely arrive in better condition. There is isotopic evidence that some knots in the Florida wintering population do in fact complete their molt prior to arrival in Florida (Newstead, unpubl. data). Staying longer further north would also reduce the risk of exposure to tropical storms during the peak of hurricane season (Niles et al., 2012).

In Texas, knots are known to utilize the extensive tidal flats of the Laguna Madre when water levels allow (Newstead, 2014), and when red tides do occur, they tend to be most severe and extensive on the Gulf beach, only occasionally affecting the Laguna Madre. Also, the Laguna Madre complex and the interspersed flats of the Rio Grande Delta extend over 400 km from Corpus Christi, Texas southward to La Pesca, Tamaulipas, Mexico. Aerial radiotelemetry documented that knots move extensively throughout this system during the nonbreeding season (Newstead, 2014), so they could potentially avoid red tide effects by moving to unaffected parts of the same extensive system.

Red tides typically occur beginning in late summer and often persist until early to mid-winter, though in the past decade some events have been initiated or prolonged into the spring and summer seasons (Brand and Compton, 2007; Stumpf et al., 2022). Comparing models allowing seasonal variation in survival for Florida, estimates were lower in all seasons when red tide was not included as a covariate, but within all models season-specific estimates were lower in summer relative to fall and winter. This suggests that, absent red tide, survival in Florida during the extensive nonbreeding period is higher relative to the breeding period, which includes lengthy round-trip migrations for breeding adults. This finding is in contrast to Leyrer et al. (2013) for *C.c. canutus* wintering at Banc d'Arguin in Mauritania, where survival during the migratory and breeding seasons was close to 1.0, with most mortality occurring on the wintering area. Banc d'Arguin, at roughly 20.5 N latitude, is extremely arid and hot even during the boreal winter. Leyrer et al. (2013) suggested that during the period following arrival from breeding grounds, environmental and interspecific competitive constraints may depress survival at a time when birds are already under high physiological stress due to flight feather molt (Leyrer et al., 2013). Additionally, during this phase knots tend to suppress costly immune functions which may make them more vulnerable to novel stressors (Buehler et al., 2008). Climate conditions on wintering sites are more moderate in the subtropical latitudes of this study, though birds may occasionally experience stress from short bouts of cold winter temperatures in addition to a wider array of other stressors such as disturbance from heavy recreational use of beaches. Such conditions could simultaneously increase maintenance metabolism costs and place constraints on foraging opportunity. Prey depletion, or prey toxicity, from red tide events during this time period would introduce another lethal or sublethal stressor on top of those already normally experienced by knots during the nonbreeding period.

Boyd and Piersma (2001) found that relative population stability of Red Knots (*C.c. islandica*) wintering in Great Britain was maintained by alternating trends of survival and recruitment, implicating a potential role of density-dependent processes in population regulation. Knots using Delaware Bay during spring migration experienced consistently high apparent survival which was offset by consistently low recruitment between 2005-2018, resulting in a slightly positive population growth rate (Tucker et al., 2023). Using data from two large shorebird monitoring datasets, Bart et al. (2007) suggested the most likely mechanisms of North American shorebird population declines are reduction in breeding population size and poor reproduction, rather than an artifact potentially explicable by shifting distributions. This is almost certainly the case with Red Knots, as nearly all regular monitoring at key sites across the range indicate a declining trend, while no "new" sites of importance have been discovered in the meantime that balance for losses seen elsewhere. The relatively acute mortality episodes associated with red tides in this study would clearly result in reduced breeding population, but it is not known whether reproductive capacity can offset such population reductions when they occur relatively frequently.

While red tide toxins have been directly tied to the mortality of Red Knots in Texas (Rafalski, 2012) and closely related shorebirds in Florida (van Deventer et al., 2012) through necropsy and tissue sampling, only one other study has quantitatively estimated the effect of HABs on shorebird survival at the population level. Ellis et al. (2021) detected a negative effect of HABs on Piping Plover (*Charadrius melodus*) survival during the nonbreeding season along the Gulf of Mexico coast. This species is not only faithful to wintering areas generally (similar to knots) but even more highly faithful to specific individual territories with small home ranges (Drake et al., 2001; Cohen et al., 2008; Newstead, 2014) and may have a greater disinclination to move away from an area affected by red tide or other factors that may negatively affect survival. Our study provides additional evidence that HABs can negatively impact shorebird populations even when sudden mass mortality events are not observed or perhaps do not occur.

Another HAB dinoflagellate, *Aureoumbra lagunensis*, creates "brown tides" in the Laguna Madre of Texas which could be affecting knots in other ways. Though this organism does not produce potent toxins, it is considered disruptive to ecosystems because of its ability to bloom at low light and nutrient levels, and create a positive feedback mechanism that results in losses to seagrasses and benthic organisms (Gobler and Sunda, 2012). One brown tide event in the 1990s persisted in the Laguna Madre for nearly eight years, the longest HAB ever recorded (Buskey et al., 2001), and blooms have recurred intermittently and at varying spatial extents since then (DeYoe et al., 2007). Major die-offs of *Mulinia lateralis*, formerly the dominant bivalve mollusk in the Laguna Madre, have been coincident with these blooms (Montagna et al., 1993). The diet of Red Knots during the winter months in the Laguna Madre has not been described, but given that *M. lateralis* is a dominant prey item in other parts of the species' range, it is likely that these crashes in local populations would also impact prey availability, and potentially survival, for knots.

While this study focused on populations affected by HABs in the Gulf of Mexico, blooms have been suggested as a potential cause of several significant mortality events on the Atlantic coast of South America, affecting the long-distance migrant *rufa* population wintering in Tierra del Fuego. In Uruguay in April 2007, approximately 1300 knots were found dead in a single event that may have been associated with a HAB, though samples were not collected to confirm the cause of mortality (Aldabe et al., 2015). The loss of ~6% of the total *rufa* population in a single documented event, and the possibility that this may not have been a one-off event but could even occur with some regularity in remote parts of its range provides a potential partial explanation for the dramatic collapse of the Red Knot population that winters on the Atlantic coast of South America. During mortality events in 1997 and 2000 in southern Brazil, Buehler et al. (2010) described similar condition of Red Knots immediately prior to mortality – disorientation, lethargy, unresponsiveness – as witnessed in red tide events in Texas (Newstead, pers. obs.) and Florida, but pathology reports were inconclusive as to the primary cause of death.

Further, Red Knots that winter along the Pacific coasts of Central and South America (the majority of which are suspected to use the focal locations of this study as stopovers; Newstead, unpublished data) may also be encountering increased frequency and intensity of HABs (Band-Schmidt et al., 2019), including several recent events in Ecuador (Torres, 2015; Borbor-Cordova et al., 2019) and Chile (Mardones et al., 2010; Paredes et al., 2019). Several dinoflagellate species that produce paralytic or diarrhetic shellfish poisons can reach bloom concentrations resulting in fish kills and other toxic effects in areas of Central and South America known to be important stopovers. Among these, *Gymnodinium catenatum*, the *Alexandrium tamarense* complex, and *Dinophysis* spp. produce toxins that become highly concentrated in bivalve species such as wedge clams, *Donax hanleyanus*, and blue mussels, *Mytilus edulis* (Carreto et al., 1986; Mee et al., 1986; Méndez and Carreto, 2018), both known to be favored prey items of red knots. The distribution and frequency of HABs appear to be increasing in Central and South America (Band-Schmidt et al., 2019), as well as in the Gulf of Mexico (Tominack et al., 2020).

The magnitude of the geolocator effect was a ~3% reduction in seasonal survival (or ~8% over a year). While many studies reporting tracking device effects on survival have focused on the short-term (often one-year return rates) with projects having highly variable numbers of birds with and without devices, the results of this study are consistent with others (Rodriguez-Ruiz et al., 2019; Pakanen et al., 2020) finding that negative effects of some tracking devices may be statistically undetectable in the short term but accrue to the level of significance over the course of longer-term studies. The use of tracking devices on wildlife has yielded transformative new insights into our understanding of life histories and factors affecting distribution and movements of animals, especially Red Knots (Niles et al., 2010; Burger et al., 2012; Niles et al., 2012; Newstead et al., 2013; Tomkovich et al., 2013; Piersma et al., 2021). However, consideration must be given to the potential costs of such deployments on survival, reproduction, movement, and other concerns. As new findings are added to the literature and technological advances lead to ever smaller and more efficient tracking devices, researchers should continue to assess the potential benefits to be gained for species conservation relative to the potential impacts to birds when planning new studies.

This study provides the first long-term apparent survival estimates for Red Knot populations in the Gulf, and strong evidence that HABs are negatively affecting populations in Texas and Florida. Preventing such large-scale events presents many challenges, although where their apparent causes are linked to excessive nutrients these factors can be mitigated by better managing anthropogenic landscape changes along the coast and through the watershed. Since HABs are considered a “co-stressor” associated with climate change (Griffith and Gobler, 2020), these findings indicate the impacts to knots could become even more severe in the future.

Accurate estimation of population size of these three Gulf wintering groups has not been possible, and is hindered by several factors including the potential shift of some portion of the Florida wintering population to the southeast US (Pelton et al., 2022), logistical difficulties in accessing habitats used by the Louisiana and Texas populations during winter, and the fact that some knots that pass through the northern Gulf in spring likely wintered somewhere further south. These are all surmountable obstacles provided adequate support for dedicated and coordinated monitoring programs. While we have presented estimates of one key demographic parameter (survival) for these populations, a better understanding of processes and rates of recruitment is needed to evaluate population trajectories.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

Ethical approval was not required for the study involving animals in accordance with the local legislation and institutional requirements because Researchers conducting field work were not affiliated with institutions with their own ethics review process. USFWS Recovery Permits (which were obtained for this work) require extensive explanation of capture/handling procedures and contingencies.

Author contributions

DN: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Writing – original draft, Writing – review & editing. SD: Formal analysis, Methodology, Writing – review & editing. BB: Methodology, Supervision, Writing – review & editing. LN: Conceptualization, Writing – review & editing. JB: Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. Partial funding for captures and resight efforts were provided by grants from US Fish & Wildlife Service (201818G909, F16AP00835, F19AC00712, F19AP00937) and Florida Fish and Wildlife Conservation Commission (9713-720-6328).

Acknowledgments

Thanks to Humphrey Sitters, Amanda Dey, and Bill Howe for initiating the study of knots in Texas, and to Delaina LeBlanc of Barataria-Terrebonne National Estuary Program for assistance with logistics, capture, and resighting since the beginning of the study in Louisiana. Thanks to Jeannine Parvin and Lena Usyk for managing and corralling encounter data from www.bandedbirds.org, and our sincerest appreciation to all the dedicated biologists and volunteers for their time assisting in catches and reporting resights over many years and many sites. This is manuscript 24-103 of the Caesar Kleberg Wildlife Research Institute.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fevo.2024.1375412/full#supplementary-material

References

Abbott, B. C., Siger, A., and Spigelstein, M. (1975). "Toxins from the blooms of *Gymnodinium breve*," in *Proceedings of the First International Conference on Toxic Dinoflagellate Blooms*. Ed. V. R. LoCicero (Massachusetts Science and Technology Foundation, Wakefield, Massachusetts), 355–366.

Aldabe, J., Rocca, P. I., González, P. M., Caballero-Sadi, D., and Baker, A. J. (2015). Migration of endangered Red Knots *Calidris canutus rufa* in Uruguay: important sites, phenology, migratory connectivity and a mass mortality event. *Wader Study* 122, 221–235. doi: 10.18194/ws.00024

Baker, A. J., González, P. M., Morrison, R. I. G., and Harrington, B. (2013). "Red Knot (*Calidris canutus*), *v.2.0*," in *The Birds of North America*. Ed. A. F. Poole (Cornell Lab of Ornithology, Ithaca, NY, USA). doi: 10.2173/bna

Baker, A. J., González, P. M., Piersma, T., Niles, L. J., de Lima Serrano do Nascimento, I., Atkinson, P. W., et al. (2004). Rapid population decline in red knots: fitness consequences of decreased refuelling rates and late arrival in Delaware Bay. *Proc. R. Soc London* 271, 875–882. doi: 10.1098/rspb.2003.2663

Band-Schmidt, C. J., Durán-Riveroll, L. M., Bustillos-Guzmán, J. J., Leyva-Valencia, I., López-Cortés, D. J., Núñez-Vázquez,, et al. (2019). Paralytic toxin producing dinoflagellates in Latin America: ecology and physiology. *Front. Mar. Sci.* 6. doi: 10.3389/fmars.2019.00042

Barron, D. G., Brawn, J. D., and Weatherhead, P. J. (2010). Meta-analysis of transmitter effects on avian behavior and ecology. *Methods Ecol. Evol.* 1, 180–187. doi: 10.1111/j.2041-210X.2010.00013.x

Bart, J., Brown, S., Harrington, B., and Morrison., R. I. G. (2007). Survey trends of North American shorebirds: population declines or shifting distirbutions? *J. Avian Biol.* 38, 73–82. doi: 10.1111/j.0908-8857.03698.x

Borbor-Cordova, M. J., Torres, G., Mantilla-Saltos, G., Casierra-Tomala, A., Bermúdez, J. R., Renteria, W., et al. (2019). Oceanography of harmful algal blooms on the Ecuadorian Coast, (1997–2017): Integrating remote sensing and biological data. *Front. Mar. Sci.* 6. doi: 10.3389/fmars.2019.00013

Boyce, M. S. (1977). Population growth with stochastic fluctuations in the life table. *Theor. Popul. Biol.* 12, 366–373. doi: 10.1016/0040-5809(77)90050-8

Boyd, H., and Piersma, T. (2001). Changing balance between survival and recruitment explains population trends in Red Knots *Calidris canutus islandica* wintering in Britain 1969-1995. *Ardea* 89, 301–317.

Brand, L. E., and Compton, A. (2007). Long-term increase in Karenia brevis abundance along the Southwest Florida Coast. *Harmf. Algae* 6, 232–252. doi: 10.1016/j.hal.2006.08.005

Bricelj, V. M., Haubois, A.-G., Sengco, M. R., Pierce, R. H., Culter, J. K., and Anderson, D. M. (2012). Trophic transfer of brevetoxins to the benthic macrofaunal community during a bloom of the harmful dinoflagellate *Karenia brevis* in Sarasota Bay, Florida. *Harmf. Algae* 16, 27–34. doi: 10.1016/j.hal.2012.01.001

Bridge, E. S., Thorup, K., Bowlin, M. S., Chilson, P. B., Diehl, R. H., and Fléron, R. W. (2010). Technology on the move: recent and forthcoming innovations for tracking migratory birds. *BioScience* 61, 689–698. doi: 10.1525/bio.2011.61.9.7

Buchanan, J. B., Johnson, J. A., Bishop, M. A., DeCicco, L. H., Hajdukovich, N., Salzer, L. J., et al. (2019). Does radio transmitter attachment influence resighting rates of Red Knots in the Pacific Flyway, USA? *Wader Study* 126, 151–154. doi: 10.18194/ws

Buchanan, J. B., Lyons, J. E., Salzer, L. J., Carmona, R., Arce, N., Wiles, G. J., et al. (2012). Among-year site fidelity of Red Knots during migration in Washington. *J. Field Ornithol.* 83, 282–289. doi: 10.1111/jofo.2012.83.issue-3

Buehler, D. M., Bugoni, L., Dorrestein, G. M., González, P. M., Pereira-Jr, J., Proença, L., et al. (2010). Local mortality events in migrating sandpipers (*Calidris*) at a staging site in southern Brazil. *Wader Study Group Bull.* 117, 150–156.

Buehler, D. M., Piersma, T., Matson, K., and Tieleman, I. (2008). Seasonal redistribution of immune function in a migrant shorebird: annual-cycle effects override adjustments to thermal regime. *Am. Nat.* 172, 783–796. doi: 10.1086/592865

Burger, J., Niles, L. J., Porter, R. R., Dey, A. D., Koch, S., and Gordon, C. (2012). Migration and over-wintering of red knots (Calidris canutus rufa) along the atlantic coast of the united states. *Condor* 114 (2), 302–313. doi: 10.1525/cond.2012.110077

Burnham, K. P., and Anderson, D. R. (2002). *Model Selection and Multimodel Inference: a Practical Information-theoretic Approach. 2nd ed* (New York: Springer-Verlag).

Buskey, E. J., Liu, H., Collumb, C., and Bersano, J. G. F. (2001). The decline and recovery of a persistent Texas brown tide algal bloom in the Laguna Madre (Texas, USA). *Estuaries* 24, 337–346. doi: 10.2307/1353236

Calvert, A. M., Amirault, D. L., Shaffer, F., Elliot, R., Hanson, A., McKnight, J., et al. (2006). Population assessment of an endangered shorebird: the Piping Plover (*Charadrius melodus melodus*) in eastern Canada. *Avian Conserv. Ecol. - Écol. Conserv. Des. oiseaux* 1, 4. doi: 10.5751/ACE-00091-010304

Carreto, J. I., Benavides, H. R., Negri, R. M., and Glorioso, P. D. (1986). Toxic red-tide in the Argentine Sea. Phyoplankton distribution and survival of the toxic dinoflagellate Gonyaulax excavata in a frontal area. *J. Plankton Res.* 8, 15–28. doi: 10.1093/plankt/8.1.15

Castle, K. T., Flewelling, L. J., Bryan, J. II, Kramer, A., Lindsay, J., Nevada, C., et al. (2013). Coyote (*Canis latrans*) and domestic dog (*Canis familiaris*) mortality and morbidity due to a *Karenia brevis* red tide in the Gulf of Mexico. *J. Wildl. Dis.* 49, 955–964. doi: 10.7589/2012-11-299

Chan, Y.-C., Tibbitts, T. L., Lok, T., Hassell, C. J., Peng, H.-B., Ma, Z., et al. (2019). Filling knowledge gaps in a threatened shorebird flyway through satellite tracking. *J. Appl. Ecol.* 56, 2305–2315. doi: 10.1111/1365-2664.13474

Cohen, J. B., Karpanty, S. M., Catlin, D. H., Fraser, J. D., and Fischer, R. A. (2008). Winter ecology of Piping Plovers at Oregon Inlet, North Carolina. *Waterbirds* 31, 472–479. doi: 10.1675/1524-4695-31.3.472

Conklin, J. R., and Battley, P. F. (2010). Attachment of geolocators to Bar-tailed Godwits: a tibia-mounted method with no survival effects or loss of units. *Wader Study Group Bull.* 117, 56–58.

COSEWIC. (2007). *COSEWIC assessment and status report on the Red Knot Calidris canutus in Canada* (Ottawa: Committee on the Status of Endangered Wildlife in Canada). Available at: www.sararegistry.gc.ca/status/status_e.cfm. Vii + 58 pp.

Costantini, D., and Møller, A. P. (2013). A meta-analysis of the effects of geolocator application on birds. *Curr. Zool.* 59, 697–706. doi: 10.1093/czoolo/59.6.697

Cummins, J. M., Jones, A. C., and Stevens, A. A. (1971). Occurrence of toxic bivalve molluscs during a *Gymnodinium breve* "red tide". *Trans. Am. Fish. Soc* 100, 112–116. doi: 10.1577/1548-8659(1971)100

DeYoe, H. R., Buskey, E. J., and Jochem, F. J. (2007). Physiological responses of *Aureoumbra lagunensis* and *Synechococcus* sp. to nitrogen addition in a mesocosm experiment. *Harmf. Algae* 6, 48–55. doi: 10.1016/j.hal.2006.06.001

Drake, K. R., Thompson, J. E., Drake, K. L., and Zonick, C. (2001). Movements, habitat use, and survival of nonbreeding Piping Plovers. *Condor* 103, 259–267. doi: 10.1093/condor/103.2.259

Duijns, S., Anderson, A. M., Aubry, Y., Dey, A., Flemming, S. A., Francis, C. M., et al. (2019). Long-distance migratory shorebirds travel faster towards their breeding grounds, but fly faster post-breeding. *Sci. Rep.* 9, 9420. doi: 10.1038/s41598-019-45862-0

Duijns, S., Niles, L. J., Dey, A., Aubry, Y., Friis, C., Koch, S., et al. (2017). Body condition explains migratory performance of a long-distance migrant. *Proc. R. Soc B* 284, 20171374. doi: 10.1098/rspb.2017.1374

Duriez, O., Ens, B. J., Choquet, R., Pradel, R., and Klaassen, M. (2012). Comparing the seasonal survival of resident and migratory oystercatchers: carry-over effects of habitat quality and weather conditions. *Oikos* 121, 862–873. doi: 10.1111/j.1600-0706.2012.20326.x

Elliott, K. H., McFarlane-Tranquilla, L., Burke, C. M., Hedd, A., Montevecchi, W. A., and Anderson, W. G. (2012). Year-long deployments of small geolocators increase corticosterone levels in murres. *Mar. Ecol. Prog. Ser.* 466, 1–7. doi: 10.3354/meps09975

Elliott-Smith, E., Bidwell, M., Holland, A. E., and Haig, S. M. (2015). *Data from the 2011 International Piping Plover Census* (Reston, Virginia, USA: U.S. Geological Survey).

Ellis, K. S., Anteau, M. J., Cuthbert, F. J., Gratto-Trevor, C. L., Jorgensen, J. G., Newstead, D. J., et al. (2021). Impacts of extreme environmental disturbances on piping plover survival are partially moderated by migratory connectivity. *Biol. Conserv.* 264, 109371. doi: 10.1016/j.biocon.2021.109371

Faaborg, J., Holmes, R. T., Anders, A. D., Bildstein, K. L., Dugger, K. M., Gauthreaux, J. S.A., et al. (2010). Recent advances in understanding migration systems of New World land birds. *Ecol. Monogr.* 80, 3–48. doi: 10.1890/09-0395.1

Fernández, G., and Lank, D. B. (2006). Effects of habitat loss on shorebirds during the non-breeding season: current knowledge and suggestions for action. *Ornitol. Neotro.* 19, 633–640.

Gauthier, G., Pradel, R., Menu, S., and Lebreton, J.-D. (2001). Seasonal survival of Greater Snow Geese and effect of hunting under dependence in sighting probability. *Ecology* 82, 3105–3119. doi: 10.1890/0012-9658(2001)082[3105:SSOGSG]2.0.CO;2

Gill, J. A., and Sutherland, W. J. (2000). "Predicting the consequences of human disturbance from behavioural decisions," in *Behaviour and Conservation*. Eds. L. M. Morris and W. J. Sutherland (Cambridge Univ. Press, Cambridge, UK), 51–64.

Gobler, C. J., and Sunda, W. G. (2012). Ecosystem disruptive algal blooms of the brown tide species, *Aureococcus anophagefferens* and *Aureoumbra lagunensis*. *Harmf. Algae* 14, 360–345. doi: 10.1016/j.hal.2011.10.013

González, P. M., Baker, A. J., and Echave, M. E. (2006). Annual survival of red knots (Calidris canutus rufa) using the san antonio oeste stopover site is reduced by domino effects involving late arrival and food depletion in delaware bay. *El hornero* 21 (2), 109–117.

Goss-Custard, J. D., Clarke, R. T., Durell, S. E. A., dit., L., Caldow, R. W. G., and Ens, B. J. (1995). Population consequences of winter habitat loss in a migratory shorebird. II. model predictions. *J. Appl. Ecol.* 32, 337–351.

Green, M., Piersma, T., Jukema, J., De Goeij, P., Spaans, B., and van Gils, J. A. (2002). Radio-telemetry observations of the first 650 km of the migration of Bar-tailed Godwits *Limosa lapponica* from the Wadden Sea to the Russian Arctic. *Ardea* 90, 71–80.

Griffith, A. W., and Gobler, C. J. (2020). Harmful algal blooms: A climate change co-stressor in marine and freshwater ecosystems. *Harmf. Algae* 91, 1–12. doi: 10.1016/j.hal.2019.03.008

Hallegraeff, G. M. (1993). A review of harmful algal blooms and their apparent global increase. *Phycology* 322, 79–99. doi: 10.2216/i0031-8884-32-2-79.1

Harrington, B. A., Hagan, J. M., and Leddy, L. E. (1998). Site fidelity and survival differences between two groups of new world Red Knots (*Calidris canutus*). *Auk* 105, 439–445. doi: 10.1093/auk/105.3.439

Hitchcock, C. L., and Gratto-Trevor, C. (1997). Diagnosing a shorebird local population decline with a stage-structured population model. *Ecology* 78, 522.534. doi: 10.1890/0012-9658(1997)078[0522:DASLPD]2.0.CO;2

Iwamura, T., Possingham, H. P., Chadès, I., Minton, C., Murray, N. J., Rogers, D. I., et al. (2013). Migratory connectivity magnifies the consequences of habitat loss from sea-level rise for shorebird populations. *Proc. R Soc. B* 280, 20130325. doi: 10.1098/rspb.2013.0325

Johnson, M. D., Sherry, T. W., Holmes, R. T., and Marra, P. P. (2006). Assessing habitat quality for a migratory songbird wintering in natural and agricultural habitats. *Conserv. Biol.* 20, 1433–1444. doi: 10.1111/j.1523-1739.2006.00490.x

Kennish, M. J. (2002). Environmental threats and environmental future of estuaries. *Environ. Conserv.* 29, 78–107. doi: 10.1017/S0376892902000061

Kvitek, R., and Bretz, C. (2005). Shorebird foraging behavior, diet, and abundance vary with harmful algal bloom toxin concentrations in invertebrate prey. *Mar. Ecol. Prog. Ser.* 293, 303–309. doi: 10.3354/meps293303

Lande, R. (1993). Risks of population extinction from demographic and environmental stochasticity and random catastrophes. *Am. Nat.* 142, 911–927. doi: 10.1086/285580

Landsberg, J. H. (2002). The effects of harmful algal blooms on aquatic organisms. *Rev. Fish. Sci.* 10, 113–390. doi: 10.1080/20026491051695

Landsberg, J. H., Flewelling, L. J., and Naar, J. (2009). *Karenia brevis* red tides, brevetoxins in the food web, and impacts on natural resources: decadal advancements. *Harmf. Algae* 8, 598–607. doi: 10.1016/j.hal.2008.11.010

Lerma, L. (2013). The effects of a red tide, *Karenia brevis* episode on the benthic macroinvertebrate communities of South Padre Island, TX. University of Texas at Brownsville, Brownsville, Texas.

Leyrer, J., Lok, T., Brugge, M., Spaans, B., Sandercock, B. K., and Piersma, T. (2013). Mortality within the annual cycle: seasonal survival patterns in Afro-Siberian Red Knots *Calidris canutus canutus*. *J. Ornithol.* 154, 933–943. doi: 10.1007/s10336-013-0959-y

Leyrer, J., Spaans, B., Camara, M., and Piersma, T. (2006). Small home ranges and high site fidelity in red knots (*Calidris c. canutus*) wintering on the Banc d'Arguin, Mauritania. *J. Ornithol.* 147, 376–384. doi: 10.1007/s10336-005-0030-8

Lyons, J. E., Winn, B., Keyes, T., and Kalasz, K. S. (2018). Post-breeding migration and connectivity of Red Knots in the western Atlantic. *J. Wildl. Manage.* 82, 383–396. doi: 10.1002/jwmg.21389

Magaña, H. A., Contreras, C., and Villareal, T. A. (2003). A historical assessment of *Karenia brevis* in the western Gulf of Mexico. *Harmf. Algae* 2, 163–171. doi: 10.1016/S1568-9883(03)00026-X

Mardones, J., Clement, A., Rojas, X., and Aparicio, C. (2010). Alexandrium catenella during 2009 in Chilean waters, and recent expansion to coastal ocean. *Harmf. Algae News* 41, 8-9.

McGowan, C. P., Hines, J. E., Nichols, J. D., Lyons, J. E., Smith, D. R., Kalasz, K. S., et al. (2011). Demographic consequences of migratory stopover: linking red knot survival to horseshoe crab spawning abundance. *Ecosphere* 2, 1–22. doi: 10.1890/ES11-00106.1

McKellar, A. E., Ross, R. K., Morrison, R. I. G., Niles, L. J., Porter, R. R., Burger, J., et al. (2015). Shorebird use of western Hudson Bay near the Nelson River during migration, with a focus on the Red Knot. *Wader Study* 122, 1–11. doi: 10.18194/ws.00020

Mee, L. D., Espinosa, M., and Diaz, G. (1986). Paralytic shellfish poisoning with a *Gymnodinium catenatum* red tide on the Pacific coast of Mexico. *Mar. Environ. Res.* 19, 77–92. doi: 10.1016/0141-1136(86)90040-1

Méndez, S. M., and Carreto, J. I. (2018). "Harmful Algal Blooms in the Río de la Plata Region," in *Plankton Ecology of the Southwestern Atlantic: From the Subtropical to the Subantarctic Realm*. Eds. M. Hoffmeyer, M. E. Sabatini, F. Brandini, D. L. Calliari and N. H. Santinelli (Springer International Publishing, Switzerland). doi: 10.1007/978-3-319-77869-3

Mondain-Monval, T. O., du Feu, R., and Sharp, S. P. (2020). The effects of geolocators on return rates, condition, and breeding success in Common Sandpipers *Actitis hypoleucos*. *Bird Study* 67, 217–223. doi: 10.1080/00063657.2020.1808592

Montagna, P. A., Stockwell, D. A., and Kalke, R. D. (1993). Dwarf surfclam *Mulinia lateralis* (Say 1822) populations and feeding during the Texas brown tide event. *J. Shellfish Res.* 12, 433–442.

Musmeci, L. R., Bala, L. O., Tschopp, A., Hernandez, M. A., and Coscarella, M. A. (2022). Red knot (*Calidris canutus rufa*) site fidelity at Península Valdés, Argentina. *Wilson J. Ornithol.* 134, 302-309. doi: 10.1676/19-00108

Newstead, D. J. (2014). *Habitat use of North Padre Island and Laguna Madre habitats by Piping Plovers (*Charadrius melodus*) and Red Knots (*Calidris canutus*) in the vicinity of current and proposed wind energy development* (Austin, Texas: Report to Texas Parks & Wildlife Department, project E-137-R). 47 pp.

Newstead, D. J., Niles, L. J., Porter, R. R., Dey, A., Burger, J., and Fitzsimmons, O. N. (2013). Geolocation reveals mid-continent migratory routes and Texas wintering areas of Red Knots *Calidris canutus rufa*. *Wader Study Group Bull.* 120, 53–59.

Niles, L. J., Burger, J., Porter, R. R., Dey, A. D., Koch, S., Harrington, B., et al. (2012). Migration pathways, migration speeds and non-breeding areas used by northern hemisphere wintering Red Knots *Calidris canutus* of the subspecies *rufa*. *Wader Study Group Bull.* 119, 195–203.

Niles, L. J., Burger, J., Porter, R. R., Dey, A. D., Minton, C. D. T., Gonzalez, P. M., et al. (2010). First results using light level geolocators to track Red Knots in the Western Hemisphere show rapid and long intercontinental flights and new details of migration pathways. *Wader Study Group Bull.* 117, 123–130.

Niles, L. J., Sitters, H. P., Dey, A. D., Atkinson, P. W., Baker, A. J., Bennett, K. A., et al. (2008). Status of the Red Knot, *Calidris canutus rufa*, in the Western Hemisphere. *Stud. Avian Biol.* 36, 1–185.

NOAA National Centers for Environmental Information. (2014). *Physical and biological data collected along the texas, mississippi, alabama, and florida gulf coasts in the gulf of mexico as part of the harmful algal BloomS observing system from 1953-08-19 to 2023-07-06 (NCEI accession 0120767)* (NOAA National Centers for Environmental Information). Dataset. Available at: https://www.ncei.noaa.gov/archive/accession/0120767.

Norris, D. R. (2005). Carry-over effects and habitat quality in migratory populations. *Oikos* 109, 178–186. doi: 10.1111/j.0030-1299.2005.13671.x

Norris, D. R., and Marra, P. P. (2007). Seasonal interactions, habitat quality, and population dynamics in migratory birds. *Condor* 109 (3), 535–547. doi: 10.1093/condor/109.3.535

Pakanen, V.-M., Rönkä, N., Thomson, R. L., Blomqvist, D., and Koivula, K. (2020). Survival probability in a small shorebird decreases with the time an individual carries a tracking device. *J. Avian Biol.* 51, e02555. doi: 10.1111/jav.02555

Pakanen, V.-M., Rönkä, N., Thomson, R. L., and Koivula, K. (2015). No strong effects of leg-flagged geolocators on return rates or reproduction of a small long-distance migratory shorebird. *Ornis Fennica* 92, 101–111. doi: 10.51812/of.133872

Paredes, J., Varela, D., Martínez, C., Zúñiga, A., Correa, K., Villarroel, A., et al. (2019). Population Genetic Structure at the Northern Edge of the Distribution of Alexandrium catenella in the Patagonian Fjords and Its Expansion Along the Open Pacific Ocean Coast. *Front. Mar. Sci.* 5. doi: 10.3389/fmars.2018.00532

Pelton, M. M., Padula, S. R., Garcia-Walther, J., Andrews, M., Mercer, R., Porter, R., et al. (2022). Kiawah and Seabrook islands are a critical site for the *rufa* Red Knot. *Wader Study* 129, 105–118. doi: 10.18194/ws

Pierce, R. H., and Henry, M. S. (2008). Harmful algal toxins of the Florida red tide (*Karenia brevis*): natural chemical stressors in South Florida coastal ecosystems. *Ecotoxicology* 17, 623–631. doi: 10.1007/s10646-008-0241-x

Piersma, T., Kok, E. M. A., Hassell, C. J., Peng, H.-B., Verkuil, Y. I., Lei, G., et al. (2021). When a typical jumper skips: itineraries and staging habitats used by Red Knots (*Calidris canutus piersmai*) migrating between northwest Australia and the New Siberian Islands. *Ibis* 163, 1235–1251. doi: 10.1111/ibi.12964

Piersma, T., Lok, T., Chen, Y., Hassell, C. J., Yang, H.-Y., Boyle, A., et al. (2016). Simultaneous declines in summer survival of three shorebird species signals a flyway at risk. *J. Appl. Ecol.* 53, 479–490. doi: 10.1111/1365-2664.12582

Powell, L. A. (2007). Approximating variance of demographic parameters using the delta method: a reference for avian biologists. *Condor* 109, 949–954. doi: 10.1093/condor/109.4.949

Rafalski, A. V. (2012). Evaluation of brevetoxin accumulation and degradation in coastal mammals, birds, and fish found moribund on Texas beaches during red tide blooms. Texas A&M University – Corpus Christi, Corpus Christi, Texas.

Robinson, W. D., Bowlin, M. S., Bisson, I., Shamoun-Baranes, J., Thorup, K., Diehl, R. H., et al. (2010). Integrating concepts and technologies to advance the study of bird migration. *Front. Ecol. Environ.* 8, 354–361. doi: 10.1890/080179

Rodriguez-Ruiz, J., Mougeot, F., Parejo, D., de la Puente, J., Bermejo, A., and Aviles, J. M. (2019). Important areas for the conservation of the european roller coracias garrulus during the non-breeding season in southern africa. *Bird Conserv. Int.* 29 (1), 159–175. doi: 10.1111/ibi.12317

Rogers, D. I., Piersma, T., and Hassell, C. J. (2006). Roost availability may constrain shorebird distribution: Exploring the energetic costs of roosting and disturbance around a tropical bay. *Biol. Conserv.* 133, 225–235. doi: 10.1016/j.biocon.2006.06.007

Rolton, A., Vignier, J., Volety, A. K., Pierce, R. H., Henry, M., Shumway, S. E., et al. (2016). Effects of field and laboratory exposure to the toxic dinoflagellate *Karenia brevis* on the reproduction of the eastern oyster, *Crassostrea virginica*, and subsequent development of offspring. *Harmf. Algae* 57, 13–26. doi: 10.1016/j.hal.2016.04.011

Rushing, C. S., Hostetler, J. A., Sillett, T. S., Marra, P. P., Rotenberg, J. A., and Ryder, T. B. (2017). Spatial and temporal drivers of avian population dynamics across the annual cycle. *Ecology* 98, 2837–2850. doi: 10.1002/ecy.1967

Scarpignato, A. L., Harrison, A.-L., Newstead, D. J., Niles, L. J., Porter, R. R., van den Tillaart, M., et al. (2016). Field-testing a new miniaturized GPS-Argos satellite transmitter (3.5 g) on migratory shorebirds. *Wader Study* 123, 240–246. doi: 10.18194/ws.00046

Schwarzer, A. C., Collazo, J. A., Niles, L. J., Brush, J. M., Douglass, N. J., and Percival, H. F. (2012). Annual survival of Red Knots (*Calidris canutus rufa*) wintering in Florida. *Auk* 129, 725–733. doi: 10.1525/auk.2012.11269

Shumway, S. E., Allen, S. M., and Boersma, P. D. (2003). Marine birds and harmful algal blooms: sporadic victims or under-reported events? *Harmf. Algae* 2, 1–17.

Simberloff, D. (1988). The contribution of population and community biology to conservation science. *Annu. Rev. Ecol. Systematics* 19 (1), 473–511.

Stantial, M. L., Cohen, J. B., Loring, P. H., and Paton, P. W. C. (2019). Radio transmitters did not affect apparent survival rates of adult Piping Plovers (*Charadrius melodus*). *Waterbirds* 42, 205–209. doi: 10.1675/063.042.0207

Stumpf, R. P., Li, Y., Kirkpatrick, B., Litaker, R. W., Hubbard, K. A., Currier, R. D., et al. (2022). Quantifying *Karenia brevis* bloom severity and respiratory irritation impact along the shoreline of Southwest Florida. *PloS One* 17, e0260755. doi: 10.1371/journal.pone.0260755

Stutchbury, B. J. M., Tarof, S. A., Done, T., Gow, E. A., Kramer, P. M., Tautin, J., et al. (2009). Tracking long-distance songbird migration by using geolocators. *Science* 323, 896. doi: 10.1126/science.1166664

Summerson, H. C., and Peterson, C. H. (1990). Recruitment failure of the bay scallop, *Argopecten irradians concentricus*, during the first red tide, *Ptychodiscus brevis*, outbreak recorded in North Carolina. *Estuaries* 13, 3, 322–331. doi: 10.2307/1351923

Sutherland, W. J., Alves, J. A., Amano, T., Chang, C. H., Davidson, N. C., Finlayson, C. M., et al. (2012). A horizon scanning assessment of current and potential future threats facing migratory shorebirds. *Ibis* 154, 663–679. doi: 10.1111/j.1474-919X.2012.01261.x

Tester, P. A., Turner, J. T., and Shea, D. (2000). Vectorial transport of toxins from the dinoflagellate *Gymnodinium breve* through copepods to fish. *J. Plankton Res.* 22, 47–62. doi: 10.1093/plankt/22.1.47

Tominack, S. A., Coffey, K. Z., Yoskowitz, D., Sutton, G., and Wetz, M. S. (2020). An assessment of trends in the frequency and duration of *Karenia brevis* red tide blooms on the South Texas coast (western Gulf of Mexico). *PloS One* 15, e0239309. doi: 10.1371/journal.pone.0239309

Tomkovich, P. S., Porter, R. R., Loktionov, E. Y., and Niles, L. J. (2013). Pathways and staging areas of Red Knots Calidris canutus rogersi breeding in southern Chukotka, Far Eastern Russia. *Wader Study Group Bull.* 120, 181–193.

Torres, G. (2015). Evaluación de mareas rojas durante 1968-2009 en Ecuador. *Acta Oceanográfica del Pacífico* 20, 89–98.

Tucker, A. M., McGowan, C. P., Lyons, J. E., DeRose-Wilson, A., and Clark, N. A. (2021). Species-specific demographic and behavioral responses to food availability during migratory stopover. *Popul. Ecol.* 2021, 1–16.

Tucker, A. M., McGowan, C. P., Nuse, B. L., Lyons, J. E., Moore, C. T., Smith, D. R., et al. (2023). Estimating recruitment rate and population dynamics at a migratory stopover site using an integrated population model. *Ecosphere* 14, 1–16. doi: 10.1002/ecs2.4439

Tuma, M. E., and Powell, A. N. (2021). The southeastern U.S. as a complex of use sites for nonbreeding *rufa* Red Knots: fifteen years of band encounter data. *Wader Study* 128, 265–273. doi: 10.18194/ws

U.S. Fish and Wildlife Service [USFWS]. (2014a). *Threatened species status for the Rufa red knot*. 73706–73748, 79 Federal Register 238 (2014 December 11).

U.S. Fish and Wildlife Service [USFWS]. (2014b). Rufa *red knot background information and threats assessment. Supplement to Endangered and Threatened Wildlife and Plants; Final Threatened Status for the* Rufa *red knot (*Calidris canutus rufa*) [Document No. FWS-R5-ES-2013-0097; RIN AY17]* (Pleasantville, New Jersey, USA: U.S. Fish and Wildlife Service).

U.S. Fish and Wildlife Service [USFWS]. (2021). Rufa *Red Knot Critical Habitat Methods [Document No. FWS-R5-ES-2021-0032-0009]* (Pleasantville, New Jersey, USA: U.S. Fish and Wildlife Service). 12 pp.

van Deventer, M., Atwood, K., Vargo, G. A., Flewelling, L. J., Landsberg, J. H., Naar, J. P., et al. (2012). *Karenia brevis* red tides and brevetoxin-contaminated fish: a high risk factor for Florida's scavenging shorebirds? *Botanica Marina* 55, 31–37. doi: 10.1515/bot.2011.122

van Dolah, F. M. (2000). Marine algal toxins: origins, health effects, and their increased occurrence. *Environ. Health Perspect.* 108, 133–141. doi: 10.1289/ehp.00108s1133

van Gils, J. A., Piersma, T., Dekinga, A., and Battley, P. F. (2006). Modelling phenotypic flexibility: an optimality analysis of gizzard size in Red Knots (*Calidris canutus*). *Ardea* 9), 409–420.

Van Hemert, C., Dusek, R. J., Smith, M. M., Kaler, R., Sheffield, G., Divine, L. M., et al. (2021). Investigation of algal toxins in a multispecies seabird die-off in the Bering and Chukchi Seas. *J. Wildl. Dis.* 57, 399–407. doi: 10.7589/JWD-D-20-00057

Van Hemert, C., Harley, J. R., Baluss, G., Smith, M. M., Dusek, R. J., Lankton, J. S., et al. (2022). Paralytic shellfish toxins associated with Arctic Tern mortalities in Alaska. *Harmf. Algae* 117, 102270. doi: 10.1016/j.hal.2022.102270

van Irsel, J., Frauendorf, M., Ens, B. J., van de Pol, M., Troost, K., Oosterbeek, K., et al. (2022). State-dependent environmental sensitivity of reproductive success and survival in a shorebird. *Ibis* 164, 692–710. doi: 10.1111/ibi.13038

Walsh, J. J., Jolliff, J. K., Darrow, B. P., Lenes, J. M., Milroy, S. P., Remsen, A., et al. (2006). Red tides in the Gulf of Mexico: Where, when, and why? *J. Geophys. Res.* 111, 1–46. doi: 10.1029/2004JC002813

Warnock, N., and Takekawa, J. Y. (2003). Use of radio telemetry in studies of shorebirds: past contributions and future directions. *Wader Study Group Bull.* 100, 1–14.

Webster, M. S., Marra, P. P., Haig, S. M., Bensch, S., and Holmes, R. T. (2002). Links between worlds: unraveling migratory connectivity. *Trends Ecol. Evol.* 17, 76–83. doi: 10.1016/S0169-5347(01)02380-1

Weiser, E. L., Lanctot, R. B., Brown, S. C., Alves, J. A., Battley, P. F., Bentzen, R., et al. (2016). Effects of geolocators on hatching success, return rates, breeding movements, and change in body mass in 16 species of Arctic-breeding shorebirds. *Move. Ecol.* 4, 1–19. doi: 10.1186/s40462-016-0077-6

White, G. C., and Burnham, K. P. (1999). Program MARK: Survival estimation for populations of marked animals. *Bird Study* 46, 120–139.

Wilmers, C. C., Nickel, B., Bryce, C. M., Smith, J. A., Wheat, R. E., and Yovovich, V. (2015). The golden age of bio-logging: how animal-borne sensors are advancing the frontiers of ecology. *Ecology* 96, 1741–1753. doi: 10.1890/14-1401.1

EXHIBIT C

MacDonald, A. et al. 2024
Uniting rufa Red Knot resighting data throughout the western Atlantic Flyway offers myriad opportunities for survival analysis
10th Western Hemisphere Shorebird Group Meeting. Sackville, NB, Canada
August 11-16, 2024

Uniting *rufa* Red Knot resighting data throughout the western Atlantic Flyway offers myriad opportunities for survival analysis

La unión de los datos de reencuentro de Calidris canutus rufa *en todo el hemisferio occidental ofrece múltiples oportunidades para el análisis de la supervivencia*

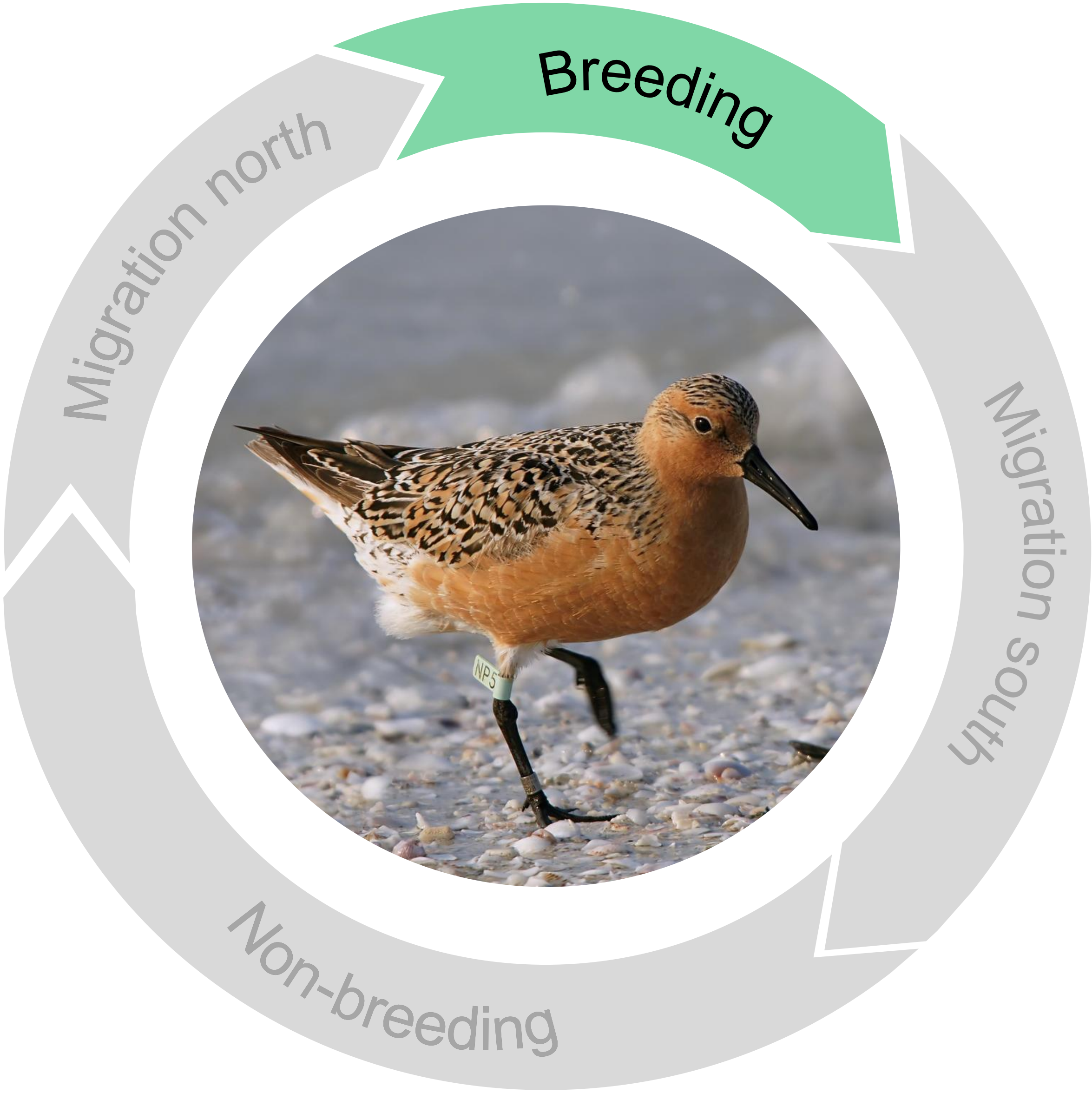
Amie MacDonald, Yves Aubry, Henrietta Bellman, Janell Brush, Christophe Buidin, Jacquie Clark, Nigel Clark, Audrey DeRose-Wilson, Amanda Dey, Theo Diehl, Stephanie Feigin, Christian Friis, Patricia González, Brian Harrington, Kevin Kalasz, Timothy Keyes, Stephanie Koch, Patrick Leary, James Lyons, Natalia Martínez Curci, David Mizrahi, Jason Mobley, David Newstead, Lawrence Niles, Erica Nol, Julie Paquet, Mark Peck, Yann Rochepault, Roberta Rodrigues, Felicia Sanders, Fletcher Smith, Bryan Watts, and Paul Smith

BIRDS CANADA
OISEAUX CANADA

Environment and
Climate Change Canada

Environnement et
Changement climatique Canada

Joe Smith



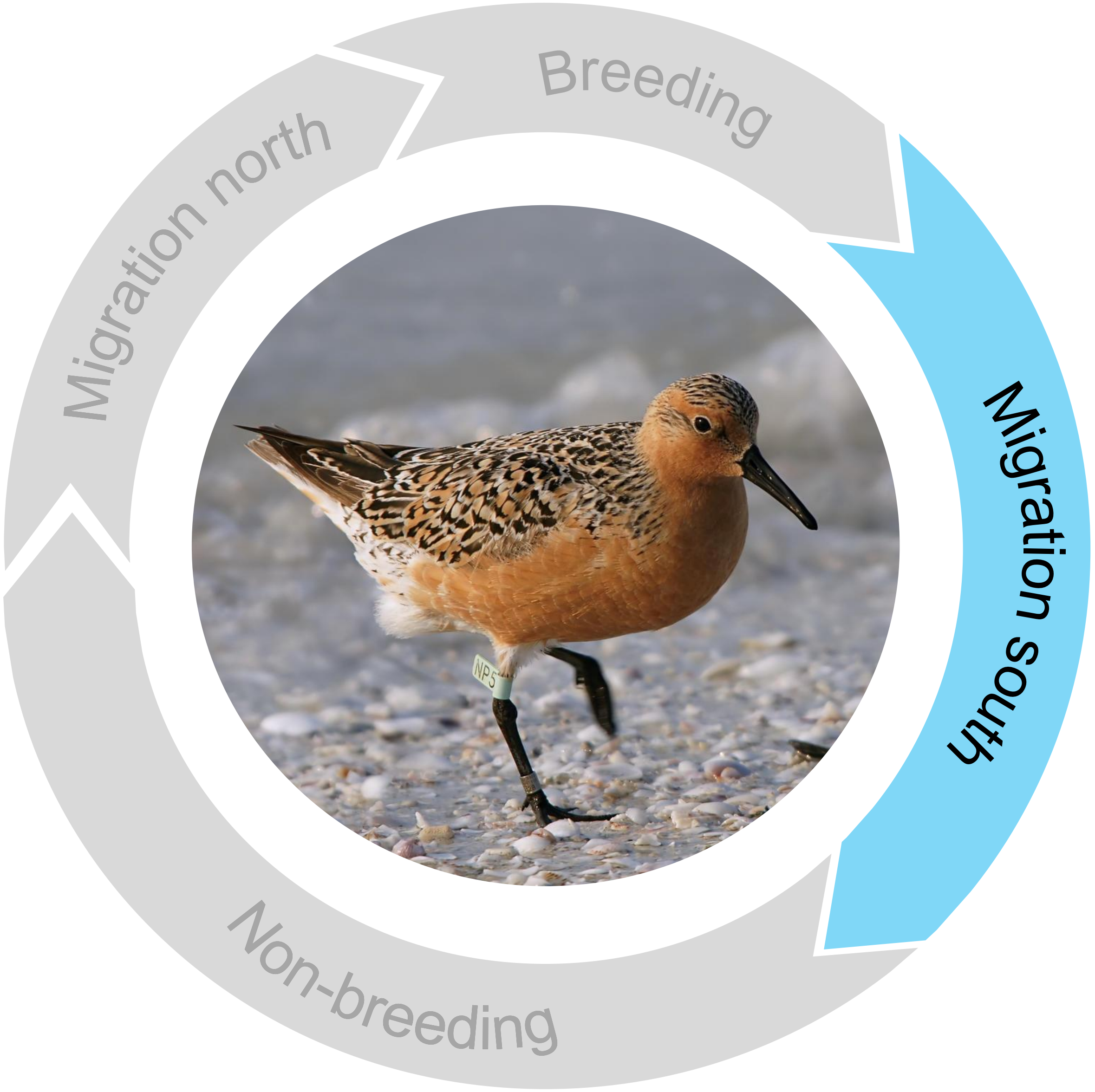
Breeding

Migration south

Non-breeding

Migration north

Joe Smith



Breeding

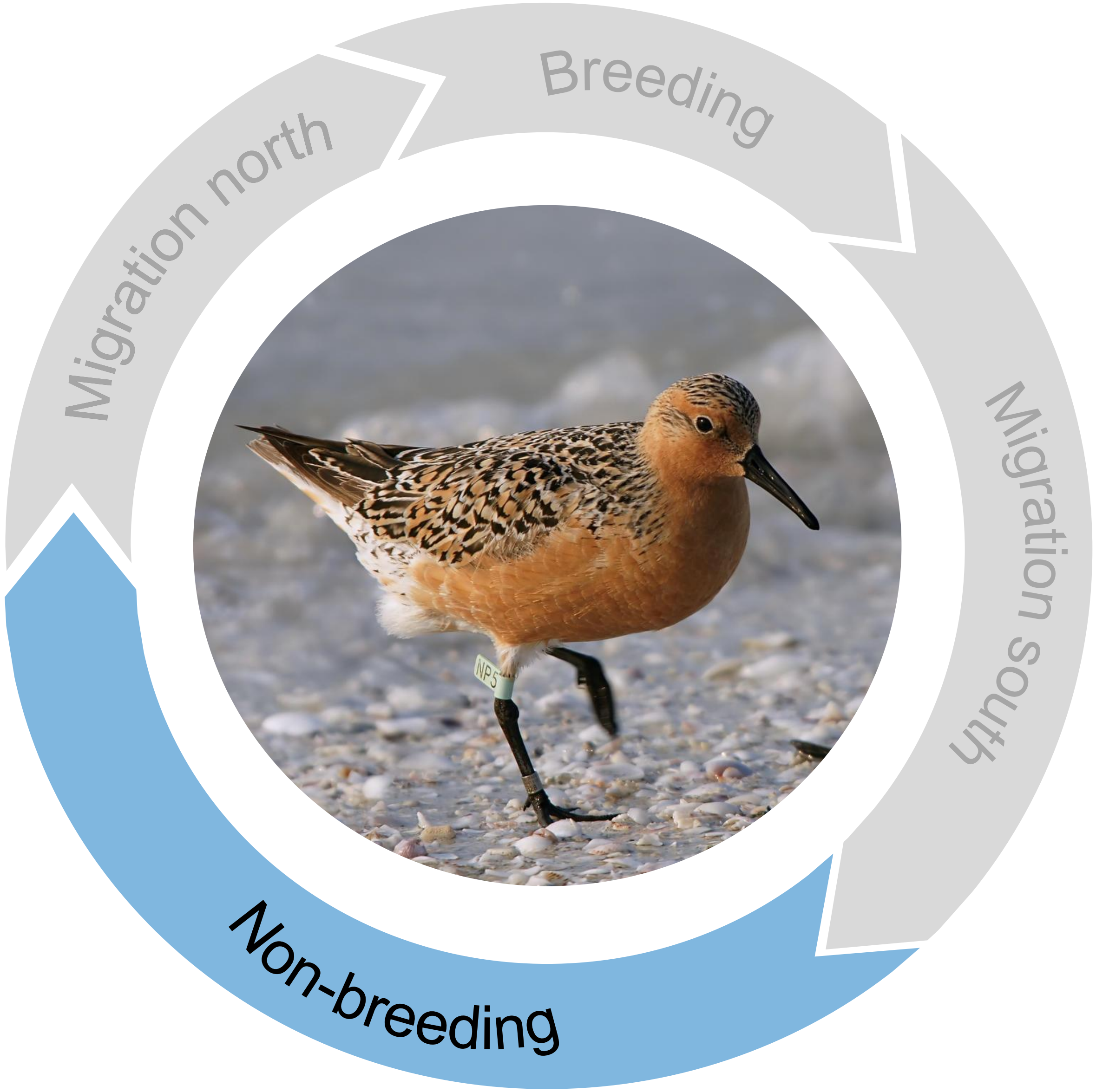
Migration south

Non-breeding

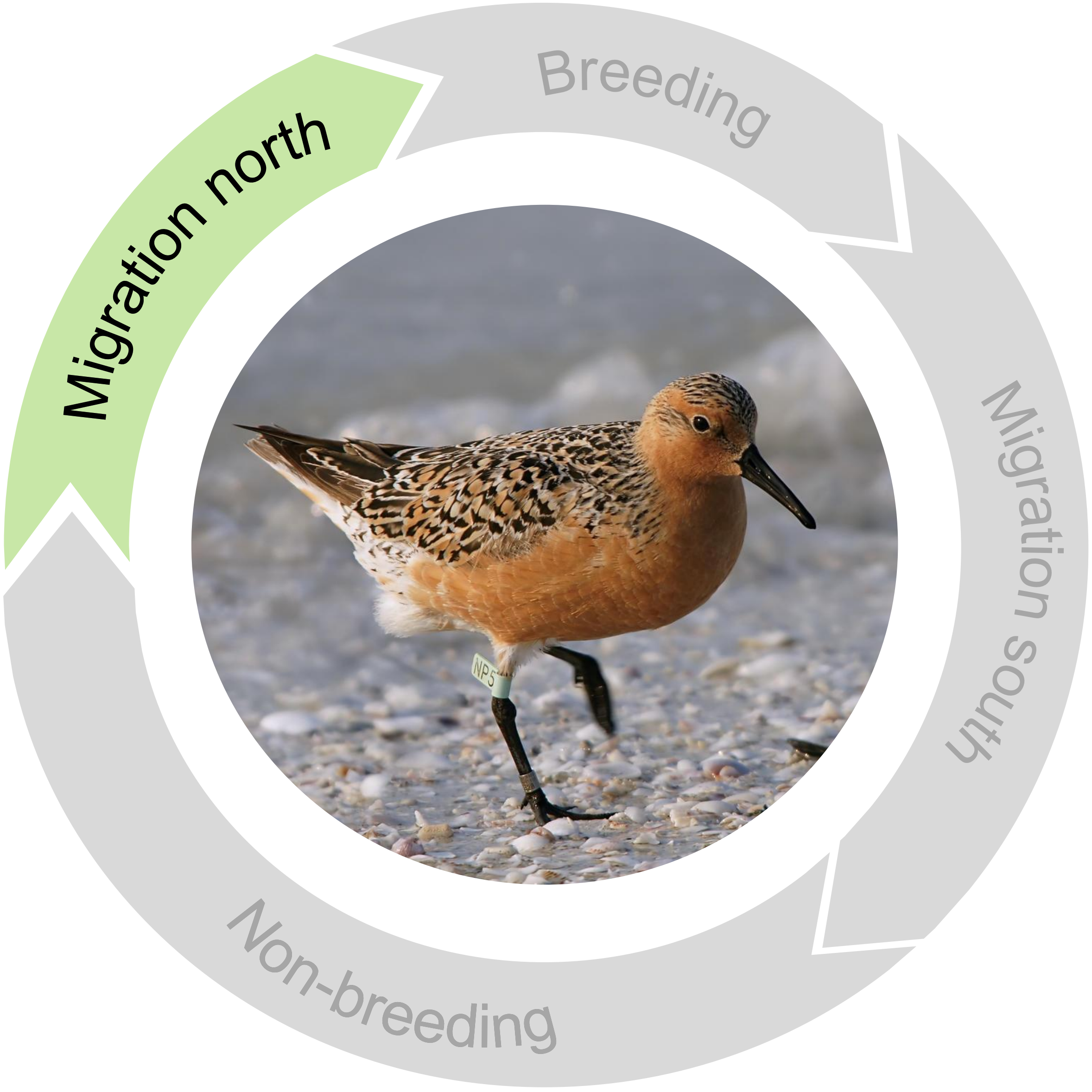
Migration north


NESTING RANGE
James Bay
Mingan
Delaware Bay
KEY WINTERING AREAS
STOPOVER SITES

Joe Smith


Breeding
Migration south
Non-breeding
Migration north
NP5

Joe Smith



Migration north

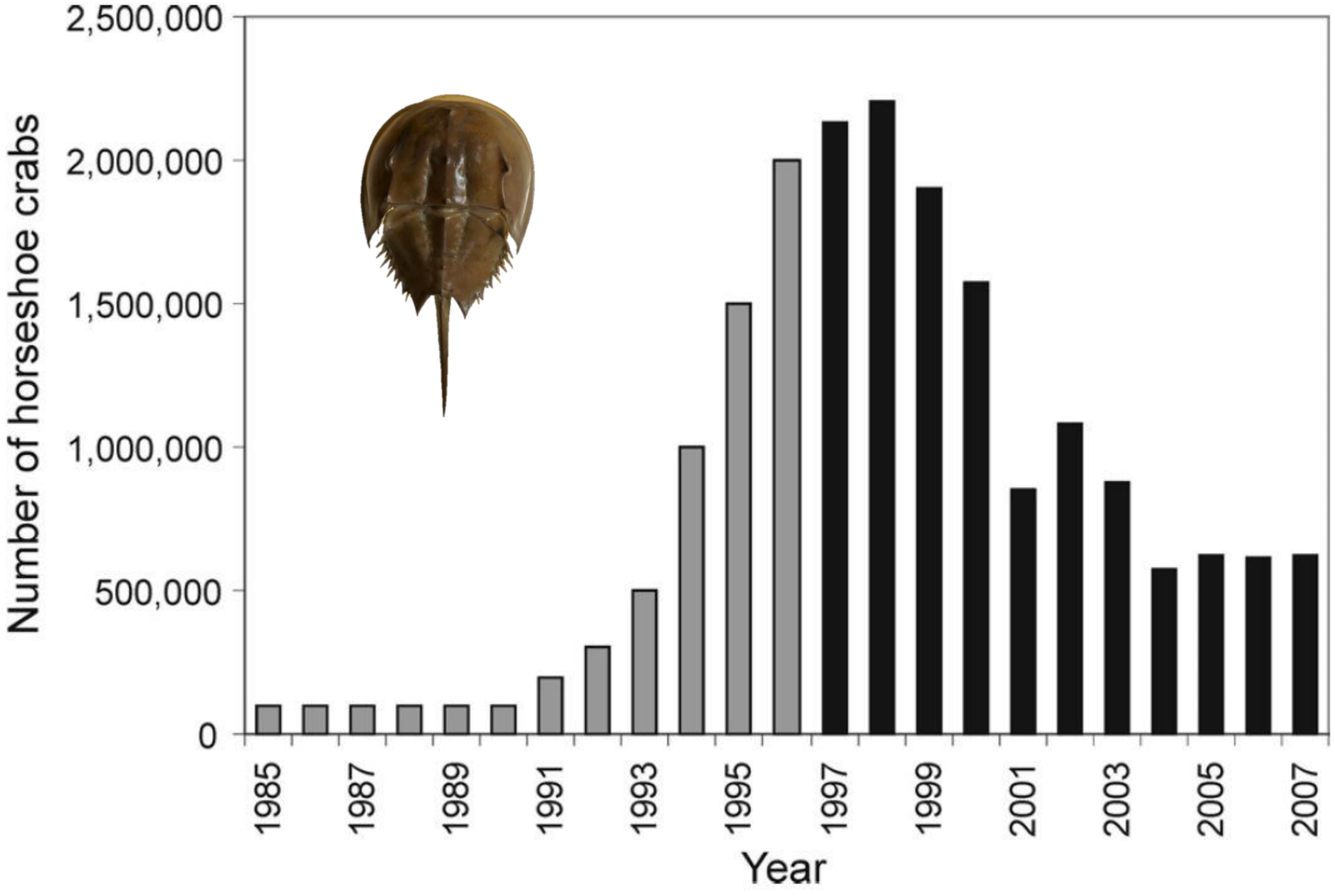
Breeding

Migration south

Non-breeding

Joe Smith

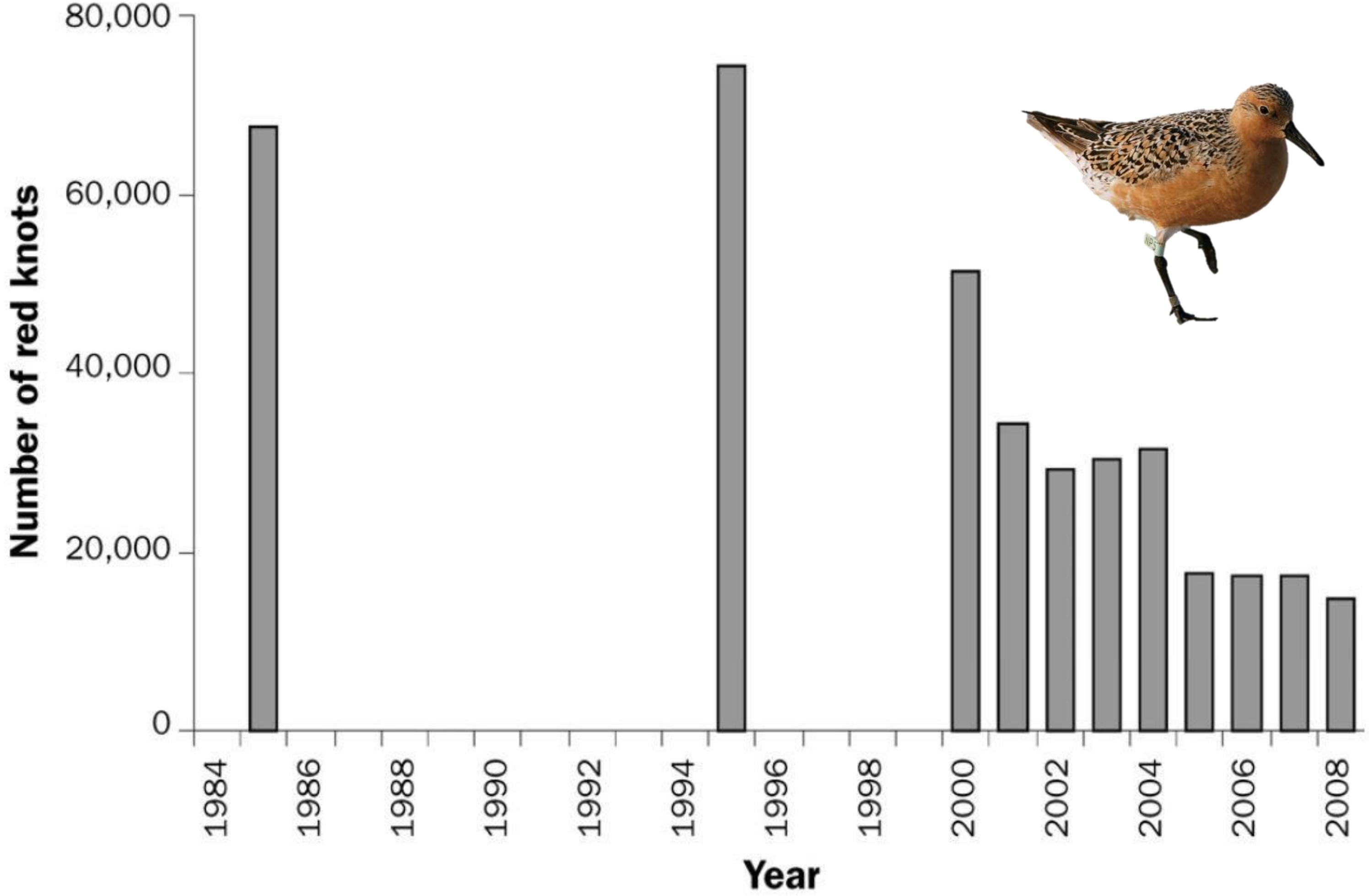
Horseshoe crabs overharvested in late 1990s



Niles et al. (2009)

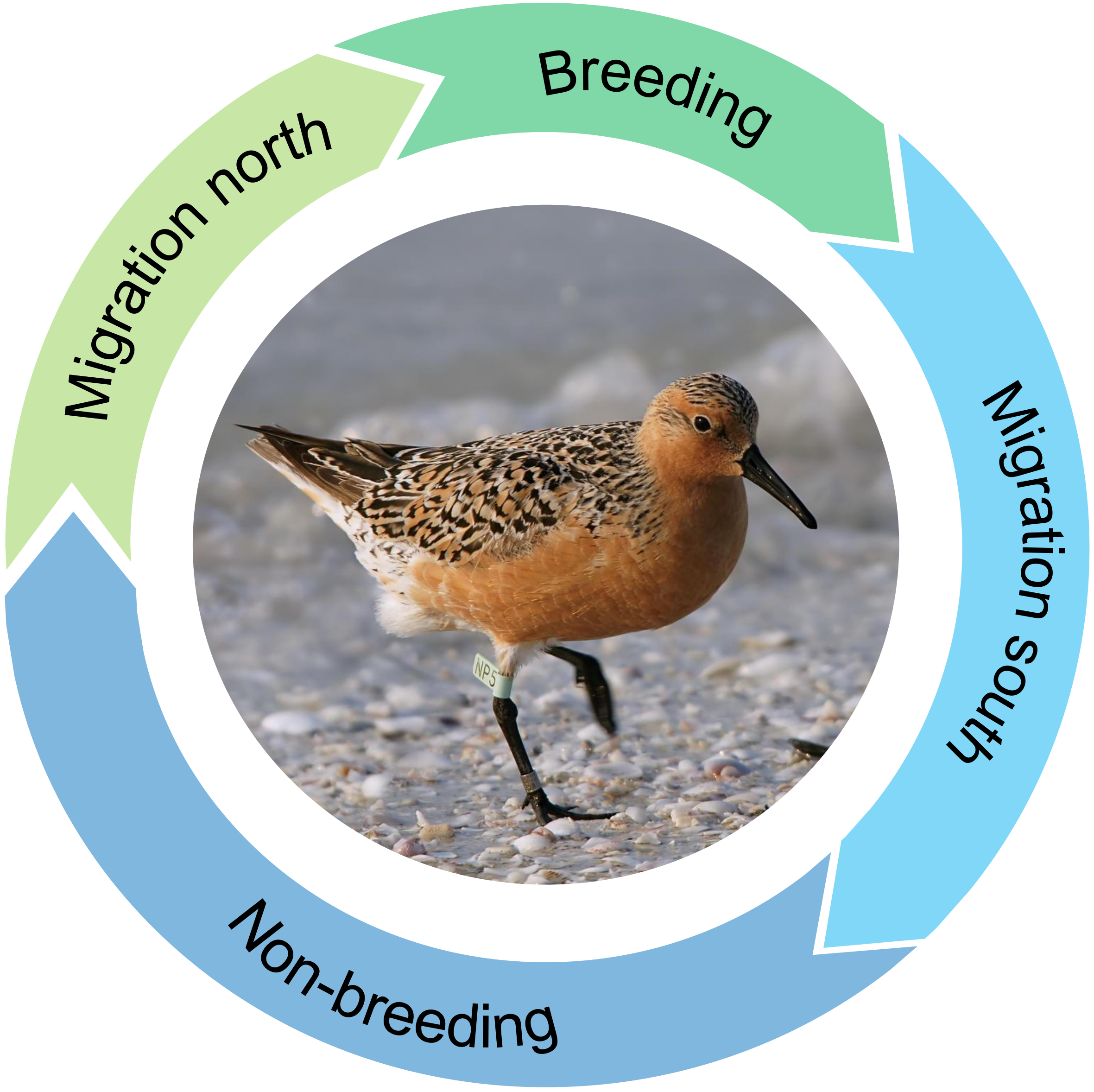
Red Knots declined in 2000s

Joe Smith



Niles et al. (2009)

Joe Smith



Breeding

Migration south

Non-breeding

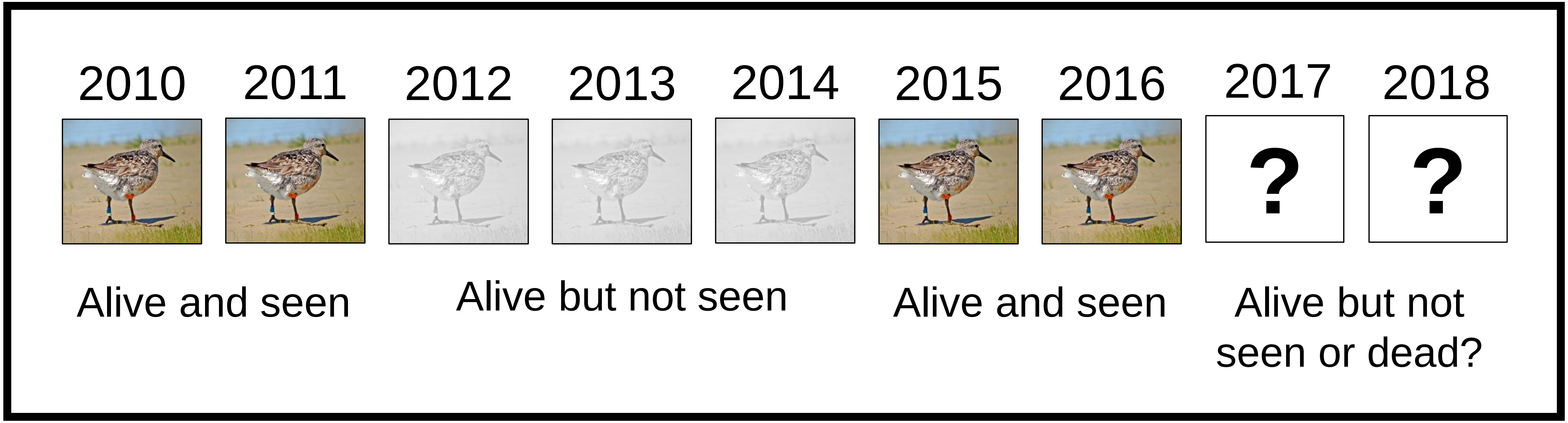
Migration north

Many *rufa* red knots are marked with coded leg flags


Yves Aubry

Jean Iron

Analyzing flag resighting data in mark-recapture models permits estimation of demographic parameters



2010 2011 2012 2013 2014 2015 2016 2017 2018

? ?

Alive and seen

Alive but not seen

Alive and seen

Alive but not seen or dead?

Bayesian analysis offers flexibility to build models to address various questions about red knot survival

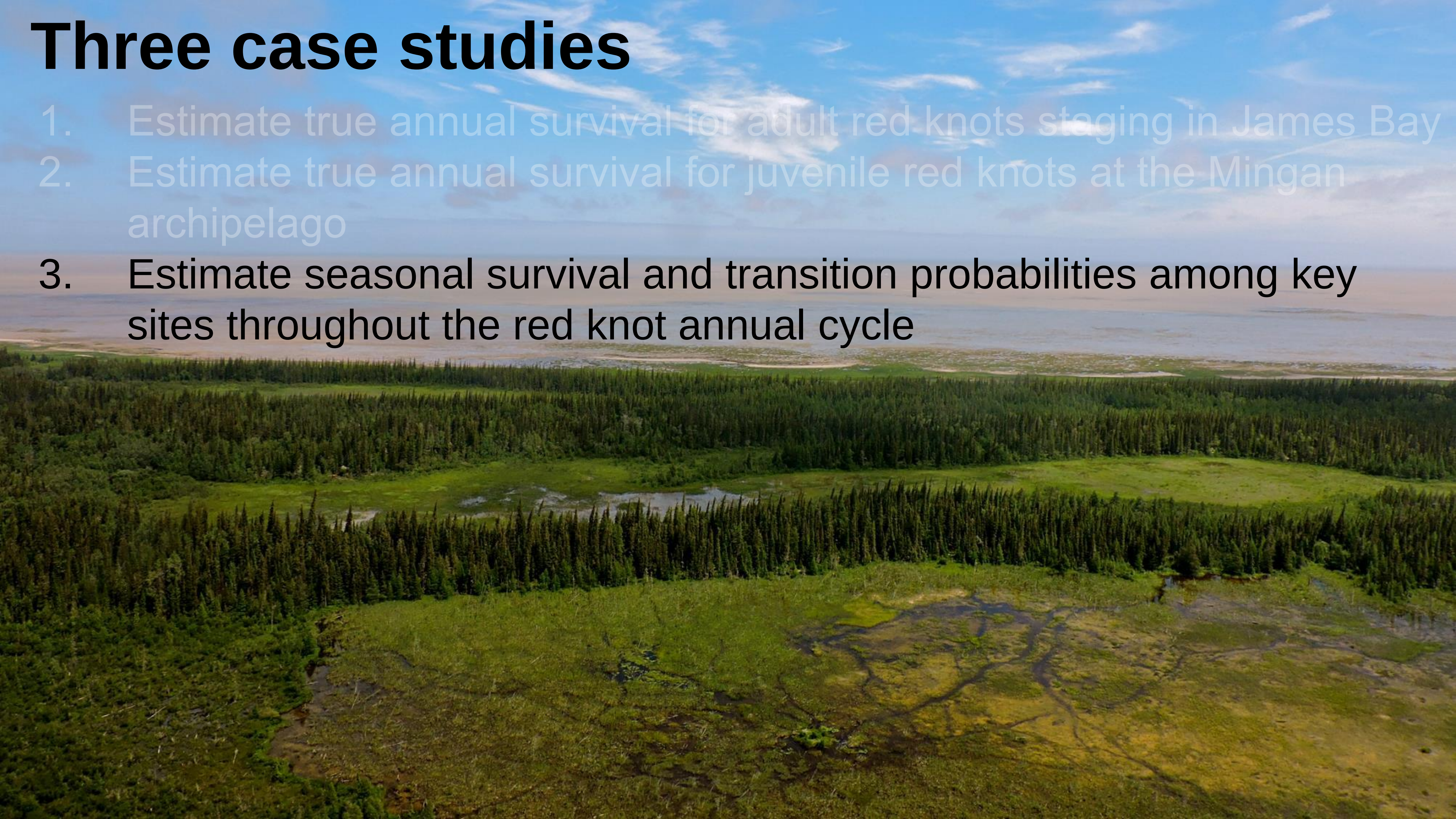
Three case studies

1. Estimate true annual survival for adult red knots staging in James Bay
2. Estimate true annual survival for juvenile red knots at the Mingan archipelago
3. Estimate seasonal survival and transition probabilities among key sites throughout the red knot annual cycle

Three case studies

1. Estimate true annual survival for adult red knots staging in James Bay
2. Estimate true annual survival for juvenile red knots at the Mingan archipelago
3. Estimate seasonal survival and transition probabilities among key sites throughout the red knot annual cycle

Three case studies



1. Estimate true annual survival for adult red knots staging in James Bay
2. Estimate true annual survival for juvenile red knots at the Mingan archipelago
3. Estimate seasonal survival and transition probabilities among key sites throughout the red knot annual cycle

Survival rates of Red Knots staging in James Bay

Resights from James Bay + Resights from outside James Bay

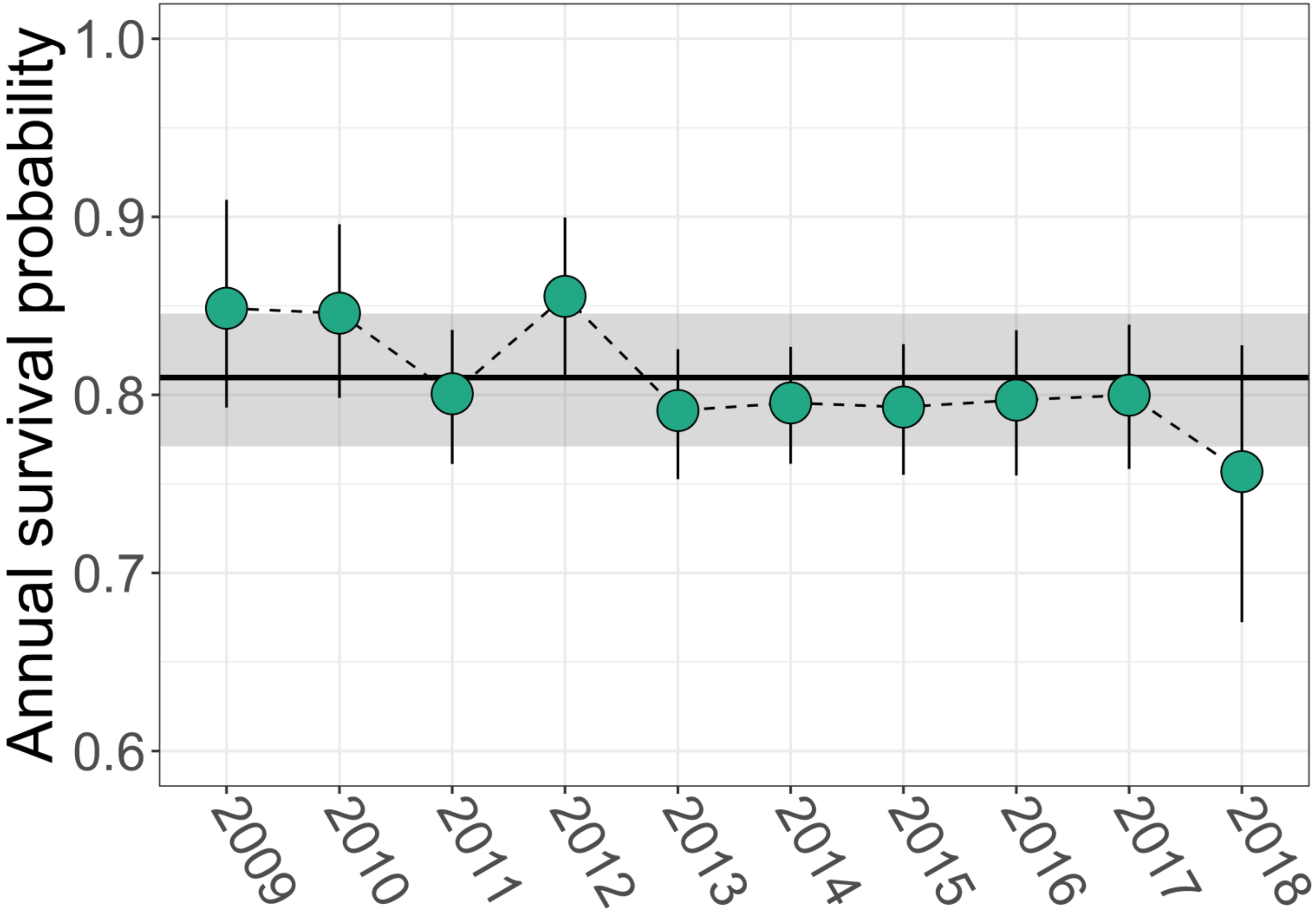
↓

Barker model

↓

True adult survival

Survival rates of Red Knots staging in James Bay



Resights from James Bay

+

Resights from outside James Bay

↓

Barker model

↓

True adult survival

=

0.810 (0.771 – 0.846)

Survival rates of juvenile Red Knots at the Mingan Archipelago

Juvenile banding records + Resights from Mingan

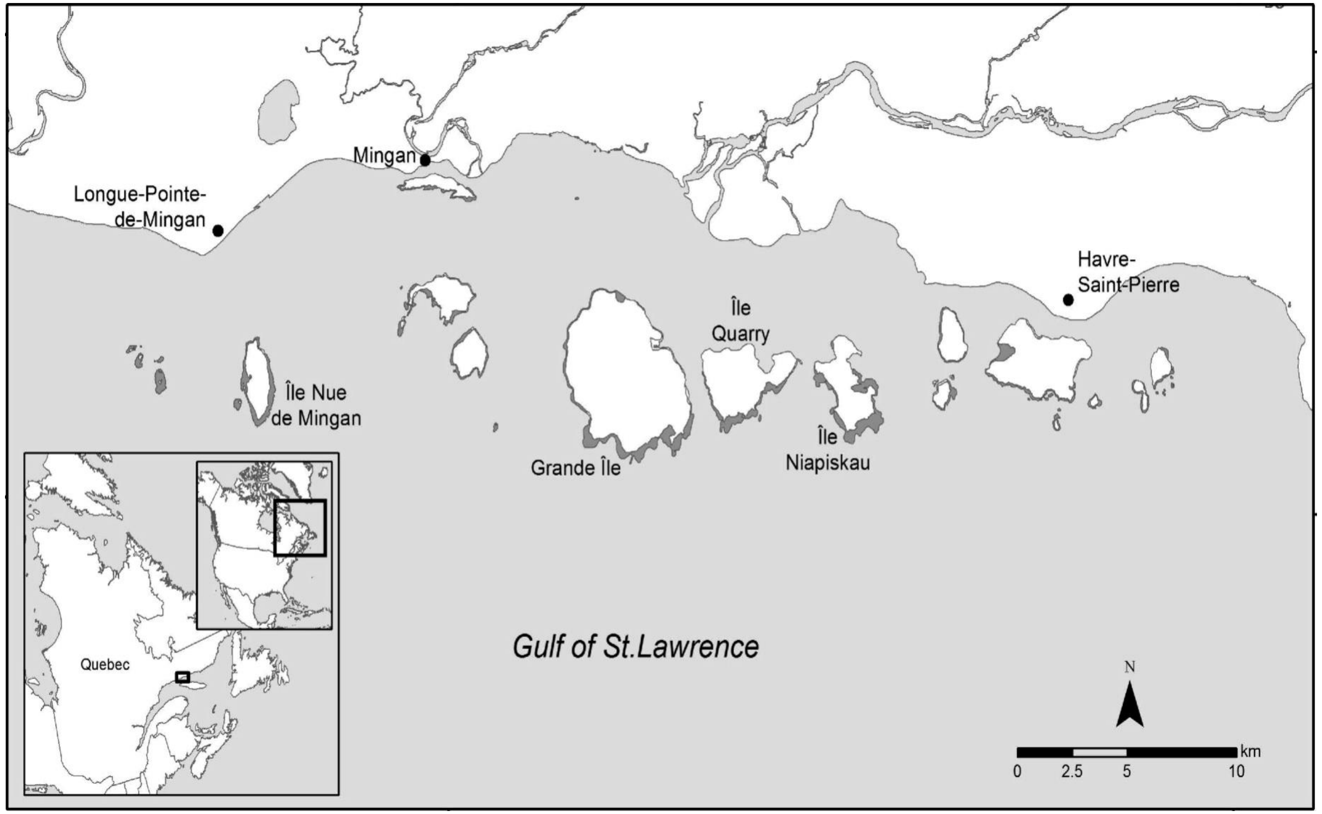
\+

Resights from outside Mingan → Barker model

↓

Juvenile survival


Ian Davies

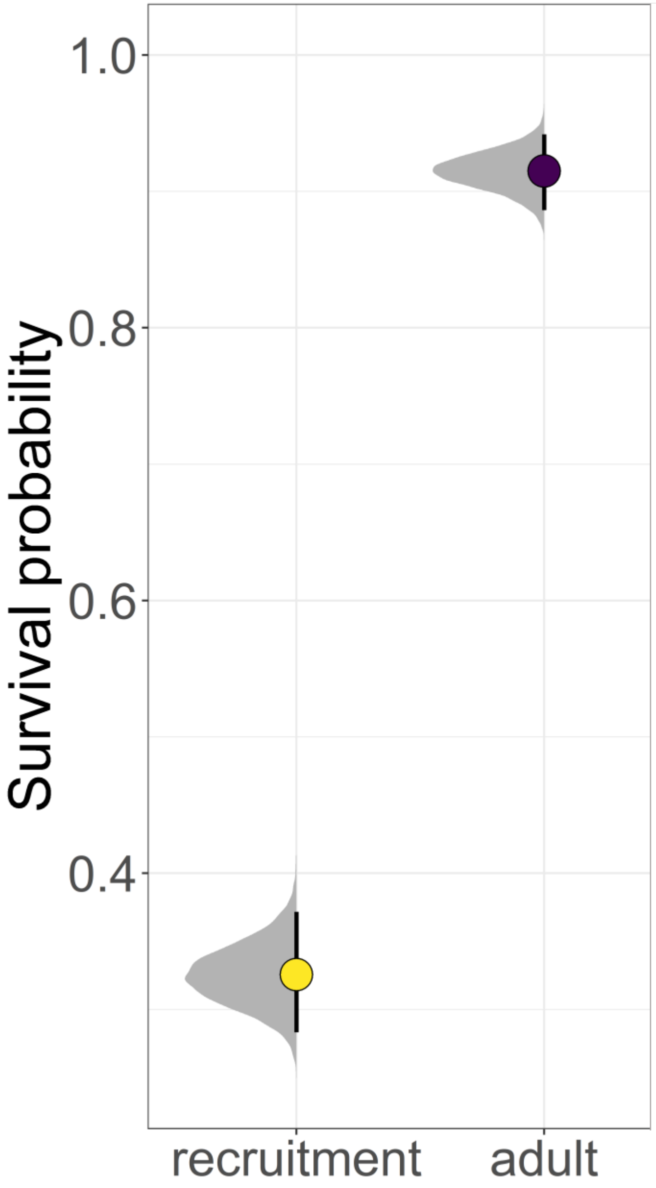


Lyons et al. (2017)

Survival rates of juvenile Red Knots at the Mingan Archipelago

Juvenile banding records + Resights from Mingan + Resights from outside Mingan → Barker model → Juvenile survival


Ian Davies



Survival rates of juvenile Red Knots at the Mingan Archipelago

Juvenile banding records + Resights from Mingan

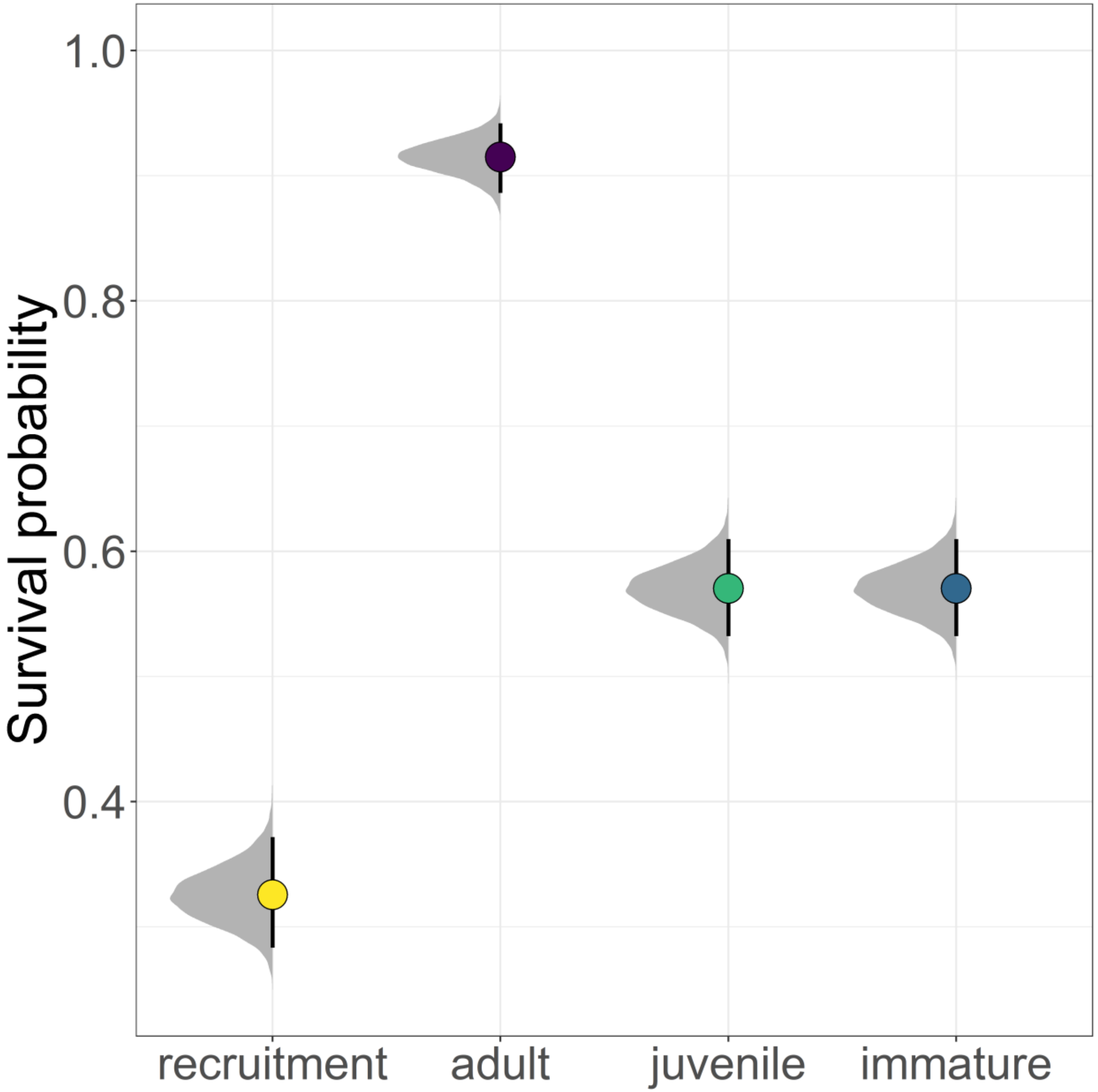
\+

Resights from outside Mingan → Barker model

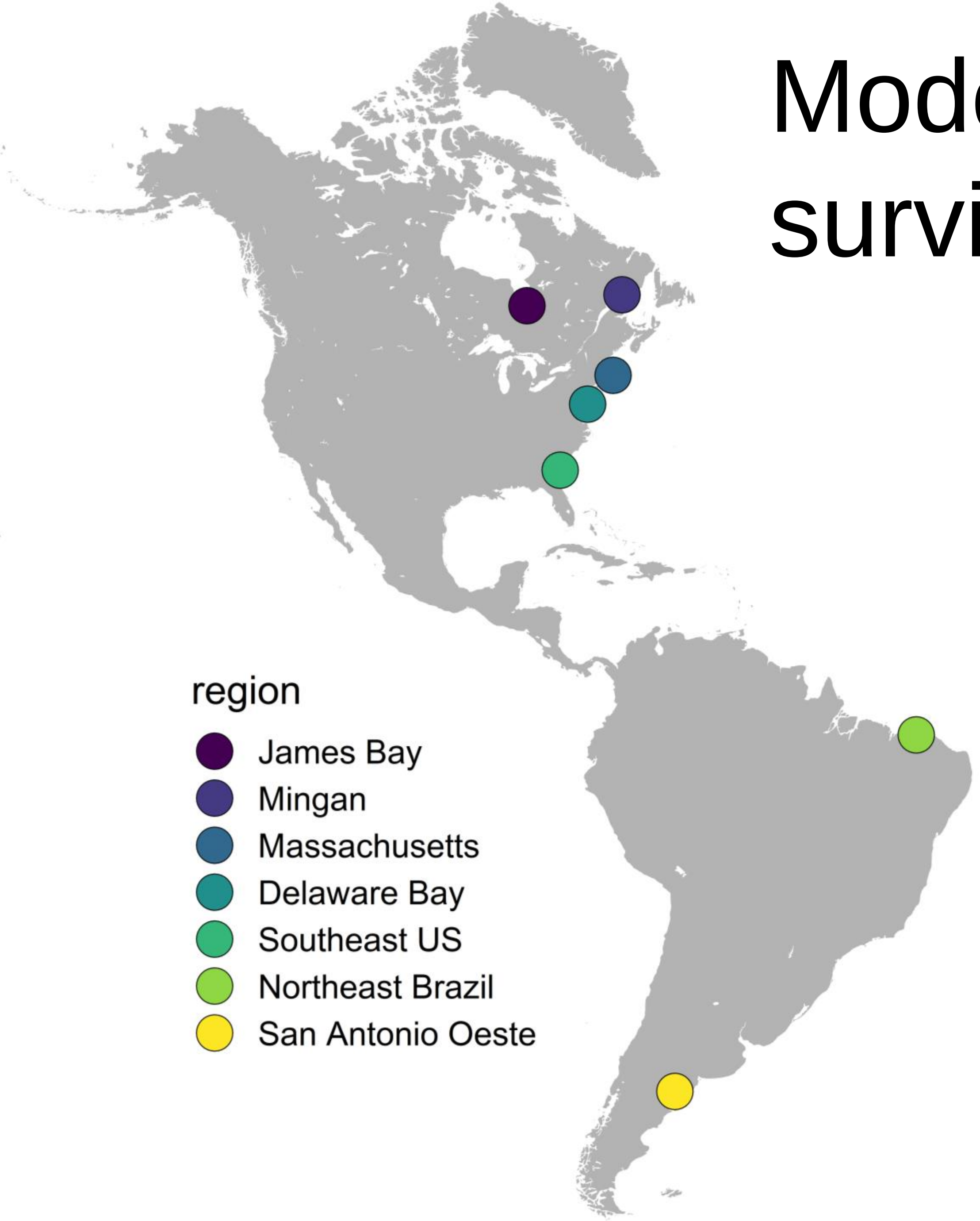
↓

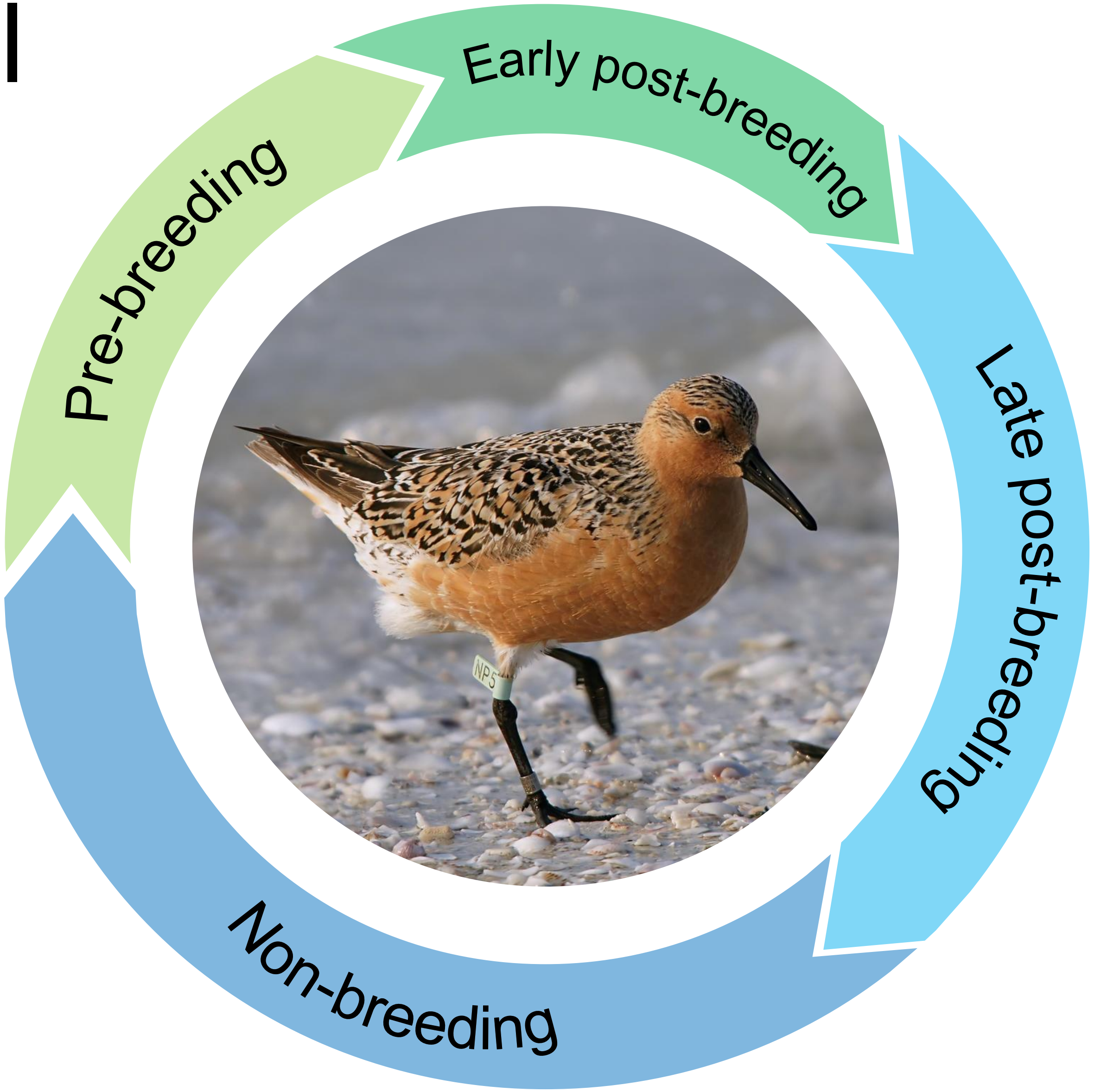
Juvenile survival


Ian Davies

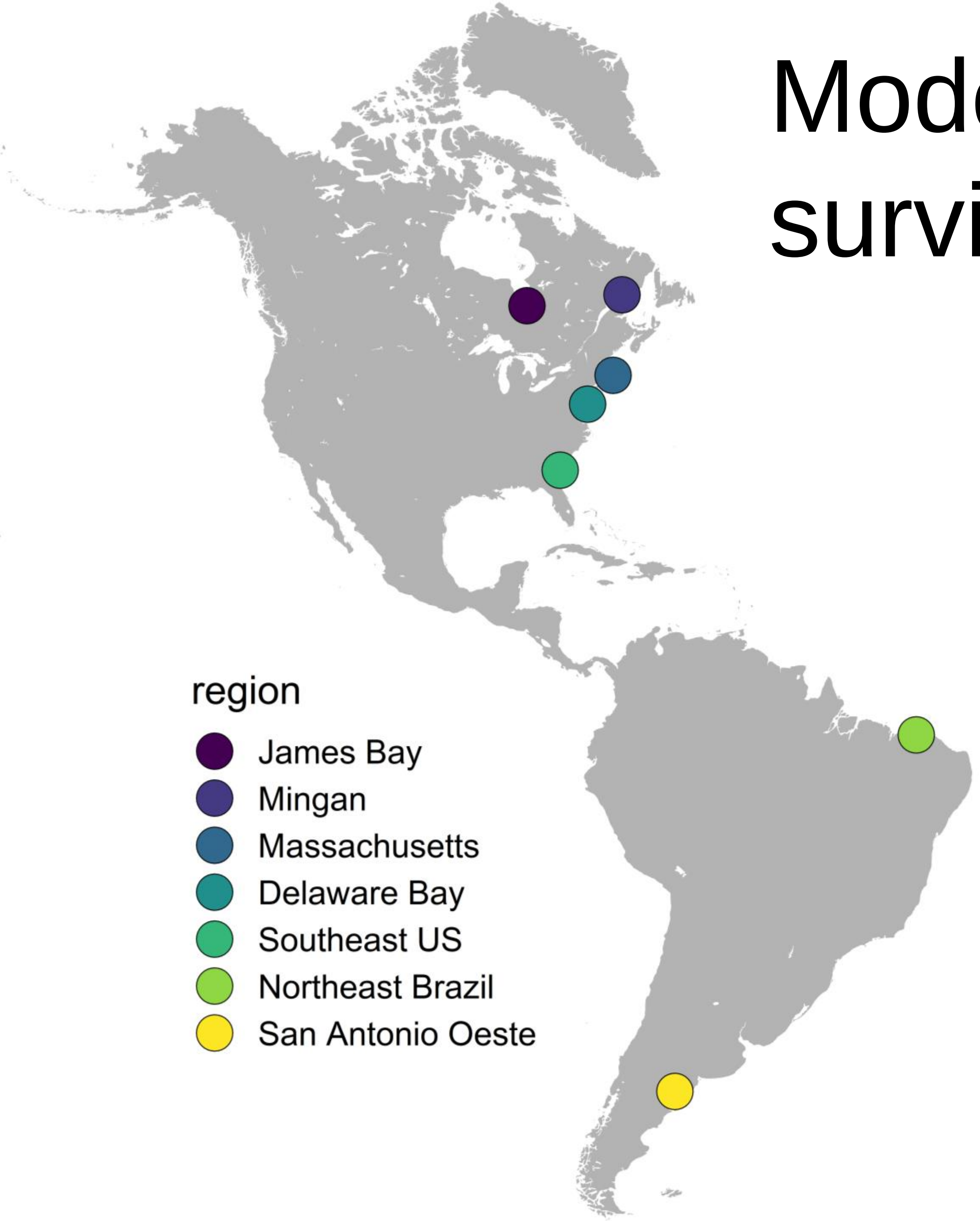


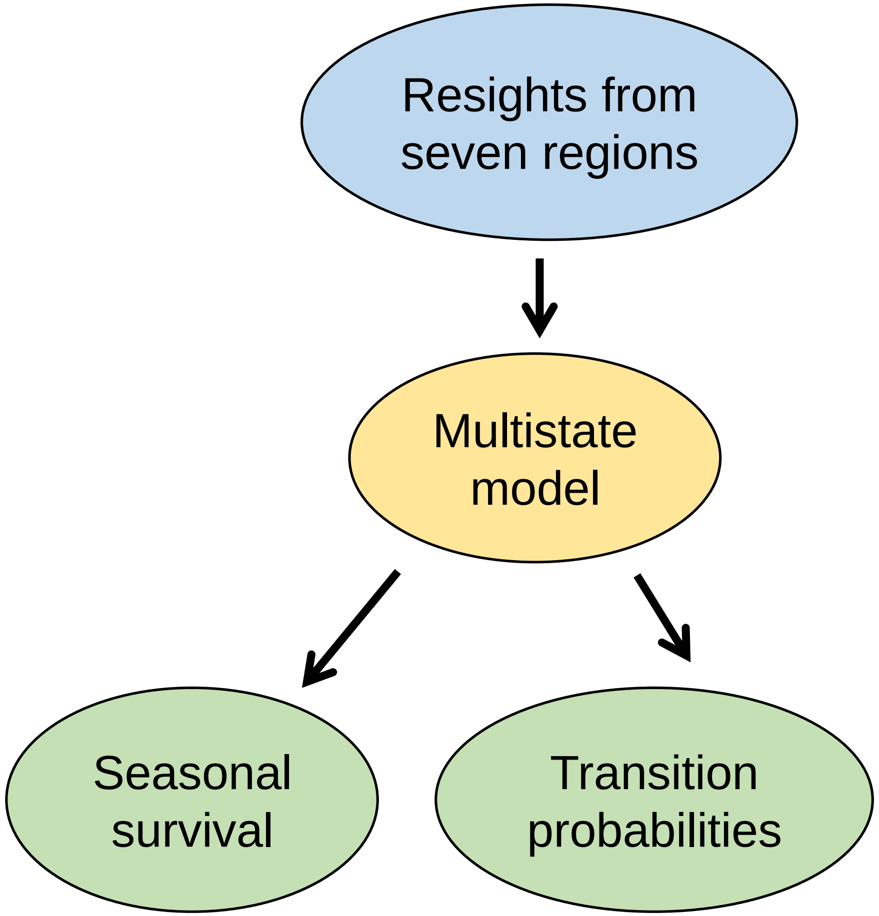
Modelling Red Knot seasonal survival





Modelling Red Knot seasonal survival





Resights from seven regions

Multistate model

Seasonal survival

Transition probabilities

In summary:

7 locations

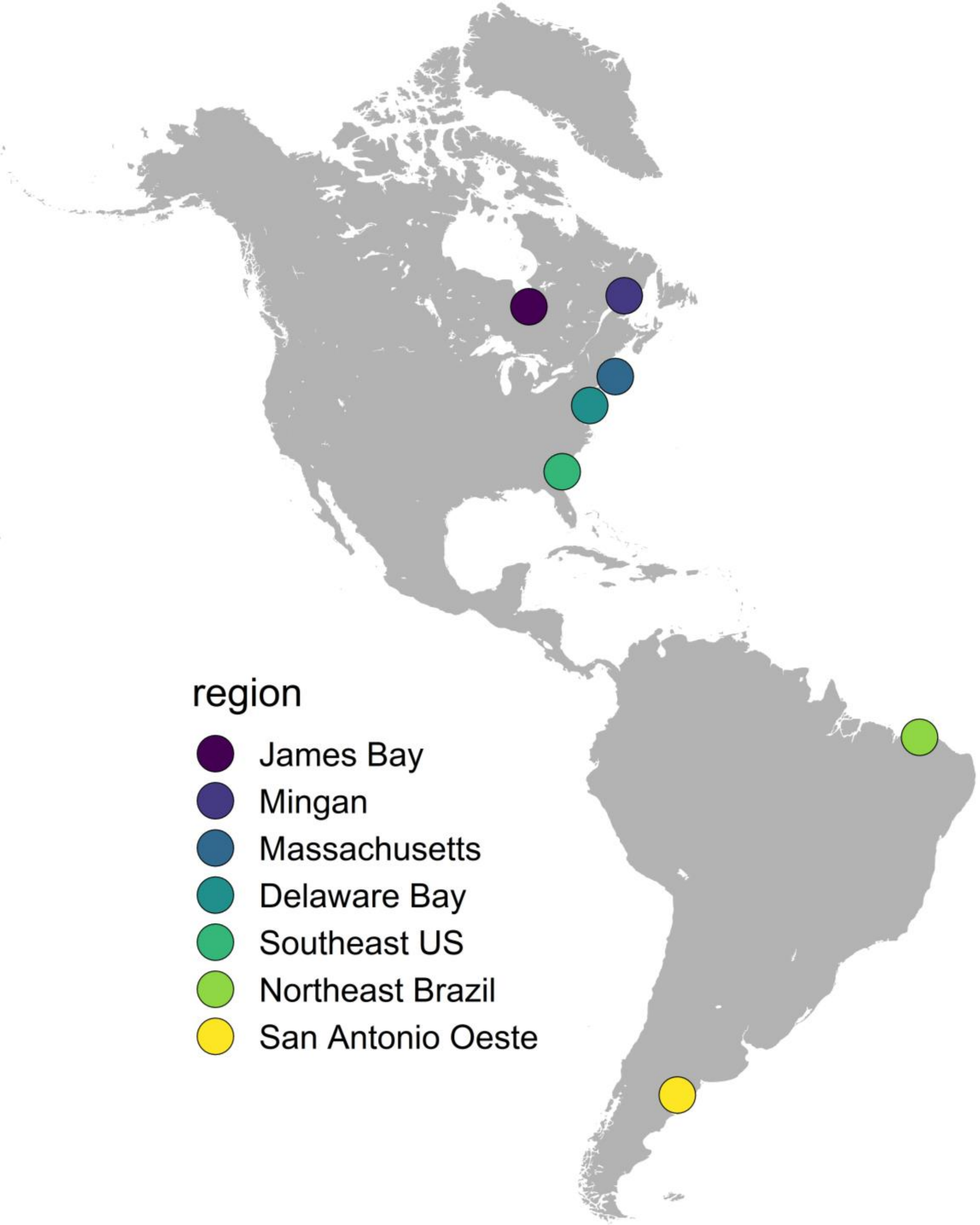
4 seasons

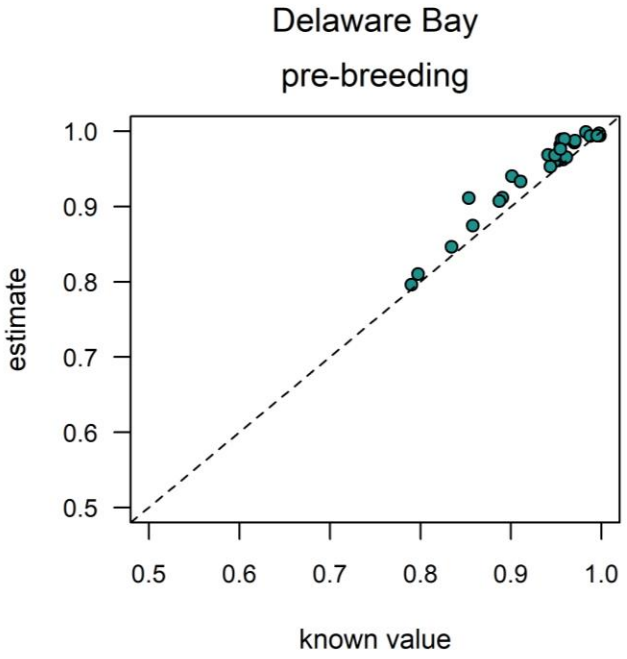
15 survival probabilities

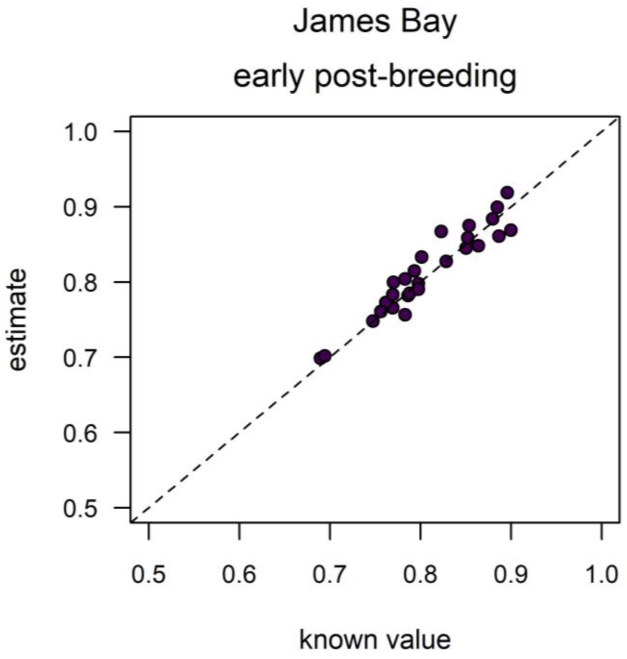
47 transition probabilities

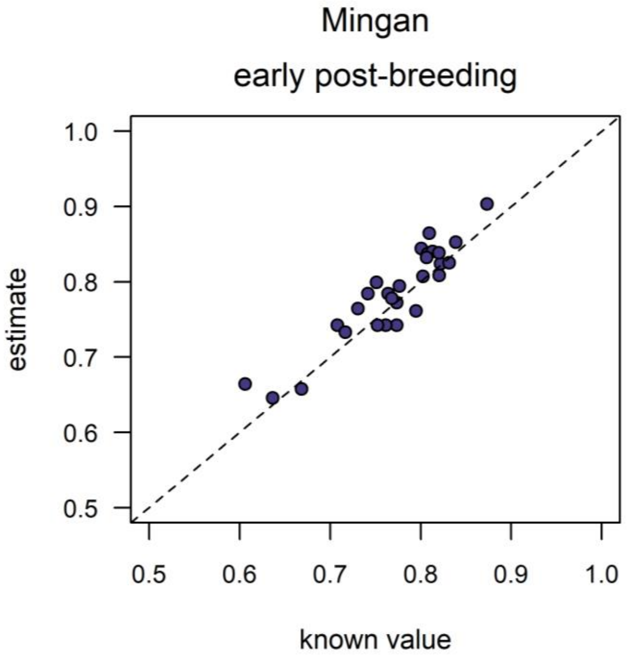
15 resighting probabilities

Karl Bardon

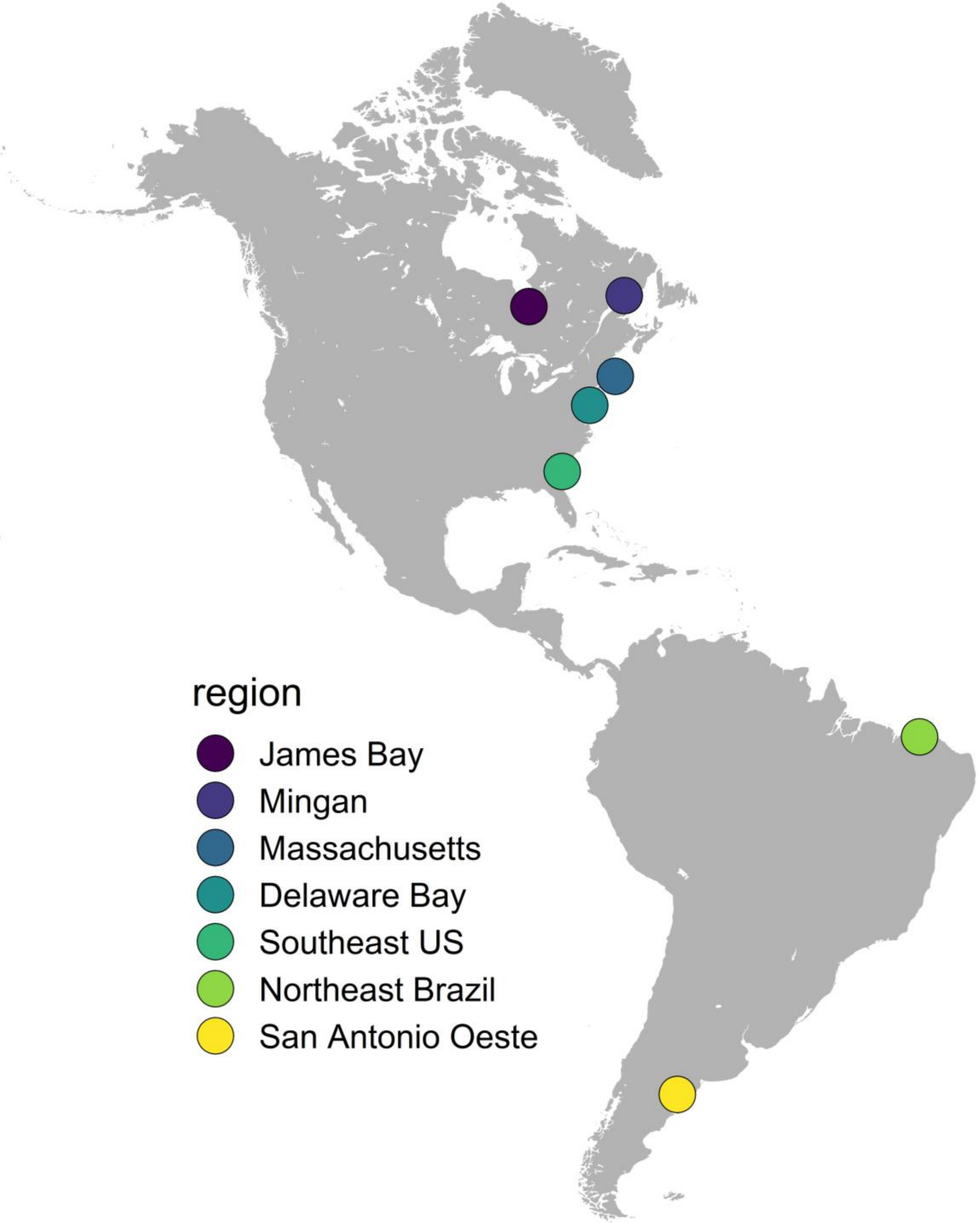


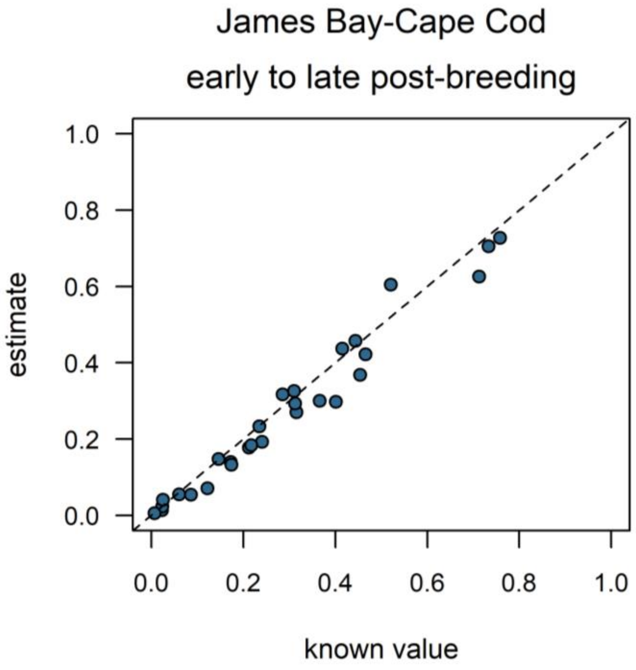


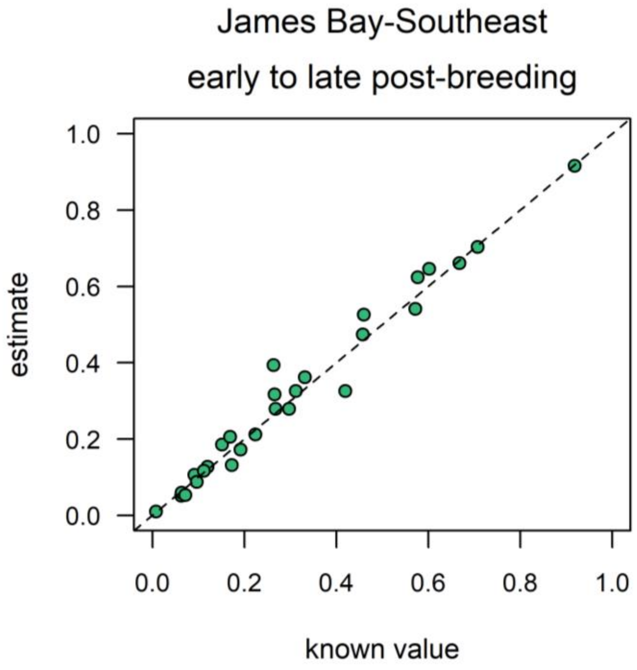


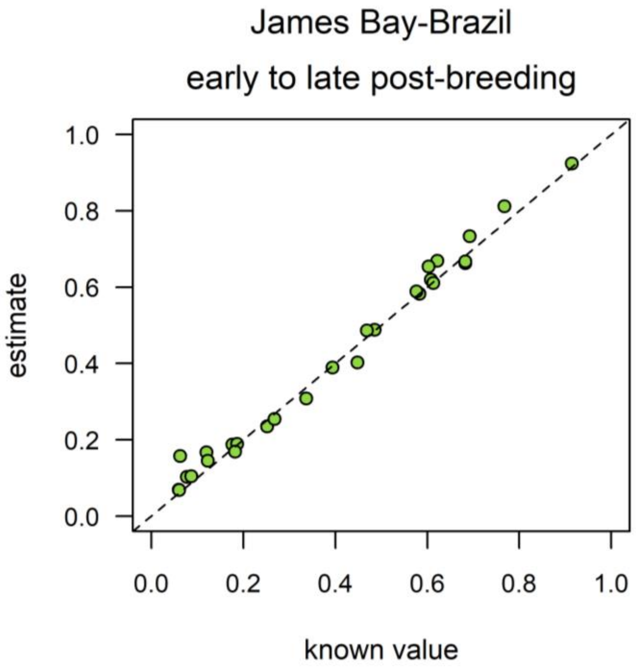


Max mean difference = **4%**




James Bay-Cape Cod
early to late post-breeding


James Bay-Southeast
early to late post-breeding

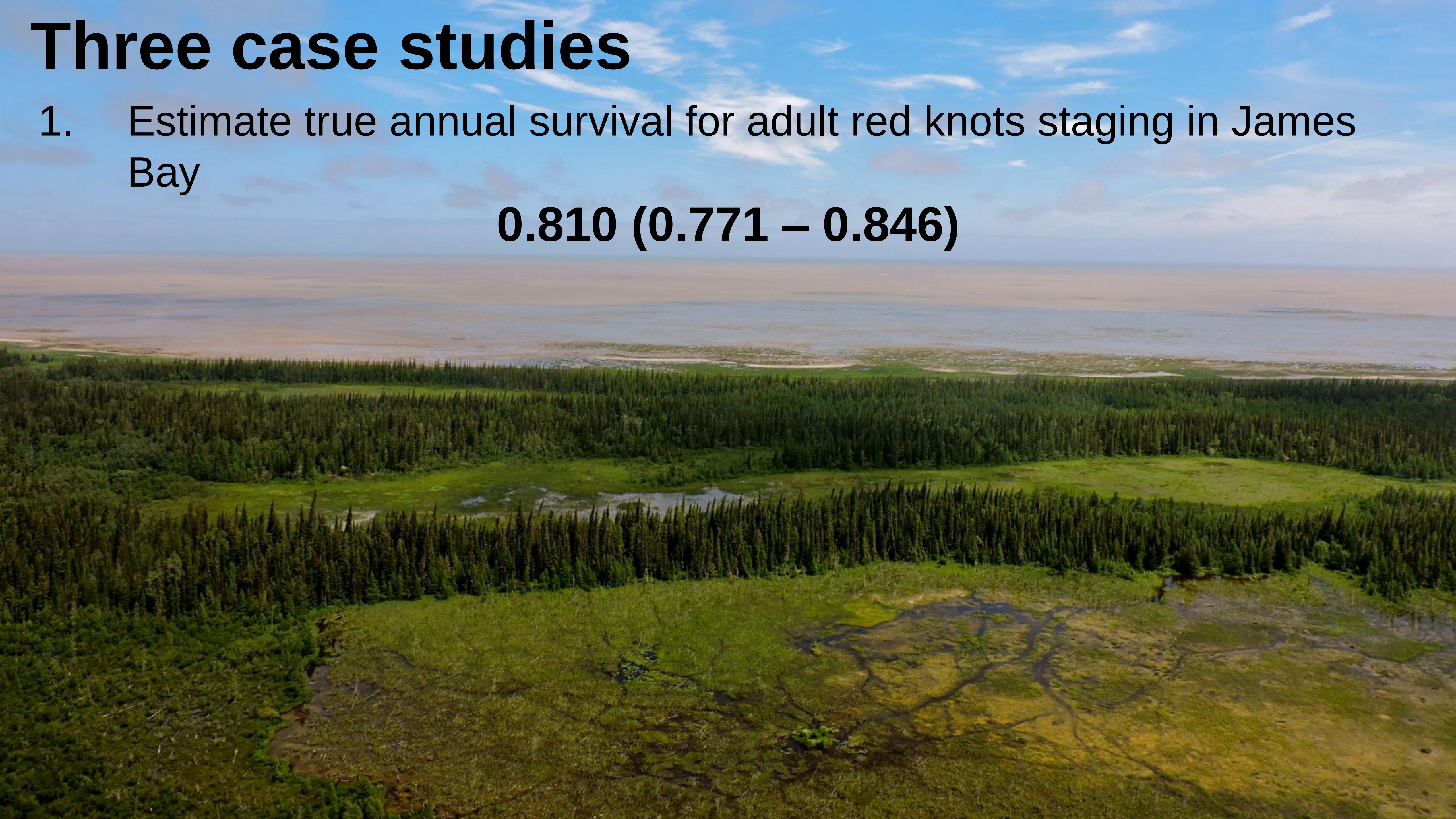

James Bay-Brazil
early to late post-breeding

Max mean difference = **6%**

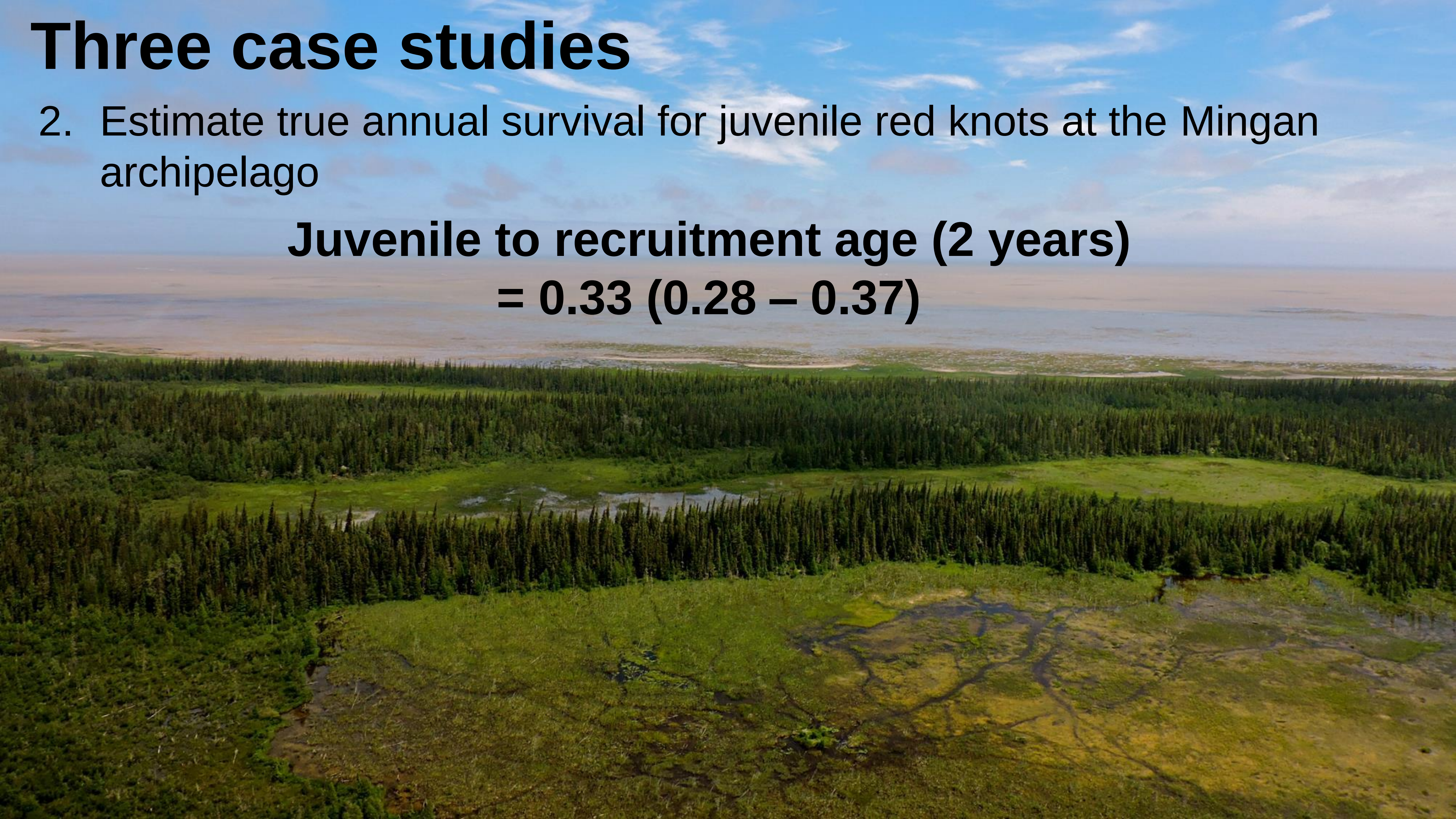
Three case studies

1. Estimate true annual survival for adult red knots staging in James Bay

0.810 (0.771 – 0.846)



Three case studies



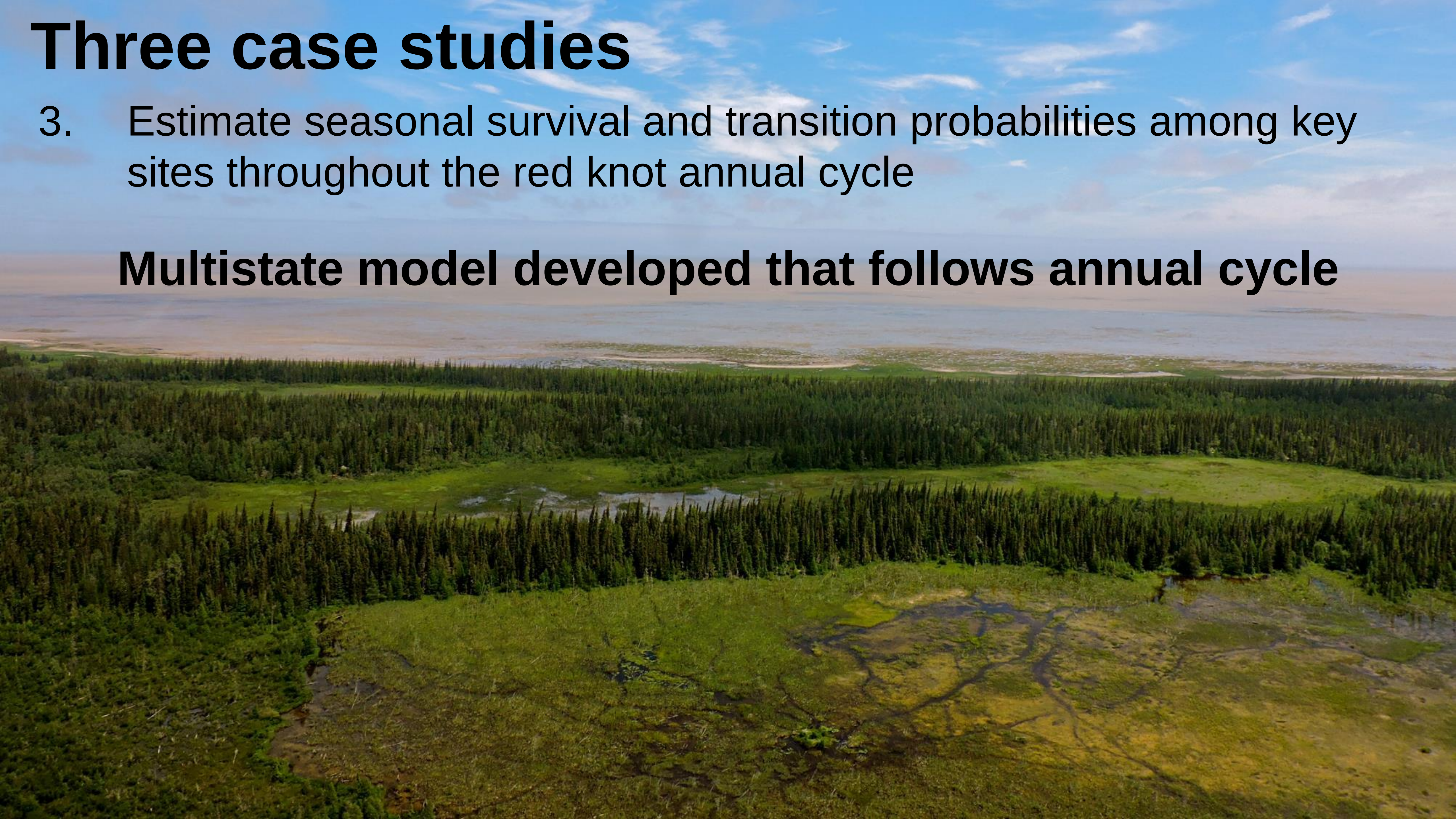
2. Estimate true annual survival for juvenile red knots at the Mingan archipelago

Juvenile to recruitment age (2 years)
= 0.33 (0.28 – 0.37)

Three case studies

3. Estimate seasonal survival and transition probabilities among key sites throughout the red knot annual cycle

Multistate model developed that follows annual cycle



Acknowledgements

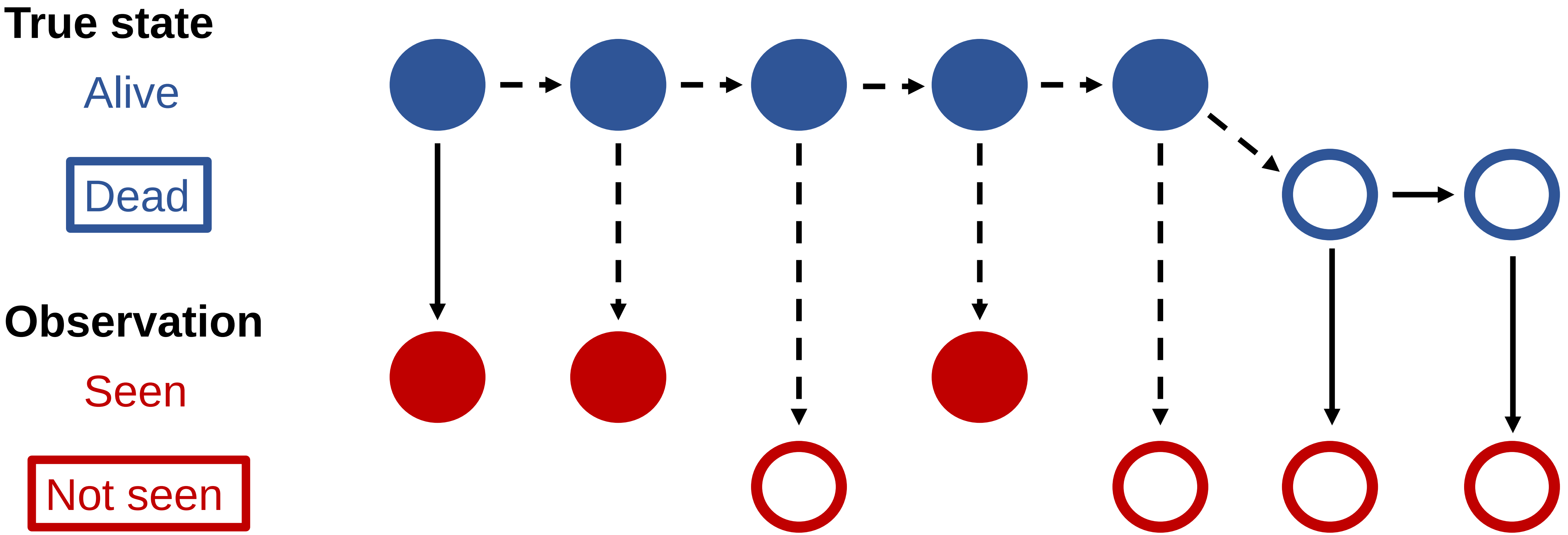
Lena Usyk (bandedbirds.org)
Jim Hines
Many, many red knot banders and surveyors
All collaborators supporting various red knot banding and resighting programs
Joseph Smith

Thank you!

¡Gracias!

amacdonald@birdscanada.org

Analyzing flag resighting data in mark-recapture models permits estimation of demographic parameters



True state

Alive

Dead

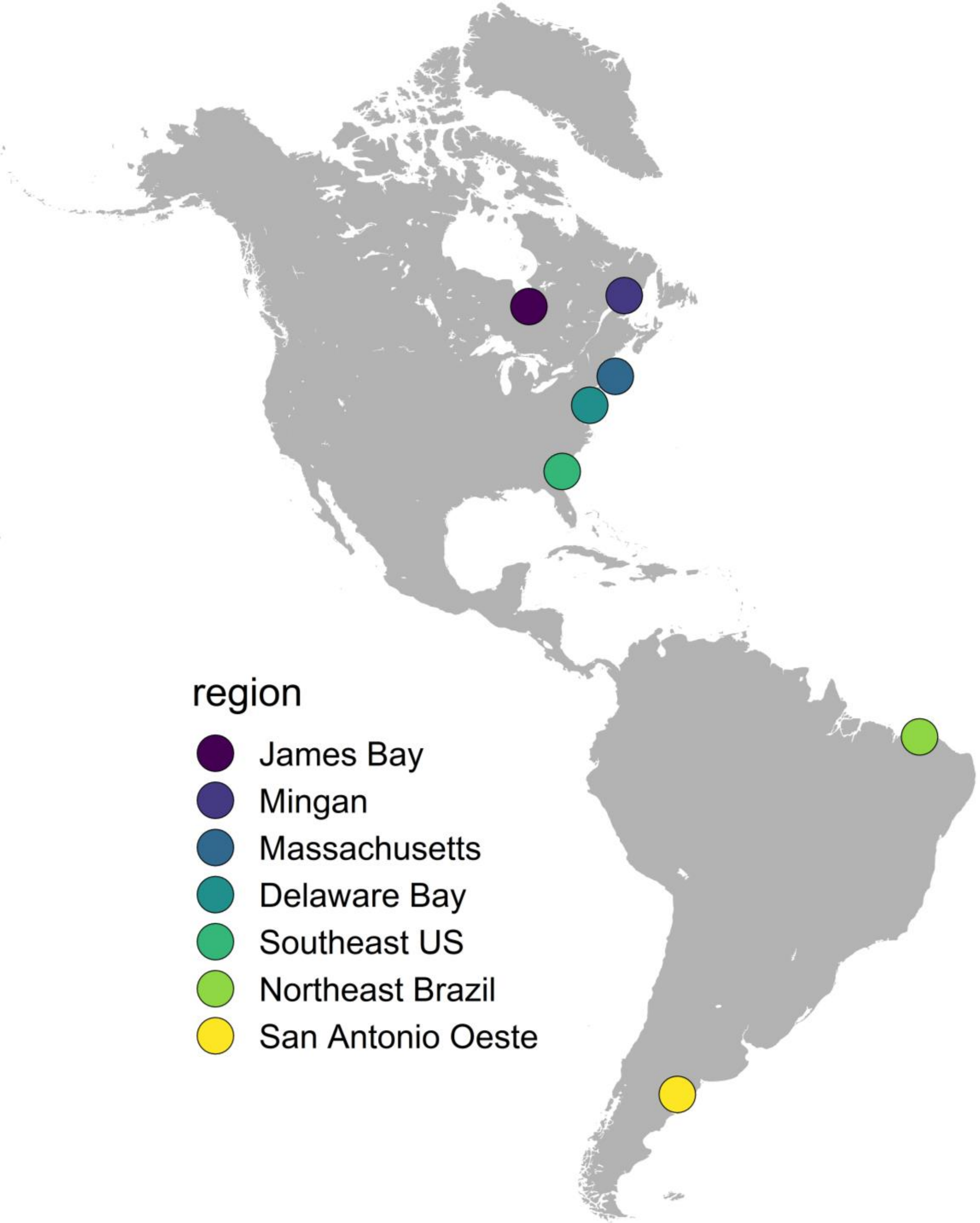
Observation

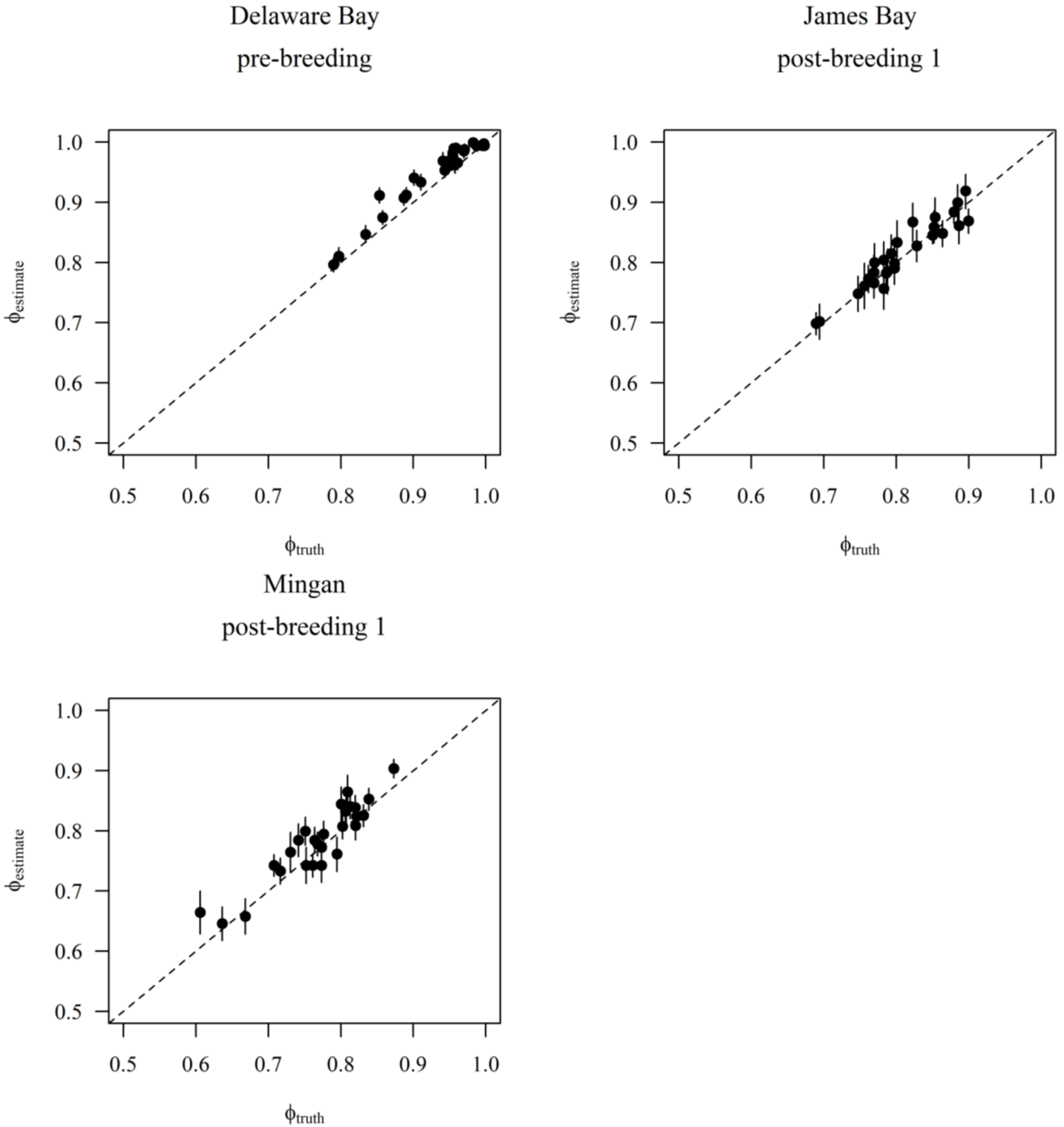
Seen

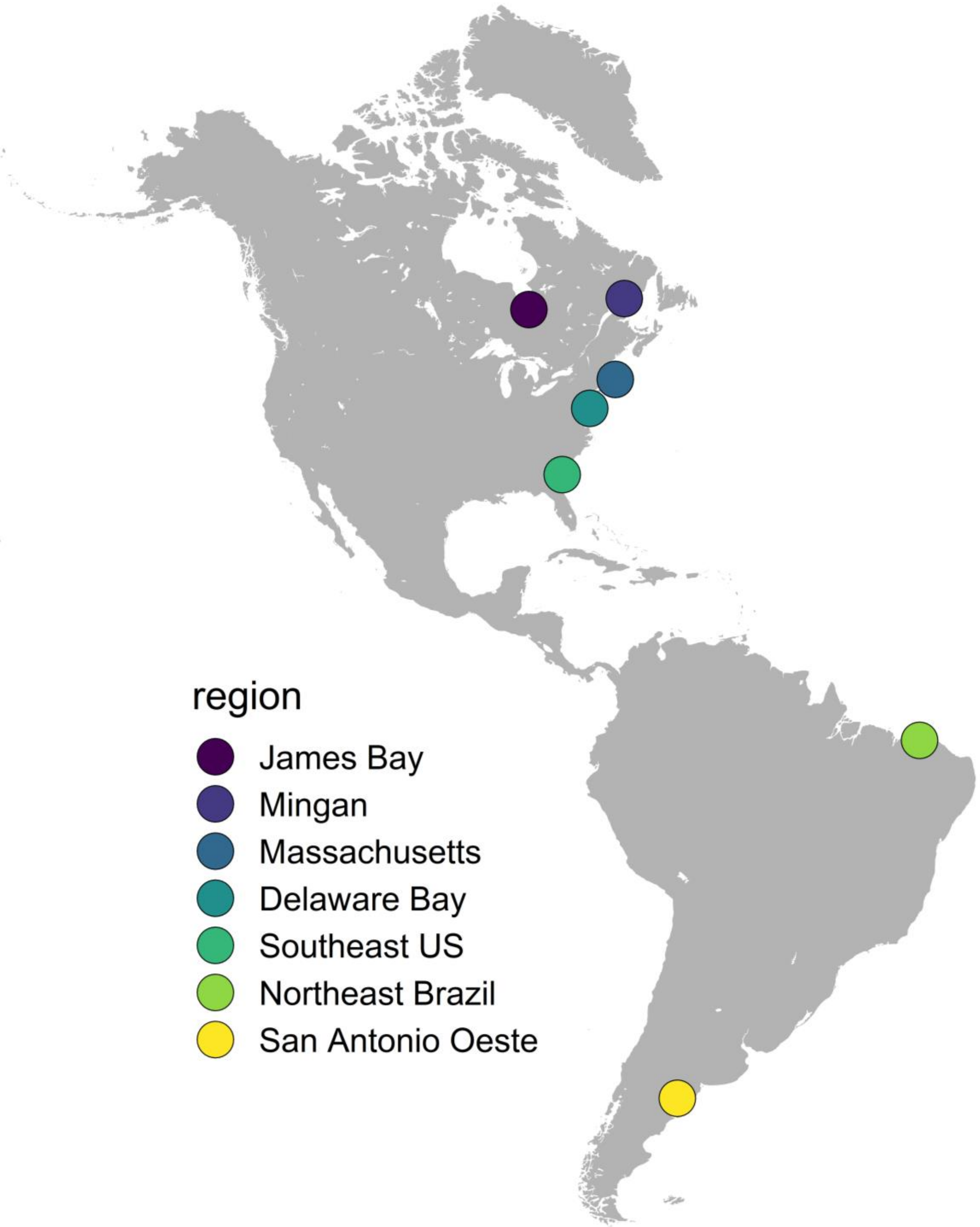
Not seen

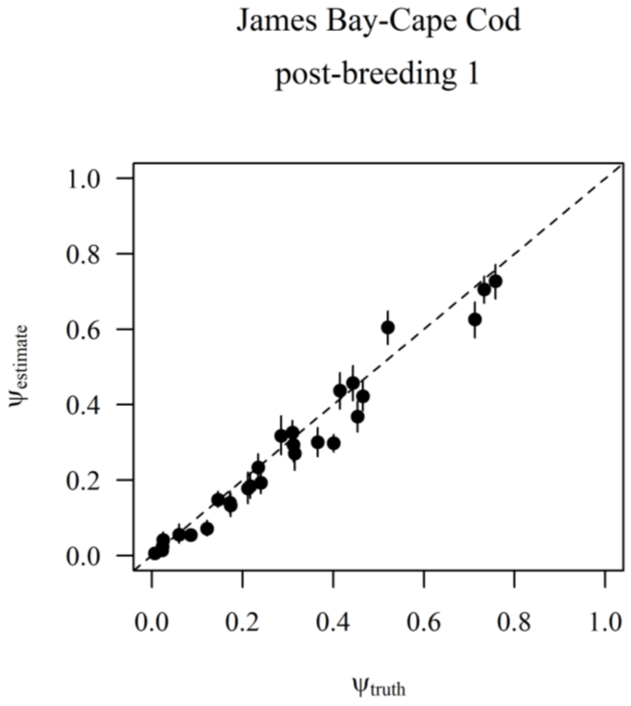
Kéry and Schaub (2012)

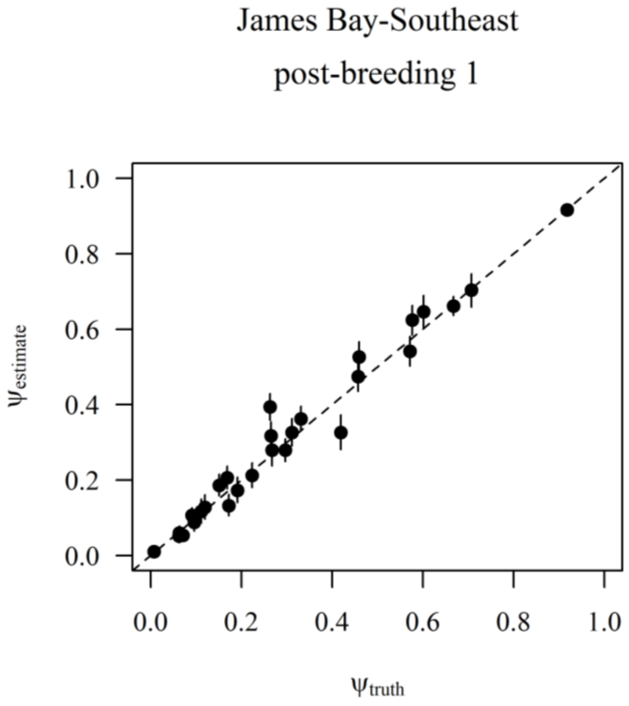
Bayesian analysis offers flexibility to build models to address various questions about red knot survival

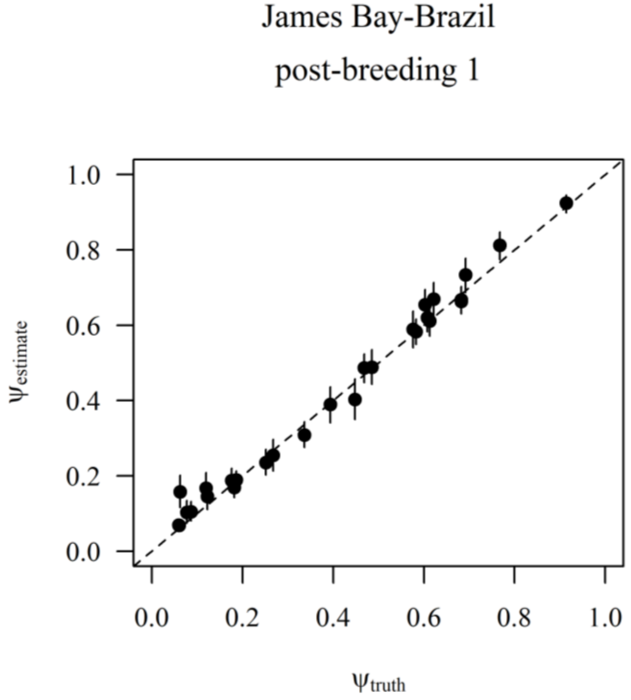

region
James Bay
Mingan
Massachusetts
Delaware Bay
Southeast US
Northeast Brazil
San Antonio Oeste


Delaware Bay
pre-breeding
James Bay
post-breeding 1
Mingan
post-breeding 1
$\phi_{estimate}$
ϕ_{truth}
0.5
0.6
0.7
0.8
0.9
1.0




James Bay-Cape Cod
post-breeding 1


James Bay-Southeast
post-breeding 1


James Bay-Brazil
post-breeding 1