EARTHJUSTICE

September 25, 2023

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: Use of the Adaptive Resource Management Model to Recommend Horseshoe
        Crab Bait Harvest Quotas

Dear Members of the Horseshoe Crab Management Board:

New Jersey Audubon and Defenders of Wildlife urge the Horseshoe Crab Management Board
("Board") to maintain the prohibition on harvesting female Delaware Bay-origin horseshoe crabs
for bait. These comments present extensive new technical analysis by Dr. Kevin Shoemaker
demonstrating that the adaptive resource management ("ARM") model[1] does not accurately
represent the impact of the horseshoe crab harvest on red knots or horseshoe crabs. As a result of
the ARM model's flaws—many of which are intrinsic to its core structure and functionality—
utilizing the model to inform management decisions will not safeguard against "limiting the red
knot stopover population or slowing recovery"[2] or violating the Endangered Species Act
("ESA"). These comments also explain why the Board must not use the ongoing stakeholder
survey to initiate a resumption of the female harvest.

The ARM model ostensibly represents the connection between horseshoe crabs and red knot
shorebirds (*Calidris canutus rufa*). Each year, red knots fly from as far south as Tierra del Fuego
to the Arctic Circle, where they breed—a round trip that can span 19,000 miles. At a critical
point in their northbound migration, after depleting much of their energy, most red knots stop at
Delaware Bay as horseshoe crabs emerge from the water to spawn on the beach. By feasting on a
superabundance of horseshoe crab eggs, red knots can double their body weight in under two

---

[1] Unless otherwise stated, in these comments, the "ARM model" refers to the version approved by the Horseshoe
Crab Management Board in 2022.
[2] 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)* 25 (2021) ("ARM Report") (providing the objective
statement for the ARM Framework).

weeks.[3] With their energy restored, red knots have an improved likelihood of completing their migration and breeding successfully.[4]

In the late twentieth century, horseshoe crabs were severely overharvested. As their numbers fell, eggs on the beach grew scarcer, with devastating impacts on red knots. In 2015, the U.S. Fish and Wildlife Service ("USFWS") listed red knots as threatened under the ESA, citing "[r]educed food availability in Delaware Bay due to commercial harvest of the horseshoe crab . . . [as] a primary causal factor in red knot population declines in the 2000s."[5] The Atlantic States Marine Fisheries Commission ("ASMFC"), through its Horseshoe Crab Management Board, has regulated the harvest of horseshoe crabs for use as bait since 1998, but both red knots and horseshoe crabs remain perilously depleted compared to historical levels. Last year, the Board approved the use of a revised ARM model to process data about horseshoe crab and red knot demographics and recommend horseshoe crab bait harvest quotas.[6] The ARM framework's objectives include "ensur[ing] that the abundance of horseshoe crabs is not limiting the red knot stopover population or slowing recovery."[7]

In advance of the Board's decision to approve the revised ARM model, New Jersey Audubon and Defenders of Wildlife submitted analysis by independent consultant Dr. Kevin Shoemaker demonstrating that the model falls far short of that objective. Among other deficiencies that Dr. Shoemaker identified, the model recognizes scarcely any correlation between the abundance of horseshoe crabs and red knots. Despite the historical role of horseshoe crab overharvest in the decline of red knots, the model predicts red knot abundance will increase even if all horseshoe crabs vanish entirely from Delaware Bay. This deficiency heavily influences the harvest quotas that the model recommends. While the previous model never recommended allowing a bait harvest of female Delaware Bay-origin horseshoe crabs, the revised model is nearly certain to recommend a significant female harvest every year. Citing public concern, the Board maintained a male-only bait harvest for 2023. That is, the Board approved the ARM model but did not immediately adopt its recommended harvest quotas.[8]

The numerous flaws in the ARM model that Dr. Shoemaker previously identified thoroughly demonstrated that the model is unfit for recommending horseshoe crab harvest levels. That conclusion was evident even though Dr. Shoemaker could analyze only one component of the model because the rest was being withheld from public review. New Jersey Audubon and Defenders of Wildlife cautioned that additional flaws may emerge when the public gained access to the entire model and urged the Board to abstain from voting on the model until that time. The

---

[3] New Jersey Department of Environmental Protection, *Wildlife Populations: Red Knot* 1-2 (2020), https://dep.nj.gov/wp-content/uploads/dsr/trends-red-knot.pdf.

[4] Sjoerd Duijns et al., *Body Condition Explains Migratory Performance of a Long-Distance Migrant*, 284 Proceedings of the Royal Society of London B 20171374, at 4-6 (2017).

[5] FWS, "Endangered and Threatened Wildlife and Plants; Final Threatened Status for the Rufa Red Knot," 79 Fed. Reg. 73,706, 73,707 (Dec. 11, 2014).

[6] ASMFC, Press Release, "Horseshoe Crab Board Sets 2023 Specifications for Horseshoe Crabs of Delaware Bay-Origin & Adopts ARM Framework Revision via Addendum VIII" (Nov. 10, 2022), http://www.asmfc.org/uploads/file/636d41cepr33_HSC2023DEBaySpecs_AddendumVIII_Approval.pdf.

[7] *See* ARM Report 25.

[8] See ASMFC, *supra* note 6.

entire model was finally released the evening before the Board approved it, and Dr. Shoemaker has now performed a comprehensive review.

Dr. Shoemaker's new analysis paints an even starker picture of the ARM model's unsuitability for managing the horseshoe crab bait harvest. Collectively, his two analyses make abundantly clear that the ARM model does not accurately represent the relationship between horseshoe crabs and red knots or the population status and trajectory of either species individually. As a result, the model cannot anticipate the consequences of its own harvest recommendations. Implementing the model's recommendations—especially its recommendation to resume a female horseshoe crab bait harvest—would place red knots at extraordinary risk and potentially destabilize the horseshoe crab population as well.

While red knots face a variety of threats, including beach development and climate change, the availability of horseshoe crab eggs is a key determinant of their survival and reproductive success. The Board cannot use the existence of other threats to deflect its responsibility to ensure that horseshoe crab levels do not limit the Delaware Bay stopover population or slow the recovery of red knots. To the contrary, the existence of other threats should impel the Board to exercise more precaution when setting harvest quotas.

These comments present Dr. Shoemaker's analysis and other material to make four principal points, all of which support the overarching conclusion that the Board cannot defensibly use the ARM model to set bait harvest quotas for Delaware Bay-origin horseshoe crabs:

1. **The availability of horseshoe crab eggs on the beach, not trawl survey data, is the most direct and meaningful determinant of red knot survival.**

   - The ARM model entirely ignores the most important source of data—the number of horseshoe crab eggs per square meter of beach (referred to as egg "density"). Egg density is the most direct measure of whether there are enough horseshoe crabs to fulfill the nutritional needs of red knots. Dr. Shoemaker shows that *egg density is strongly correlated with red knot survival*.
   - The ARM model's cornerstone is the relationship between two factors that bear virtually no relation: female horseshoe crab abundance data derived from trawl surveys and red knot abundance. The absence of a meaningful correlation between these data likely results from the difficulty of collecting and evaluating horseshoe crab abundance data using trawl surveys. It does not indicate that no significant correlation exists between the two species. But the ARM model mistakenly concludes that red knot population trajectories are not strongly related to horseshoe crab populations and thus that increasing the horseshoe crab harvest would scarcely impact red knots, even as it ignores egg density data that strongly show the opposite.
   - By failing to recognize the dependence of red knots on horseshoe crabs, the ARM model predicts the abundance of red knots will increase even if all horseshoe crabs suddenly disappear from Delaware Bay. By contrast, the correlation between egg density and red knot survival reveals a grave threat: if horseshoe

3

crab egg density stagnated at the lowest recently observed level (to say nothing of entirely disappearing), red knots would quickly plummet to near-zero levels.

2. **The ARM model overestimates and misrepresents the health and resilience of red knots and horseshoe crabs at Delaware Bay.**

- In order to serve as a legitimate basis for managing the ecosystem, the ARM model would need to accurately characterize the demographics of red knots and horseshoe crabs. In many key respects, the model misrepresents these demographics. As a result, its recommended harvest quotas are largely untethered from the actual condition of red knots and horseshoe crabs and would have dangerous impacts that the model cannot predict.

Red Knots

- The ARM model inaccurately concludes that the red knot lifespan is roughly three times what the data show (15 years instead of 5 years). Thus, the model assumes that red knots have many more breeding opportunities than they actually do. The model seriously underestimates the impact that one or two poor breeding years—due to a scarcity of horseshoe crab eggs, for example—can have on lifetime reproductive success and, by extension, the persistence of the species.
- When estimating red knot abundance, the ARM model draws a large number of conclusions from a very small dataset of population counts. This causes the model to falsely detect trends in the data even when no trends are present. Dr. Shoemaker tested the model with 50 sets of random, white-noise data that lacked any trend; the model spuriously detected a non-negligible trend in red knot abundance more than 80% of the time.
- The component of the model that estimates the red knot population fails standard "goodness-of-fit" tests, meaning that it does not conform to the empirical data. This failure further suggests that the model does not represent actual ecological processes. Thus, the recommended harvest quotas are unsubstantiated numbers bearing minimal connection to the condition of the ecosystem.

Horseshoe Crabs

- The ARM model estimates horseshoe crab abundance by processing data from three trawl surveys. The data from these surveys are not significantly correlated, suggesting that they largely reflect random fluctuations rather than meaningful biotic signals. By consolidating these results into a single, Delaware Bay-wide population estimate, the model manufactures a veneer of certainty that conceals the underlying prevalence of random noise.
- Beyond the inherent limitations of the trawl survey data, the model fails to adjust for confounding factors, such as water depth and temperature, that impact the survey results. When Dr. Shoemaker adjusted for these factors and reanalyzed the data, there was no conclusive trend in horseshoe crab abundance, undercutting the ARM model's claim of a modest positive trajectory.

4

- Dr. Shoemaker's new analysis supplements the extensive analysis submitted last year that explained how the model generates highly overoptimistic horseshoe crab population projections.

3. **Implementing the ARM model's recommendations would pose a profound risk of violating the Endangered Species Act.**

- ASMFC would violate the ESA by authorizing horseshoe crab harvest at levels that would "take" red knots, a federally protected species. Taking a species includes harming it, which in turn includes "significant habitat modification or degradation where it actually kills or injures wildlife by significantly impairing essential behavioral patterns, including breeding, feeding or sheltering."[9] The ESA's "take" prohibition extends to governmental authorizations to third parties to undertake actions that result in the incidental take of ESA-listed species because those authorizations "solicit" or "cause" prohibited take.[10]
- The ARM model is not informative as to whether any particular horseshoe crab harvest level would result in an unlawful take of red knots. The model does not accurately represent the status of horseshoe crabs and red knots, and it is oblivious to the dependence of red knots on horseshoe crabs. Since the model does not represent ecological conditions, the Board cannot rely on it to assess ecological impacts or ensure compliance with the law.
- USFWS's evaluation of the ARM framework provides no meaningful information about the likelihood of an ESA violation. In stating that the model's harvest recommendations would "pose[] negligible risk to red knot recovery and negligible risk of take," USFWS merely characterized the model's own outputs. Since the model claimed that its recommended harvest quotas would be harmless, the agency concluded that no take would be likely. USFWS's statement hinges on the accuracy of the model, which is deeply flawed.

4. **The ongoing stakeholder survey cannot justify a resumption of the female horseshoe crab harvest.**

- The Board must make management decisions based on the best available science and legal requirements. The vulnerability of red knots and horseshoe crabs, together with the ARM model's inability to generate accurate predictions of the effects on red knots of horseshoe crab harvest levels, mandate that the Board take a risk-averse approach and, at a minimum, maintain the prohibition on harvesting females and refrain from increasing male harvest quotas.
- To the extent that the Board also considers public opinion, the public has already spoken on this issue. When the Board accepted public comment last year on whether to adopt the new ARM model, more than 34,000 people expressed their opposition, compared to only 5 who expressed support. The overwhelming

---

[9] 16 U.S.C. § 1532(19) (defining take); *id.* § 1538(a)(1)(B) (take prohibition); 50 C.F.R. § 17.3 (defining harm).
[10] *Strahan v. Coxe*, 127 F.3d 155, 163 (1st Cir. 1997); 16 U.S.C. § 1538(g).

message was clear: female Delaware Bay-origin horseshoe crabs should not be harvested for bait.

- The Board has since decided to conduct a stakeholder survey to gauge the level of support for the bait harvest of female horseshoe crabs. Unlike the public comment solicitation, this survey is open only to an undisclosed, hand-selected group of respondents.
- Whatever the survey's outcome, it cannot justify reauthorizing a female bait harvest. The Board must not discount public comments and scientific and legal imperatives through opaque engagement with its selected survey respondents.

The remainder of these comments elaborate upon each of those four points. Dr. Shoemaker's new analysis immediately follows these comments. These comments and analysis supplement the comments that New Jersey Audubon and Defenders of Wildlife submitted prior to the adoption of the ARM model (the "Addendum VIII comments"). The Addendum VIII comments—including expert reports by Dr. Shoemaker and Dr. Romuald Lipcius—are incorporated by reference and attached.

Respectfully submitted,

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

Technical Comments

I. THE AVAILABILITY OF HORSESHOE CRAB EGGS ON THE BEACH, NOT TRAWL SURVEY DATA, IS THE MOST DIRECT AND MEANINGFUL DETERMINANT OF RED KNOT SURVIVAL.

The ARM model is irreparably distorted by its core finding that the abundance of female horseshoe crabs has virtually no impact on red knots. That finding defies both historical observation and empirical data, and it subverts the very purpose of utilizing a model to inform horseshoe crab harvest quotas. By contrast, the density of horseshoe crab eggs on the beach correlates strongly with red knot survival. The Board must fully account for the vital correlation between the two species when making management decisions.

A. Horseshoe Crab Eggs Are Critical to the Survival of Red Knots at Delaware Bay.

The relationship between horseshoe crabs and red knots is an extraordinary example of the interconnectedness of life on Earth. Each year, red knots fly from as far as the southern tip of South America to breed in the Arctic Circle. For most red knots, this epic journey coincides with another ecological marvel: the emergence of horseshoe crabs from the waters of Delaware Bay to spawn on the beach. Historically, an enormous population of horseshoe crabs has produced a vast resource of eggs. This bounty of eggs serves as a critical food source for red knots. Having already flown thousands of miles at significant physiological expense, red knots can consume enough eggs in less than two weeks to double their body weight and gain the energy to complete their migration and breed successfully.[11] Horseshoe crab eggs may be especially important for the most southern-wintering red knots, whose migrations are the longest and most energy-intensive.[12] Only with a superabundance of horseshoe crabs can red knots access the eggs: horseshoe crabs lay their eggs too deeply in the sand for red knots to reach, but successive waves of spawning crabs churn the sand, elevating a portion of the eggs toward the surface.[13]

The importance of horseshoe crab eggs shapes red knots' migratory paths, and the plethora of crabs has historically drawn red knots to Delaware Bay.[14] USFWS has labeled the overharvest of horseshoe crabs in Delaware Bay a "primary causal factor" in red knots' decline.[15] A key objective of the ARM framework is to "ensure that the abundance of horseshoe crabs is not limiting the red knot stopover population or slowing recovery."[16]

---

[11] Lawrence Niles et al., *Effects of Horseshoe Crab Harvest in Delaware Bay on Red Knots: Are Harvest Restrictions Working?*, 59 BioScience 153, 154 (2009); New Jersey Department of Environmental Protection, *Wildlife Populations: Red Knot* 1-2 (2020), https://dep.nj.gov/wp-content/uploads/dsr/trends-red-knot.pdf; Duijns, *Body Condition Explains Migratory Performance* at 4-6.

[12] *See* FWS, *Draft Recovery Plan for the Rufa Red Knot* 13-14 (May 2021).

[13] Niles, *Effects of Horseshoe Crab Harvest* 155.

[14] The utilization of other horseshoe crab-rich stopover sites in South Carolina further bolsters the importance of horseshoe crabs to red knots.

[15] 79 Fed. Reg. at 73,707.

[16] ARM Report 25. More information about the role of horseshoe crab eggs in red knot migration is available in New Jersey Audubon and Defenders of Wildlife's comments on Addendum VIII (attached).

In light of the well-established reliance of red knots on horseshoe crabs, achieving the ARM framework's objective requires the restoration of adequate horseshoe crab egg resources. But instead, the ARM model concludes—contrary to decades of observation and belying the ARM framework's own objective statement—that red knot abundance bears almost no connection to the abundance of horseshoe crabs.[17] The model would predict that red knot numbers would most likely increase even if horseshoe crabs disappeared entirely.[18] According to the model, horseshoe crabs, including the egg-laying females, could be harvested in large numbers, and red knots would barely notice the difference.

The ARM model is wrong. As described below, and building on decades of observation, the fate of red knots is significantly correlated with the fate of horseshoe crabs. The model's contrary—and counterfactual—conclusion does not represent the dynamics of the ecosystem and results from flaws in how the model is structured and processes data.

> B. <u>The density of horseshoe crab eggs on the beach strongly correlates with red knot survival and demands central consideration in management decisions.</u>

The ARM model entirely ignores the most direct measure of whether there are enough horseshoe crab eggs for red knots: the density of eggs at or near the surface of the beach. Data on egg density have reliably and consistently been collected for decades. Peer-reviewed, published research shows that egg density has declined by an order of magnitude since the 1980s.[19]

Building on that peer-reviewed research, Dr. Shoemaker found a significant positive correlation between egg density and red knot survival. The data show that higher egg density has historically tracked with higher red knot survival rates. The reverse is also true: projecting forward from this correlation, multiple years of low egg density would likely decimate the red knot population.[20]

Instead of using egg density data, the ARM model uses data that are, at best, a remote proxy of food availability for red knots: the abundance of female horseshoe crabs, as estimated from trawl surveys conducted in the open sea. The ARM model illogically assumes that the ecosystem is meeting the needs of red knots based on horseshoe crab trawl surveys, even as horseshoe crab egg densities on the beach languish at low levels. Thus, a model with the stated purpose of protecting red knots is erroneously being used to assert that red knots hardly need protection after all.

As described above, Dr. Shoemaker previously explained that the ARM model would project a likely increase in red knot abundance even if horseshoe crabs vanished entirely from Delaware

---

[17] *See, e.g.*, ARM Report 86.
[18] *See* Kevin Shoemaker, *Review of 2021 ASMFC ARM Revision* 6-12 (Sept. 2022), *in* Addendum VIII comments (attached) ("Shoemaker 2022 Analysis").
[19] *See* Joseph A. M. Smith et al., *Horseshoe Crab Egg Availability for Shorebirds in Delaware Bay: Dramatic Reduction After Unregulated Horseshoe Crab Harvest and Limited Recovery After 20 Years of Management*, Aquatic Conservation: Marine and Freshwater Ecosystems 1, 8 (2022), https://doi.org/10.1002/aqc.3887.
[20] *See* 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* 19-27 (Sept. 2023) ("Shoemaker 2023 Analysis").

Bay.[21] He has now supplemented that finding with a projection based on the correlation between red knot abundance and horseshoe crab egg density (Figure 1).[22] The contrast between the two projections is stark and highlights the recklessness of accepting the ARM model's representation of the ecosystem. Notably, the projection based on egg density—unlike the projection based on horseshoe crab abundance—does not assume that horseshoe crabs vanish entirely but incorporates the less extreme scenario that egg density stagnates at the lowest historically observed level. Yet even under that relatively modest and more plausible scenario, the consequences for red knots would be dire.
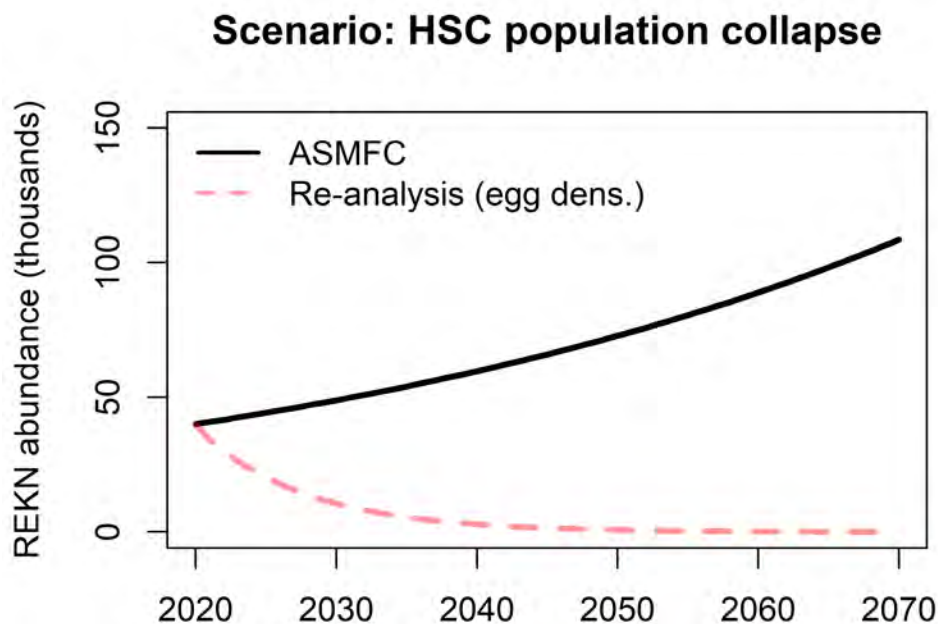

Figure 1 (appears as Figure 6 in Dr. Shoemaker's analysis). The solid black line represents the ARM model's weak correlation between red knot abundance and horseshoe crab abundance (as estimated from trawl surveys) and depicts a scenario in which horseshoe crabs completely disappear from Delaware Bay. The dashed red line represents the correlation between red knot abundance and egg density and depicts a scenario in which egg density stagnates at the lowest historically observed level.

Figure 1 shows that persistently low egg density would cause the abundance of red knots at Delaware Bay to plummet toward near-zero levels. It further undercuts the ARM model's implausible expectation that red knot abundance would increase even in the total absence of horseshoe crabs.

It bears emphasis that, while egg density is the best indicator of resource adequacy for red knots, there is almost certainly a positive correlation between the abundance of horseshoe crabs and red knots. The model's failure to find such a correlation may be attributable to trawl surveys' inaccurate measurements of horseshoe crab abundance. Section II.B, *infra*, presents Dr. Shoemaker's finding that the trawl survey data are likely more reflective of random noise than

---

[21] *See* Shoemaker 2022 Analysis 6-12.
[22] *See* Shoemaker 2023 Analysis 25.

horseshoe crab demographics. If the horseshoe crab abundance estimates are inaccurate, then the strength of their correlation with red knot abundance is meaningless.

Management decisions that affect a threatened species like the red knot, including by causing prohibited take, demand a precautionary approach. Basing management decisions on the ARM model would be risk-prone and invite calamity for red knots.

II. **THE ARM MODEL OVERESTIMATES AND MISREPRESENTS THE HEALTH AND RESILIENCE OF RED KNOTS AND HORSESHOE CRABS AT DELAWARE BAY.**

In addition to disregarding the connection between horseshoe crabs and red knots, the ARM model contains fundamental errors and deficiencies that prevent it from accurately representing the status of either species individually. As a result, its recommended harvest quotas do not reflect ecological conditions. Implementing the model's recommendations would have adverse outcomes that the model cannot accurately predict and put both red knots and horseshoe crabs at greater risk, in sharp contrast to the precautionary approach that managing an imperiled ecosystem demands.

A. The ARM model's evaluation and projections of red knot demographics are not reliable.

i. *The model artificially inflates the red knot survival rate.*[23]

The ARM model incorrectly estimates that red knots' lifespan is roughly three times as long as similarly sized shorebirds—nearly 15 years compared to 5 years. The lifespan estimate is derived from the annual survival rate, which the model estimates at 93%. Most other estimates of the survival rate for red knots (and similarly sized shorebirds) are closer to 80%.

Overestimating the survival rate results in the model underestimating the vulnerability of red knots to a single unsuccessful breeding year. To maintain a stable population, each female needs to replace herself at least once (on average) during her lifetime. For example, a female that is reproductively active for 14 years may be relatively unaffected by one or two poor breeding years. But for a female that is reproductively active for 4 years, the same conditions would significantly reduce her likelihood of reproductive success, even if, on average, she produces more offspring per year.

The model's erroneous survival rate flows from its method of tabulating red knot resightings. Researchers have long affixed leg bands to red knots, with each band having a unique, three-character code. By reading the codes from red knots that are banded and then return to Delaware Bay in subsequent years, researchers acquire data about what proportion of red knots survive from year to year.

The difficulty of reading codes from leg bands means that researchers need to account for two types of misread errors. The ARM model accounts for one type by ignoring a reading if the code

---

[23] This finding is presented at Shoemaker 2023 Analysis 8-14.

was never actually used on a leg band. But it does not account for codes that are mistaken for other existing codes. For example, assume that in year 1 of the study, a red knot is assigned the code 1AB, and in year 7, a red knot is assigned the code 7AB. In year 8, a researcher may misread "7AB" as "1AB," even though the bird assigned 1AB may have died years earlier.

To minimize misread errors, researchers can weed out codes that are sighted only once in a season. Uncorroborated by additional readings, these codes are more likely to be misreads. Dr. Shoemaker recalculated red knots' survival rate after weeding out these uncorroborated potential misreads. The resulting estimated survival rate dropped to approximately 80%, which is much more consistent with most other estimates.[24]

For further verification, Dr. Shoemaker also calculated the survival rate using readings only from red knots that were captured after previously having been banded—upon capture, the codes could be read at close range. These close-range readings constitute a much smaller data-set but would be expected to include minimal misreads. This subset of readings yielded an estimated survival rate of approximately 79%, consistent with Dr. Shoemaker's corrected overall estimate and estimates from other researchers.

The enormous overestimate of red knot survival is indicative of how profoundly the ARM model fails to represent even the basic lifecycle of the species it is supposed to protect—and why the model should not be used to make existential decisions affecting that species.

ii.   *The ARM model misrepresents trends in red knot abundance.*[25]

The ARM model has a strong tendency to detect false trends in red knot abundance, even when no trend exists. Thus, the model cannot be trusted to assess one of the most important factors: whether and to what degree the red knot population is increasing or decreasing.

This problem results from a design flaw in a key component of the model that estimates abundance and recruitment. The component, called a "state-space" model, uses annual red knot population counts to estimate various metrics related to red knot demography (all of which feed into abundance estimates). These metrics, or "parameters," include estimates of initial red knot abundance, annual recruitment, and the effect of horseshoe crab abundance on red knot recruitment. But the initial dataset is far too small to support the large number of parameters estimated from it.

More concretely, this component of the model draws from just 14 datapoints: the peak count of red knots in Delaware Bay for each of the years 2005-2018. From that limited dataset, the model estimates at least 18 different parameters. As models become "overparameterized," they bear a decreasing relationship to the truth. Dr. Shoemaker analogizes this phenomenon to a parachutist

---

[24] Allan J. Baker et al., *Rapid population decline in red knots: fitness consequences of decreased refuelling rates and late arrival in Delaware Bay*, Proceedings of the Royal Society of London, Series B: Biological Sciences 271(1541), 875-882 (2004); Theunis Piersma et al., *Simultaneous declines in summer survival of three shorebird species signals a flyway at risk*, Journal of Applied Ecology 53(2), 479-490 (2016); Verónica Méndez et al., *Patterns and processes in shorebird survival rates: a global review*, Ibis 160(4), 723-741 (2018).
[25] This finding is presented at Shoemaker 2023 Analysis 34-39.

connected to a parachute with suspension cords. As the number of suspension cords declines, the parachutist and parachute become increasingly untethered. Similarly, with insufficient datapoints, the parameters lose a strong connection to the truth. Instead, the model is likely to conclude that false, or "spurious," trends exist, even when the data indicate no such thing.

Dr. Shoemaker's tests revealed that the ARM model is highly likely to find spurious trends. To test this, he generated 50 sets of random, white-noise population count data that lacked a trend in either direction. Feeding those 50 random datasets into the model, he found that the model contrived a significant, spurious trend 42 times. That is, working from a dataset of white noise, the model was more than 80% likely to project that red knot abundance was on a trajectory to increase or decrease significantly by the year 2100.

This flaw in the model is unlikely to be resolved through the accumulation of more data in future years. While the acceptable ratio of datapoints to parameters varies, Dr. Shoemaker explains that 30-to-1 is sometimes used as a rule of thumb. The affected component of the ARM model has less than 1 datapoint per parameter. Even though one additional datapoint of red knot abundance is collected each year, it would take decades before the dataset grew large enough to support the demands that the model places up on it.

iii. The model bears little resemblance to real-world data.[26]

Based on the information that ASMFC has released, the ARM model has not undergone sufficient goodness-of-fit testing. As Dr. Shoemaker explains, such "testing is a critical validation step . . . [for] ensuring that key assumptions made during the modeling process are reasonable and justified."[27]

To fill this gap, Dr. Shoemaker performed four goodness-of-fitness tests for various aspects of the ARM model, focusing on the open robust design component of the integrated population model—a portion of the model that measures red knot survival among other parameters. Each of the four tests assessed different parameters in order to test different aspects of the model.

The model failed each of the goodness-of-fit tests by a wide margin. Dr. Shoemaker explains that these failures "cast[] additional doubt on conclusions generated from this model."[28] Basing management decisions on a model that bears so little resemblance to real data would be an exercise in arbitrary and risk-prone decision-making.

B. <u>Properly evaluated, the horseshoe crab trawl surveys do not indicate a positive trend in horseshoe crab abundance.</u>

As discussed in Section I, the ARM model is centered around the correlation between red knot abundance and female horseshoe crab abundance as measured by trawl surveys. This overreliance on trawl survey data is inherently inappropriate because the data do not correlate with red knot abundance, and the model ignores data on horseshoe crab egg density that correlate

---

[26] This finding is presented at Shoemaker 2023 Analysis 39-41.
[27] *Id.* at 39.
[28] *Id.* at 41.

strongly with red knot outcomes. That problem is compounded by several flaws in how the model uses and processes the trawl survey data. Upon correcting some of those flaws, it becomes clear the trawl surveys do not support the ARM model's optimistic assessment of the horseshoe crab population trajectory. In fact, the trawl surveys reveal no conclusive trend in either direction, bolstering the need to make precautionary management decisions for this overexploited species, especially considering that the species remains depleted relative to historical levels.

To assess the horseshoe crab population, the ARM model processes data from three different trawl surveys using a catch multiple-survey analysis ("CMSA"). While the goal is to derive a meaningful signal from the three surveys collectively, the survey data seem to be heavily influenced by random fluctuations, rendering any collective signal meaningless. In fact, there is virtually no correlation among the horseshoe crab abundance data from the three surveys.[29] The resulting unified abundance estimate provides a false veneer of certainty, masking an underlying reality of random noise.

In addition, the CMSA does not adjust for confounding factors that skew the survey data. The number of horseshoe crabs counted in the surveys can be impacted by seasonality, water temperature and depth, and other factors. But the CMSA does not adjust for these impacts, allowing the data to remain skewed.

By adjusting for these confounding factors and reanalyzing the data, Dr. Shoemaker made two striking findings: first, in contrast to the ARM model's finding that horseshoe crabs are recovering, the trawl surveys do not indicate any upward trend in the population of female horseshoe crabs in Delaware Bay.[30] And second, the three surveys are even less correlated with each other—and more likely to reflect random noise—than they previously appeared.[31]

This new analysis supplements the analysis that Dr. Shoemaker performed prior to the approval of Addendum VIII, detailing serious deficiencies in the CMSA's evaluation of horseshoe crab data. For example, Dr. Shoemaker previously showed that:

- The CMSA does not properly account for uncertainty in its horseshoe crab abundance projections.[32] It treats the potential for inherent biases—which could persistently skew the model's projections too high or too low—as if they were year-to-year variations that would cancel each other out over time. If the CMSA properly accounted for uncertainty, it would show that horseshoe crabs face a realistic risk of falling to extremely low levels even in the absence of any harvest (bait or biomedical) or discard mortality.

---

[29] *See id.* at 17-19.

[30] *See id.* at 28-33.

[31] *See id.* at 17-19. While Dr. Shoemaker adjusted the trawl survey data for confounding factors, the trawl surveys remain unsuitable for quantifying the correlation between horseshoe crabs and red knots. Even with adjusted data, the surveys appear inherently random and vastly inferior to egg density data as a corollary to red knot survival. Instead, Dr. Shoemaker's analysis reveals that the trawl survey data are completely uncorrelated, and, even using the ARM model's preferred data source, horseshoe crab abundance is not increasing.

[32] *See* Shoemaker 2022 Analysis 12-18.

- For years when horseshoe crab recruitment data were not available, the CMSA filled in numbers that are absurdly higher than the estimates from any year with empirically observed data, resulting in significantly inflated long-term abundance projections.[33]

The Addendum VIII comments also presented analysis by Dr. Romuald Lipcius highlighting many worrying trends in the trawl survey data from Virginia Polytechnic Institute (which collects the most detailed demographic information on horseshoe crabs in Delaware Bay). For example, the Virginia Tech survey data indicate that the body size of female horseshoe crabs in Delaware Bay is decreasing, the ratio of females to males is decreasing, and the number of newly mature females is disturbingly low, among other troubling developments.[34] As Dr. Lipcius explained, these are not the trends that one would expect to find in a recovering population, especially one in which females have been protected from harvest.[35]

The prior analyses, together with Dr. Shoemaker's new analysis, strongly suggest that horseshoe crabs are not recovering in Delaware Bay. They require protection for their own sake, as well as for the nourishment that their eggs provide to red knots and other species. They certainly should not be harvested at levels recommended by a model that misrepresents the condition and trajectory of both of the species that it considers.

III. IMPLEMENTING THE ARM MODEL'S RECOMMENDATIONS WOULD POSE A PROFOUND RISK OF VIOLATING THE ENDANGERED SPECIES ACT.

In their Addendum VIII comments, New Jersey Audubon and Defenders of Wildlife cautioned that, by utilizing the ARM model, ASMFC would risk violating the Endangered Species Act. The ESA prohibits any person from "tak[ing] any [endangered] species within the United States or the territorial sea of the United States."[36] Such prohibited "take" includes actions that "harm" listed species, including "significant habitat modification or degradation where it actually kills or injures wildlife by significantly impairing essential behavioral patterns, including breeding, feeding or sheltering."[37] The ESA's "take" prohibition extends to governmental authorization to third parties to conduct activities that themselves result in unauthorized incidental take, thus "solicit[ing]" or "caus[ing]" an offense.[38] By virtue of a regulation in effect at the time the red knot was listed as threatened, the statutory take prohibitions apply to the take of many USFWS-listed threatened species, including the red knot.[39]

The Addendum VIII comments explained that ASMFC would likely commit a take by authorizing a harvest of female horseshoe crabs, impairing red knots' ability to feed. While the Board did not accept the model's recommendation to authorize a female harvest for 2023, that

---

[33] *See id.* at 22-24.
[34] *See* Romuald Lipcius, *Expert Report* 6, 10 (Sept. 2022), *in* Addendum VIII comments (attached).
[35] *See id.* at 4-5.
[36] 16 U.S.C. § 1538(a)(1)(B).
[37] 50 C.F.R. § 17.3.
[38] *Strahan v. Coxe*, 127 F.3d at 163; 16 U.S.C. § 1538(g).
[39] 50 C.F.R. § 17.31(a) (applying the provisions of § 17.21 (addressing endangered species) to threatened species listed on or prior to September 26, 2019, unless USFWS has promulgated a species-specific rule); *id.* § 17.21(a), (c) ("[I]t is unlawful . . . to solicit another to commit or to cause to be committed" the taking of an endangered species.).

remains a threat for future years due to the ARM model's proclivity for recommending a substantial female harvest quota.

Because the model does not accurately represent the relationship between horseshoe crabs and red knots, it offers no useful guidance on whether *any* particular harvest level amounts to a take. Notably, while much of the discussion around the ARM model has addressed the risk of a female horseshoe crab harvest, the model is similarly unable to assess the risk posed by a male harvest. In this information void, allowing any horseshoe crab harvest is a roll of the dice.

The Endangered Species Act requires a precautionary approach. As the Supreme Court has stated, "Congress has spoken in the plainest of words, making it abundantly clear that the balance has been struck in favor of affording endangered species the highest of priorities, thereby adopting a policy which it described as 'institutionalized caution.'"[40] In line with that principle, the ARM framework's stated objective includes "*ensur[ing]* that the abundance of horseshoe crabs is not limiting the red knot stopover population or slowing recovery."[41] It would be inconsistent with Endangered Species Act requirements (and the ARM framework's objective) to utilize a model that, among other deficiencies:

- by virtually disregarding the correlation between red knots and horseshoe crabs, fails to appreciate the importance of the very resource that it is managing for;
- takes no account of egg density on the beach surface—the one datapoint that directly measures whether the horseshoe crab population is providing adequate nutrition for red knots;
- significantly misapprehends the life cycle of red knots, vastly overestimating their lifespan;
- through an overparameterized model, incorrectly concludes that there are trends in red knot abundance even when no trends exist;
- generates horseshoe crab abundance estimates from trawl surveys that are heavily influenced by random noise; and
- produces erroneously optimistic projections of horseshoe crab abundance while disregarding multiple, persistent negative trends in horseshoe crab demographics.

Without a clearer understanding of the impact of the horseshoe crab harvest, the only lawful, precautionary, and ecologically defensible approach is for the Board to set conservative (if any) bait harvest levels. Certainly, no reauthorization of a female bait harvest could be defensible under these circumstances.

Moreover, as explained in the Addendum VIII comments, ASMFC cannot rely upon USFWS's statement that the ARM model's harvest recommendations would "pose[] negligible risk to red knot recovery and negligible risk of take."[42] USFWS's evaluation was based entirely on the model's own outputs and thus harbors all of the flaws inherent in the model itself. In particular, the evaluation accepts that the correlation between horseshoe crab abundance and red knot success is minimal without considering other evidence of a correlation (like egg density).

---

[40] *Tenn. Valley Auth. v. Hill*, 437 U.S. 153, 194 (1978).
[41] ASMFC, *ARM Report* 25 (emphasis added).
[42] FWS Evaluation 3.

Unsurprisingly, it concludes that "there is a very small probability (<1%) ARM management will result in a lower abundance of red knots."[43]

By merely repackaging the ARM model's findings, the USFWS evaluation never provided significant additional information about the effects of implementing the model's recommendations. Dr. Shoemaker's new analysis highlights additional flaws in the model and, by extension, in the USFWS evaluation, demonstrating that the evaluation is even less informative than previously known. The USFWS evaluation lends no independent factual or legal support for the Board's reliance on the ARM model.

IV. THE ONGOING STAKEHOLDER SURVEY CANNOT JUSTIFY A RESUMPTION OF THE FEMALE HORSESHOE CRAB HARVEST.

It is imperative that the Board base horseshoe crab harvest quotas "on the best scientific information available"[44] and the requirements of the ESA. As detailed extensively in the comments above and the attached analyses and comments, the ARM model does not provide a firm scientific basis for setting horseshoe crab bait harvest quotas and cannot predict the impact of its recommended quotas. Implementing those quotas would therefore imperil the ecosystem in ways that the Board cannot foresee. In the absence of reliable information about what harvest levels the ecosystem can sustain, the only scientifically defensible approach is to set highly conservative harvest quotas—continuing the prohibition on harvesting females and certainly not increasing male harvest quotas from current levels.

The results of a stakeholder survey cannot alter the Board's obligation to make scientifically grounded and legally sound management decisions. But to the extent that the Board also considers public opinion, the Board must respect the overwhelming opposition to a female horseshoe crab bait harvest expressed in the comments submitted on Addendum VIII last year. The Board's comment solicitation yielded 34,631 submissions, all but 5 of which opposed the adoption of the new ARM model—a tally that reflected the public's "[o]pposition to female horseshoe crab harvest."[45] Although the Board approved the new ARM model, it appropriately rejected the model's recommendation to authorize a female harvest, "[a]cknowledging public concern about the status of the red knot population in the Delaware Bay."[46] Shortly after the Board's decision to adopt the revised ARM model but decline to adopt its recommendation for 2023, the chair of the subcommittee responsible for the ARM model wrote:

> [T]here is absolutely no appetite for female harvest from any stakeholder. Not only were the shorebird advocates strongly against any resumption of female harvest, but it appears that the bait industry is completely satisfied with male only harvest. . . . [W]hen ASMFC is asked by NGOs in the media where the pressure for female harvest is coming from, it's really coming from us scientists in our

[43] *Id.*

[44] ASMFC, *Interstate Fisheries Management Program Charter* § 6(a)(2) (Aug. 2019).

[45] *See* Memorandum from Caitlin Starks on Public Comment on Draft Addendum VIII to the Horseshoe Crab Fishery Management Plan 1 (Oct. 20, 2022), *in* ASMFC, Materials for the 2022 Annual Meeting of the Horseshoe Crab Management Board, http://www.asmfc.org/files/Meetings/2022AnnualMeeting/HorseshoeCrabBoard_Nov2022.pdf.

[46] ASMFC, *supra* note 6.

desire to find an optimal solution to the problem statement. Perhaps our problem statement is no longer applicable in this situation.[47]

While the Board maintained protections for female horseshoe crabs in 2023, it did not resolve whether those protections would extend to future years. Instead, the Board expressed interest in a process "with stakeholders and managers and scientists, to try to help better inform future goals and objectives and modeling approaches" and "to really start to talk about what our goals and objectives are for both the fishery and the ecosystem."[48] In a subsequent meeting, Board members repeatedly conveyed that the primary objective of the stakeholder engagement was to determine whether any public appetite exists for a female bait harvest—and if not, to adjust the management framework accordingly.[49] After reviewing options for stakeholder engagement, the Board opted to proceed with a survey. Unlike the public comment period, however, this survey would seek the perspectives only of hand-selected respondents, not all interested members of the public.[50]

The public has already spoken on this issue. Whatever the outcome of the stakeholder survey, the Board must respect the overwhelming opposition to a female harvest expressed in the public comments on Addendum VIII. The entire public, including everyone invited to participate in the stakeholder survey, had the opportunity to weigh in during the public comment period, but only a small fraction of commenters were invited to complete the survey. ASMFC appears to be denying requests for additional stakeholders—even longtime horseshoe crab advocates—to complete the survey, and it has denied a request to disclose the list of people who received the survey.[51] This method of secretive, restricted engagement falls far short of ASMFC's obligation to "provide adequate opportunity for public participation."[52] Public transparency is essential

---

[47] Email from John Sweka to Conor McGowan, David Smith, James Lyons, Clinton Moore, Anna Tucker, Richard Wong, Kristen Anstead, Caitlin Starks (additional recipients redacted) re Kristen's presentation to the HSC board (Nov. 17, 2022) (obtained via FOIA).

[48] Comments of Shanna Madsen 28, *Draft Proceedings of the Horseshoe Crab Management Board Hybrid Meeting: November 2022* (as approved at the May 2023 meeting), https://www.asmfc.org/files/Meetings/2023SpringMeeting/May3/HorseshoeCrabManagementBoard_May2023.pdf.

[49] *See, e.g.*, *Horseshoe Crab Board Proceedings May 2023*, at 1:00:21, https://www.youtube.com/watch?v=QFw9N1LJF-A, Comments of John Clark ("We decided to move ahead with this item to see what we want to do in the future 'cause of course, if there is no desire for female harvest, that's a whole different way to manage the species."); *id.* at 1:03:47, Comments of Shanna Madsen ("I'd like to see us start with option one, which is putting together a survey to ask that very direct question: do our constituents want us to harvest female horseshoe crabs? And if the answer is no, then I think that really helps us outline what that objective statement is."); *id.* at 1:05:48, Comments of Rick Jacobson ("I agree, too, with the previous speakers that we do have a fundamental question that we need to ask ourselves first: what is the public appetite for the harvest of female horseshoe crabs from Delaware Bay? It is a critical question.").

[50] *See id.* at 1:12:21, Comments of ASMFC Fisheries Policy Director Toni Kerns ("I just want to make it clear that it's not our intention to send this survey to the world. We intend to hit the major stakeholders. . . . We're not trying to exclude the public, but we have just done a management document where we received, how many, 34,000 comments, and we heard from the general public on their intentions, and we still want to make sure we're capturing all the stakeholders here, but we're also not looking for that many comments to have to summarize in order to provide feedback to this Board.").

[51] *See* email from Caitlin Starks to Susan Linder denying request for survey (Aug. 24, 2023); email from Toni Kerns to Susan Linder denying request for list of survey recipients (Sept. 13, 2023). The stated rationale for withholding the list of survey recipients was to preserve the anonymity of responses, but no information about responses was requested.

[52] ASMFC, *Interstate Fisheries Management Program Charter* § 1(c).

when setting harvest quotas for a public resource. The Board must not discount public comments based on feedback from a limited, undisclosed group of hand-selected survey recipients.

V. **CONCLUSION**

Independent analysis powerfully demonstrates that the ARM model is not suitable for managing the bait harvest of Delaware Bay-origin horseshoe crabs. The ARM model entirely fails to accurately represent what scientific study of the relationship between red knots and horseshoe crabs has already incontrovertibly established—that robust horseshoe crab populations capable of generating a superabundance of eggs on red knot stopover beaches are critical for the red knot's survival and reproduction. The model is oblivious to the strong correlation between red knots and horseshoe crabs and misconstrues data about each species, creating an unbridgeable chasm between its harvest recommendations and actual ecological conditions. Consistent with the Endangered Species Act and its own stated objective to protect red knots, as well as its obligation under the Interstate Fisheries Management Program Charter to base its decisions about horseshoe crab harvest quotas on the best available scientific information, the Board must not implement the model's recommendations. The Board's obligation includes, at a minimum, maintaining the zero-harvest bait quota for female horseshoe crabs and not increasing male-only harvest quotas from current levels.

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 Shoemaker, Ph.D.

Associate Professor, University of Nevada, Reno

September 2023

This is an expert review of the Atlantic States Marine Fisheries Commission (ASMFC)'s Adaptive Resource Management (ARM) framework – which has been approved for use in managing the Horseshoe Crab fishery in Delaware Bay – performed by Kevin Shoemaker, Ph.D. This document is intended to supplement the review of the ARM completed by Dr. Shoemaker in Fall 2022.

Dr. Shoemaker holds a Ph.D. and an M.S. 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

This report reviews the scientific merits of the Adaptive Resource Management (ARM) framework that has been approved for use by the Atlantic States Marine Fisheries Commission (ASMFC) as a tool for guiding management of the horseshoe crab (HSC) fishery in Delaware Bay and protecting the Federally Threatened *Rufa* Red Knot (*Calidris canutus rufa*; REKN). In Fall 2022

I completed an initial review of the ARM, in which I pointed out five major areas of concern: (1) the fitted relationship linking HSC abundance to REKN survival was functionally insignificant as a driver of REKN population dynamics, (2) the HSC simulation model did not correctly address parameter uncertainty, (3) the statistical model used to estimate HSC demographic processes (Catch Multiple Survey Analysis, or CMSA) exhibited poor fit to the data, (4) the CMSA results were compromised by a 4-year period during which a key source of data was not collected, and (5) the ARM lacked performance benchmarks (null models) to ensure that key model components (e.g., the effect of HSC abundance on REKN survival) meaningfully improved predictive performance versus simpler approaches. The purpose of this follow-up report is to evaluate components of the ARM for which the source code was unavailable for evaluation in my initial review. In particular, I focus on the Integrated Population Model (IPM) approach used by ASMFC for estimating REKN demographic parameters and for quantifying the influence of HSC abundance on the REKN population.

Delaware Bay is a critical stopover site used by REKNs and other shorebirds as they migrate to breeding grounds in the high arctic from their wintering grounds as far south as Tierra del Fuego (USFWS 2021). In particular, HSC eggs deposited on coastal beaches provide a necessary high-calorie food resource for REKNs and other migrating shorebird species as they replenish fat reserves depleted from their long migration and prepare for breeding. At the heart of ASMFC's ARM framework is a set of 'harvest functions' for setting HSC harvest recommendations on the basis of annual estimates of HSC and REKN abundance. In theory, these harvest functions are calibrated to maximize HSC harvest yields while causing minimal risk to the HSC or REKN populations. Optimization of the harvest functions is accomplished by running numerous alternative harvest scenarios using a two-species (HSC and REKN) demographic simulation model and weighing the benefits (harvest) and costs (population risks to HSCs and REKNs) of the simulated outcomes. The cornerstone of this two-species demographic simulation model is a weak (but statistically conclusive) positive effect of female HSC abundance on REKN survival, which serves as a formal, quantitative linkage between the two species. Therefore, the validity of the ARM framework depends upon proper specification of each species' demographic rates (e.g., survival and recruitment) and the degree to which the modeled HSC-REKN interaction

is an appropriate representation of the real-world biotic interaction between these species.

Building on the issues raised in my initial review, this report identifies six additional areas of concern (see below). Based on these concerns, I conclude that the ARM framework is not useful for managing risk to the REKN population due to HSC harvest. Furthermore, my results suggest that the revised ARM misrepresents the importance of HSCs to the REKN population and thereby underestimates both the existential risk to the REKN population posed by female HSC harvest and the potential for promoting REKN recovery through increased regulatory protections and conservation efforts aimed at promoting HSC population increases in the Delaware Bay region. The six primary areas of concern are summarized below, with technical details provided in the "supporting evidence and analyses" section.

(1) **Estimates of REKN survival used in the ARM appear to be artificially inflated, likely resulting in falsely optimistic estimates of population resilience**. The majority of previously reported estimates of annual survival for REKNs and similar shorebirds are in the neighborhood of 80%, corresponding to an average lifespan of approximately five years. In contrast, ASMFC reported a mean annual REKN survival estimate of 0.93, which corresponds to an expected lifespan of nearly 15 years. By nearly tripling the expected REKN lifespan *vis-a-vis* previous estimates, ASMFC is effectively classifying the REKN as a uniquely long-lived species among medium-sized shorebirds. Since individual females must only replace themselves once during their lifetime (on average) for a population to be stable, longer-lived species can afford a higher per-capita failure rate in breeding attempts than shorter-lived species. Therefore, long-lived species are expected to be more resilient to short-term fluctuations in recruitment. However, my findings strongly indicate that ASMFC's estimate of REKN survival is biased high due to the presence of misread errors (by which a flag code is mistaken for a code previously deployed on a different bird). The potential for misread errors in the study system has been previously acknowledged (Tucker et al. 2019). After correcting for potential misread errors, REKN survival estimates fall to approximately 80% annually – a rate more consistent with previous estimates for REKN and similar species. The apparent positive bias in ASMFC's survival estimates is likely to result in falsely optimistic estimates of population resilience to short-term environmental fluctuations, raising

3

concerns about the adequacy of the ARM framework for assessing population-level risks to this federally Threatened species.

(2) **Trawl-based indices of HSC abundance are inadequate for detecting robust links to REKN demography**. ASMFC documented a very weak (and not ecologically meaningful; see attachment) positive effect of female HSC abundance on REKN survival. This relationship is the cornerstone of the revised ARM framework, as it represents the primary functional link between the two focal species. The effect of HSC abundance on REKN survival was estimated using the output from a Catch Multiple Survey Analysis (CMSA) as a proxy for annual HSC abundance in Delaware Bay. The CMSA in turn was trained using data from three trawl-based surveys, conducted by Virginia Tech, New Jersey, and Delaware, respectively (in addition to data on known sources of HSC mortality). However, my reanalysis of the available data uncovered no conclusive relationship between REKN survival and any trawl-based index of HSC abundance. Notably, after including several additional years of REKN mark/resight data (I used REKN banding and resighting data from 2003 through 2022, whereas ASMFC's used data from 2005 to 2018), the effect of HSC abundance on REKN survival became negative (lower REKN survival with more female HSCs) when using the code and data provided by ASMFC. This result underscores the frailty of the foundational relationship on which ASMFC's two-species ARM is based. Trawl-based surveys are necessarily imperfect snapshots of the abundance of HSCs occupying Delaware Bay, obscured by differing survey methodologies and poorly understood aspects of HSC ecology, including seasonal and daily activities, habitat preferences, and degree of clustering on the seafloor. Moreover, the functional link between HSC abundance and REKN demographic rates is eroded by additional, poorly understood processes that govern the availability of HSC eggs for shorebirds, including variation in the timing of HSC egg deposition and the factors that dislodge eggs from their clusters, rendering them accessible to shorebirds. Therefore, the lack of a demonstrable effect of trawl-based HSC indices on REKN vital rates likely reflects the weakness of these indices and not the weakness of the underlying biotic interaction.

(3) **REKN survival is strongly sensitive to HSC egg-density, indicating that persistent**

degradation of the HSC egg resource could have dire consequences for the REKN population. Intuitively, surveys of HSC egg densities measured on the same beaches used by foraging shorebirds during their spring migration should more directly capture the biotic interaction between these two species. Although researchers have been consistently measuring the surface density of HSC eggs at multiple beaches across Delaware Bay (NJ side only) since 2000, ASMFC chose to rely on trawl-based surveys instead of egg-density surveys as a proxy for the HSC resource available to REKNs. My reanalysis of the Delaware Bay mark-resight database indicates that REKN survival is strongly and positively influenced by annual fluctuations in HSC egg density. Unlike the weak relationship documented in the ARM, the fitted relationship between HSC egg-density and REKN survival implies severe risks to the REKN population under a scenario of sustained low HSC egg densities. In contrast to ASMFC's two-species ARM, this alternative characterization of the HSC-REKN interaction is capable of explaining the observed decline in REKN populations during the late 20th century, which is widely attributed to unregulated harvest of HSCs in Delaware Bay. These new results strongly suggest that ASMFC's ARM framework misrepresents the importance of HSCs to the REKN population. As a result, the ARM not only severely underestimates the consequences of HSC population declines on the REKN population, but it severely underestimates the critical role that a rebound of the HSC population could play in the recovery of this federally Threatened species.

(4) **The ARM exaggerates the evidence for an increasing trend in the number of female HSCs in Delaware Bay.** Based on my reanalysis, neither the trawl-based surveys used by ASMFC nor the egg-density surveys (recently used to document an increasing trend in the HSC population) show strong evidence for increasing abundance of female HSCs in Delaware Bay over the last 20 years. As a case in point, the raw data (catch-per-unit-effort; CPUE) from New Jersey's ocean trawl survey (one of the data sources used by ASMFC for documenting a positive trend in HSC abundance) appears to indicate increasing female HSC abundance over time (statistically significant at alpha=0.05); however, when the raw CPUE numbers are adjusted for strong effects of seasonality, water temperature, depth and dissolved oxygen on HSC captures in Delaware Bay (NJ ocean trawl survey), the apparent positive trend in

HSC CPUE becomes inconclusive. Notably, the trawl-based indices used by ASMFC in their CMSA model did not control for these confounding factors. Therefore, the increasing trend in the HSC population reported by ASMFC and used in the ARM may be an artifact of differing survey conditions (e.g., differences in trawl depth or water temperature) rather than evidence of recovery of the HSC population over time. Furthermore, regression models combining the CPUE estimates (both adjusted and unadjusted) from all three trawl-based surveys showed no conclusive evidence for a trend in HSC abundance over time. Similarly, when the egg-density data were adjusted for known differences in survey methodologies, the apparent positive trend (reported in Smith et al. 2022) became inconclusive (note that this adjustment did not impact the estimated relationship between REKN survival and HSC egg-densities). Overall, my reanalysis suggests that the ARM framework exaggerates the potential for recovery of the female HSC population under present conditions, and thereby likely underestimates the risk of harvest to the HSC (and REKN) populations in Delaware Bay.

(5) **The statistical model (IPM) used for estimating REKN population parameters is over-parameterized and likely to yield spurious results**. The IPM framework used to train the REKN population model comprises two integrated sub-models: (1) a "state-space" model for estimating abundance and recruitment on the basis of population counts over time, and (2) a "capture-recapture" model for estimating survival rates from observation records of uniquely marked individuals. Whereas the data available for fitting the capture-recapture model (over 100,000 resighting records of tens of thousands of unique REKN individuals) was information-rich and well-suited for training complex models, the data available for training the state-space model was sparse by any standard, comprising 14 unique data points (one count per year from 2005 to 2018). In fact, the number of parameters estimated in the state-space model appears to exceed the number of data points. As an analogy, consider a parachute whose canopy is attached to its user with suspension cords. A minimum of three cords is necessary for the parachute to have any chance of operating correctly, yet many more cords are typically incorporated to ensure robust performance. Similarly, a free parameter (an "unknown") must be tethered to the truth using data points

as suspension cords. A model's claim to truth strengthens as the ratio of data points to free parameters increases; statisticians often recommend a ratio exceeding 30 or more for robust model performance. With less than 1 data point per parameter, the IPM's state-space model is occupying a danger zone statisticians refer to as "over-parameterization", or "over-fitting". Over-parameterized models have a strong tendency to produce spurious results (results that fail to replicate when confronted with new data). To confirm the tendency of the REKN IPM to yield spurious results, I generated artificial REKN count data under a model with no underlying trend (a white-noise process) and assessed how often the IPM erroneously detected a trend. After running 50 replicates (iteratively replacing the peak-count data with newly simulated white-noise), the IPM falsely detected an ecologically meaningful temporal trend (increase or decline in abundance over time) over 80% of the time. Among the unknown quantities estimated from the 14 peak-count data points are several terms critical for understanding and forecasting REKN population dynamics, including initial abundance, population trends (growth or decline), mean recruitment, and the effect of HSC abundance on recruitment. Lacking sufficient data for parameter estimation, the REKN recruitment and population trend estimates used in the ARM model are more likely to reflect random noise in the peak count data rather than the demographic reality of the REKN population. Therefore, the REKN demographic simulations used in the ARM should not be considered a robust representation of the real-world population of *Rufa* Red Knots that uses Delaware Bay each year.

(6) **The IPM exhibits poor fit to the observed REKN data**. Goodness-of-fit (GOF) testing is a critical validation step in any model-fitting workflow, ensuring that key assumptions made during the modeling process are reasonable and justified. For example, the results from a linear regression or ANOVA test can only be interpreted once the analyst confirms that important assumptions are satisfied (e.g., that model residuals are approximately normally distributed). Although the REKN IPM is much more complex than a linear regression model, assessing goodness-of-fit is no less important. In the context of hierarchical Bayesian analysis (the paradigm used by ASMFC to fit the REKN IPM), a commonly used approach is to run a Posterior Predictive Check (PPC), in which data are repeatedly simulated under the

fitted model and compared to the actual data. If a model is unable to generate data resembling what was actually observed, the model is determined to be an inadequate representation of the true processes that generated the data. In their ARM report, ASMFC mentions (but does not further document) two PPCs that were performed to assess goodness-of-fit. One of these tests – the only test included in the publicly shared IPM code – uses a PPC to assess the degree to which the state-space model adequately represented the 14 peak-count data points. However, this test has been shown to be an insufficient gauge of model adequacy. The second and final goodness-of-fit test mentioned in the ARM report (for which the result suggests moderate lack of fit) is absent from the version of the IPM code shared publicly, so it is not possible to assess what test was actually run. However, I ran three additional PPCs to assess the degree to which the IPM adequately represented the REKN resighting data from 2003 to 2022. These tests, which were applied and reported in an earlier version of the open-robust-design (ORD) model for estimating REKN survival and stopover use (Tucker et al. 2021), indicated poor fit to the data, suggesting that the IPM is an inadequate representation of key processes operating in the REKN population – including survival. The failure of the IPM to pass rigorous goodness-of-fit tests casts additional doubts on the conclusions generated from this model.

SUPPORTING EVIDENCE AND ANALYSES

The remainder of this report supplies supporting details for the six major areas of concern identified above, including results and figures from re-analyses of the data presented in the ARM report. I report additional findings in the "supplemental analyses" section located at the end of this report.

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

The majority of published survival estimates for REKNs and other medium-sized shorebirds indicate a mean annual survival of approximately 80% (Baker et al. 2004; Piersma et al. 2016; Mendez et al. 2018), corresponding to an expected lifespan of approximately five years. In contrast, ASMFC reported a mean adult REKN survival rate of 0.93 on the basis of the REKN IPM,

corresponding to an expected lifespan of nearly 15 years. By nearly tripling the expected REKN lifespan (versus previous estimates), ASMFC is effectively classifying the REKN as a longer-lived species than other similar-sized shorebirds (Mendez et al. 2018). Since a stable population requires only that individual females replace themselves once during their lifetime, longer-lived species can afford to fail in more of their breeding attempts than shorter-lived species. Therefore, longer-lived species are expected to be more resilient to short-term fluctuations in breeding success and juvenile survival than species with a shorter lifespan (Lovich et al. 2015). ASMFC argues that their characterization of the REKN life history is accurate, and that previously reported estimates may be biased low (ASMFC 2021). In contrast, my findings strongly indicate that ASMFC's estimate of REKN survival is biased high, most likely due to the presence of misread errors in the REKN resighting database.

The presence of potential misread errors in the study system has been previously acknowledged (Tucker et al. 2019). Studies with simulated and real-world data have shown that misread errors can induce biases in survival estimates (Tucker et al. 2019; Rakhimberdiev et al. 2022). Because the data used to fit the REKN IPM was adjusted for one type of potential misread error (i.e., any observed flag codes that were never deployed in Delaware Bay were discarded), the only type of misread error that ASMFC did not account for was the possibility that a flag code was mistaken in the field for a different previously deployed code (effectively ascribing that observation to a bird that may no longer be alive). This type of misread error (if present in sufficient numbers) is known to falsely inflate survival, especially for the early years of a long-term mark-resight study (Tucker et al. 2019). Tucker et al. (2019) showed that this source of bias can be corrected by discarding observations for which a flag code was sighted only once (i.e., by a single observer during a single sampling occasion) in a given season. Although this technique necessarily discards some correct observations (only a fraction of these 'singlet' observations are likely to be in error) and thereby reduces the precision of the resulting estimates (Tucker et al. 2019; Rakhimberdiev et al. 2022), Tucker et al. (2019) demonstrated that this method was effective in removing biases induced by this class of misread error. Furthermore, Tucker et al. (2019) demonstrated that, when applied to the flag-resighting data from Delaware Bay, REKN survival estimates from early in the study period dropped from 87% to 81%, suggesting that

9

these survival estimates were artificially inflated due to misread errors.

The number of leg-flag codes that can be manufactured is necessarily limited by the number and type of symbols and colors used. Notably, given the very large number of leg flags that have been deployed on REKNs in Delaware Bay since 2003, shorebird biologists have cycled through all possible flag code permutations for the flag color (lime green) most commonly deployed in Delaware Bay. Therefore, any leg-flag codes that are read or transcribed in error are more likely to be falsely attributed to a different bird in the database than to be discarded (as it would be if there were no match in the database). Furthermore, the risk of this type of error is likely to increase substantially as the years pass and as a greater fraction of flag code permutations are deployed in the field. Coupled with the fact that longer time series are likely to manifest increasingly strong biases due to misread errors (Tucker et al. 2019) the risk of biased survival estimates and spurious trends is likely to increase markedly as the database continues to grow (e.g., in future iterations of the ARM model if potential misread errors continue to be ignored).

To assess whether ASMFC's survival estimates were biased due to the inclusion of misread errors, I used REKN banding and resighting data from Delaware Bay to estimate annual REKN survival using two different statistical frameworks: Cormack-Jolly-Seber (CJS; a standard approach to survival estimation using capture-recapture data) and the open-robust-design (ORD) framework for survival estimation used by ASMFC. First, I ran standard CJS models to estimate annual survival rates as a function of the banding data only (Cooch 2008) (i.e., ignoring all flag-resighting data). This model generated separate estimates of survival and detection probability for each year, and included additional terms for transience and 'trap-response' (Pradel and Sanz-Aguilar 2012). The banding data were much less information-rich than the re-sighting observations, with far fewer observations and a much lower re-capture rate. However, misread errors should be virtually absent from the banding records (as captured birds can be examined at close range). I trained this model (and all models presented in this section) using Markov Chain Monte Carlo (MCMC) in a Bayesian framework using JAGS (Plummer 2012), which was called from R using 'JagsUI' (Kellner et al. 2019). The 'band-only' models yielded an estimated mean annual REKN survival of 79% (Fig. 1). Based on a posterior predictive check (PPC), the Bayesian p-

value for this model was 0.1, indicating reasonable fit, with the observed data slightly over-dispersed relative to the fitted model (Fig. 1).
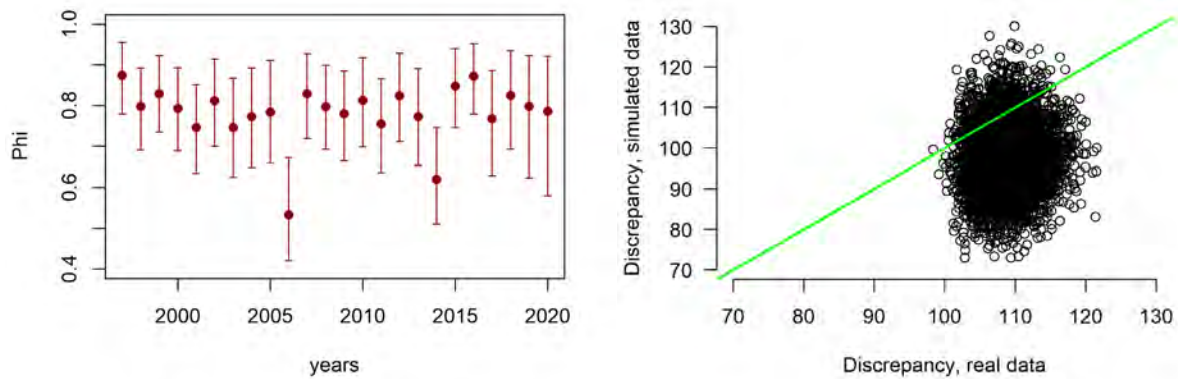
Figure 1. *Left: mean annual (apparent) REKN survival (y axis) based only on banding data (no resighting data) from Delaware Bay from 1997 to 2022, using a Bayesian CJS model. Mean estimated apparent survival was 0.79, much lower than ASMFC's estimate of 0.93. Apparent survival (Phi) is a compound parameter indicating the probability of surviving and remaining within the study area. This model accounts for the presence of transients, which can bias survival estimates low. Right: Goodness-of-fit plot for the Bayesian CJS model using only banding data from Delaware Bay. This model exhibited reasonable fit to the data, with a Bayesian p-value of 0.1.*

Next, I fitted CJS models that incorporated the resighting data along with the banding data. When potential misread errors (flag codes observed only once by a single observer in a given season) were retained for analysis, mean apparent survival across all years was ~88%, with a steady decline in survival observed over the period from approximately 2005 to 2015 (Fig. 2). When potential misread errors were removed, mean REKN survival estimates dropped to ~80% annually – a rate more consistent with previous estimates for REKNs and other similar-sized shorebirds (Fig. 2). After correcting for potential misread errors, no temporal trend in survival was apparent across the study period (Fig. 2). This pattern is consistent with the known effects of misread errors, which tend to induce a spurious negative trend in survival (more positively biased estimates going further back in time) for long-term studies (Tucker et al. 2019).
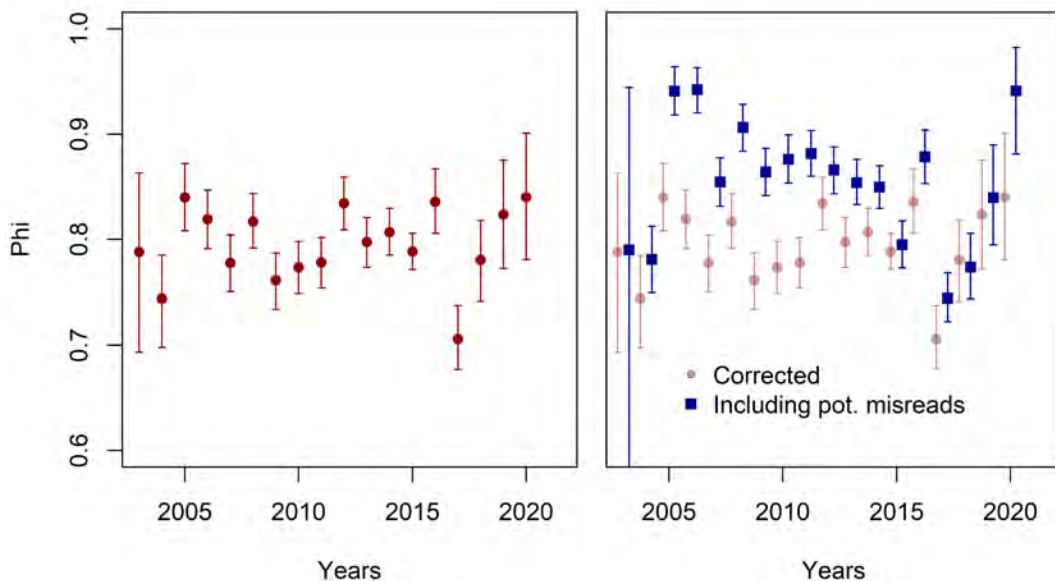
Figure 2. Left: REKN annual apparent survival estimates (Phi; error bars indicate Bayesian 95% credible intervals) based on banding and resighting records from Delaware Bay, using a Bayesian CJS model with inter-annual process variance in survival, accounting for potential transients and 'trap response' (whereby individuals are more likely to be resighted if they were resighted in the previous year). This analysis uses only birds first banded in Delaware Bay (resighting observations of birds first captured elsewhere were discarded prior to analysis, following ASMFC 2021). To correct for misread errors, only birds resighted more than once in a particular year were considered to have been resighted that year. After correcting for potential misread errors, the estimated average apparent survival (Phi) was 0.80 annually, much lower than ASMFC's estimate of 0.93. Right: Comparison of REKN apparent survival with potential misreads (blue squares, including individuals resighted only once in a given year) versus the corrected version of the data with single-resight observations removed (transparent red; same results reported in left panel). As noted by Tucker et al. (2019), misread errors are more likely to bias survival estimates high in the early years of long time series- we see this effect here, especially in the period from 2005 to 2015.

Finally, I used ASMFC's open robust design (ORD) framework to estimate annual REKN survival rates. This model, described by Tucker et al. (2022), is capable of estimating survival in addition to temporary emigration and the timing of arrival and departure from the stopover site

each year. When potential misread errors are retained in the data set, the ORD model indicated a mean REKN survival rate of 0.9 (somewhat lower than ASMFC's estimate of 0.93, but similar to the survival rate reported in Tucker et al. 2022), with survival rates generally declining across the study period, as expected for data sets with misread errors (Tucker et al. 2019) (Fig 3). When potential misread errors were removed following the methods of Tucker et al. (2019), mean apparent survival rates dropped to ~80% or below throughout most of the 20-year study period, with no apparent trend over time (Fig. 3).
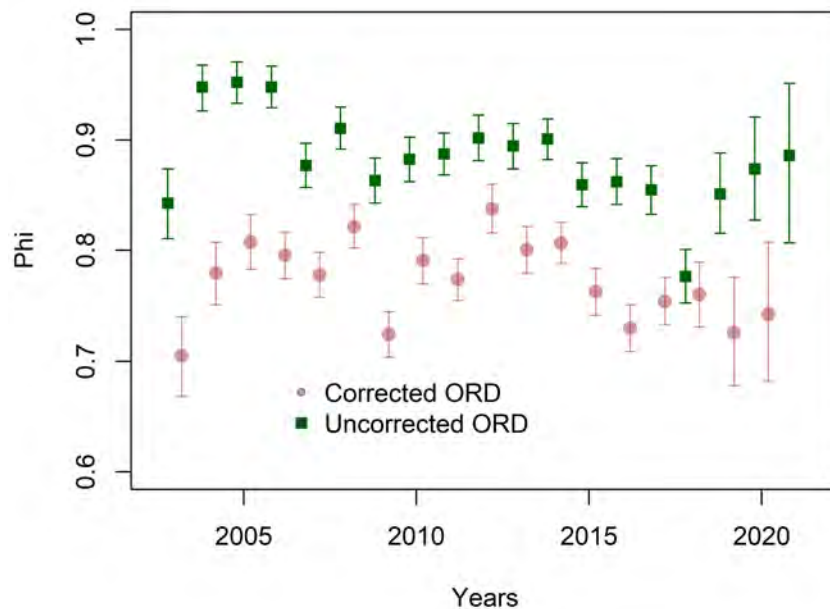
Figure 3. *Annual apparent survival (phi; y axis) estimates from the open robust design (ORD) model used by ASMFC, fitted to REKN banding and resighting data from Delaware Bay from 2003 to 2022. Red circles and confidence intervals represent estimates from the model after correction for potential misread errors (i.e., by removing instances in which a REKN was re-sighted only once in a season). Estimated survival from the uncorrected ORD model (green squares; without correction for potential misread errors) are nearly always substantially higher than the corresponding estimates after accounting for potential misread errors. In addition, the uncorrected time series (green squares) displays the characteristic (spurious) negative trend in survival typically associated with survival estimates from long time series that include misread errors (Tucker et al 2019).*

Overall, these tests strongly indicate that the REKN survival rates used by ASMFC's ARM framework are artificially inflated and do not accurately reflect the real-world population of *Rufa* Red Knots. This artificially exaggerated longevity is likely to result in falsely optimistic estimates of REKN population resilience to short-term environmental fluctuations. In reality, the REKN population is likely to be much more vulnerable to one or two bad breeding years than the ARM model would suggest. The misspecification of the REKN demographic model raises serious concerns about the adequacy of the ARM framework for assessing population-level risks to this federally protected species.

2. Trawl-based indices of HSC abundance are inadequate for modeling the biotic interaction between REKNs and HSCs

ASMFC's IPM indicated a weak (and not ecologically meaningful; see attachment) positive effect of female HSC abundance on REKN survival. This relationship is in many ways the cornerstone of the ARM framework, as it represents the primary functional interaction between the two focal species. In the IPM, the effect of HSC abundance on REKN survival was trained using output from a Catch Multiple Survey Analysis (CMSA) as a proxy for female HSC abundance in Delaware Bay. In the CMSA, the HSC population was estimated on the basis of data from three trawl-based surveys (in addition to known sources of HSC mortality), conducted by Virginia Polytechnic Institute and State University (VT), New Jersey (NJ), and Delaware (DE), respectively.

I was able to obtain the survey records from each of the three Delaware Bay trawl surveys for reanalysis up to and including data from 2022. For my reanalysis, I only analyzed data on female HSCs due to their unique importance for REKNs. For each trawl survey, I generated a 'raw' annual catch-per-unit-effort (CPUE; often used as an indicator of abundance) by dividing the total number of female HSC captures by the total survey effort (generally reported as the length of seafloor surveyed). However, raw CPUE values do not control for other factors that can affect the number of expected HSC captures, such as time of year (seasonality), water temperature, salinity, depth, and dissolved oxygen. Therefore, comparing raw CPUE estimates across years can be misleading if (for example) the surveys were conducted at different seasons, or under disparate water temperatures or depths. To control for these unwanted effects, I used

generalized linear models (GLM) and generalized additive models (GAMs) to model the number of female HSCs captured in each trawl survey as a function of seasonality (Julian date), water temperature, dissolved oxygen, salinity, and depth, using an offset term to account for differences in survey effort (tow length) among surveys (e.g., Fig. 10). Nonlinear responses were accommodated with quadratic terms or spline fits. All models assumed a negative binomial error distribution and a log-link. Models were fitted in R using the package 'glmmTMB' (Brooks et al. 2017), with goodness-of-fit assessed using the 'DHARMa' package (Hartig and Hartig 2017).

In my reanalysis I attempted to replicate the biotic interaction reported by ASMFC using trawl-based indices of female HSC abundance. Specifically, I used the REKN banding and resighting records from 2003 to 2022 (including 6 years of additional data relative to the ASMFC model, which only used data from 2005 to 2018) to model REKN apparent survival as a function of HSC several trawl-based indices of HSC abundance: (1) the CMSA results reported by ASMFC, (2) raw (unadjusted) and adjusted indices of HSC abundance from the DE, NJ and VT trawl surveys, and (3) design-based estimates of HSC abundance derived from the VT trawl survey (Wong et al 2022). In my reanalysis, I used conventional capture-recapture methods (Cormack-Jolly-Seber; CJS) in addition to the open-robust-design (ORD) framework used by ASMFC to estimate the effect of these indices on REKN survival.

Despite running multiple analyses with alternative trawl-based indices, my reanalysis efforts have uncovered no conclusive link between REKN survival and any trawl-based index of HSC abundance (including the CMSA-based indices used by ASMFC) (Table 1). Neither classical capture recapture methods (CJS) nor ASMFC's ORD method yielded evidence for a positive HSC-REKN relationship. Notably, the model that most closely resembled ASMFC's model – using the ORD framework for parameter estimation and the CMSA results as a proxy for HSC abundance – indicated a statistically significant *negative* effect of HSC abundance on REKN survival (Table 1). This surprising result is likely to be a spurious correlation, and should not be interpreted to suggest that higher HSC abundance in Delaware Bay leads to lower REKN survival (higher mortality). Critically, this result demonstrates that ASMFC's documented relationship between REKN survival and HSC abundance (upon which this two-species ARM framework is based) is unstable, underscoring the tenuousness and uncertainty of this critical relationship. Interestingly,

this relationship could not be replicated even after (1) using the ORD parameter estimation framework and the code provided by the ASMFC modelers (2) reducing the dataset to cover the same period analyzed by ASMFC (2005 through 2018), (3) using the same CMSA-based estimates of female HSC abundance used by ASMFC, and (4) including the other time-varying covariates used in the ASMFC model (arctic snow cover and spawn timing). The instability of the HSC-REKN relationship reported by ASMFC suggests both that it is unlikely to be meaningful reflection of reality and that it is a poor foundation upon which to base a two-species risk assessment framework.

Table 1. Tests of alternative HSC abundance indices as drivers of REKN survival. Gray shading reflects non-significant results (95% CI overlaps zero, suggesting coefficient could plausibly be positive or negative), green shading reflects significant positive coefficients (more HSC implies higher REKN survival), light green shading represents weakly (marginally) significant positive coefficients, and red shading reflects significant negative coefficients (more HSC implies lower REKN survival).

HSC Abundance index	HSC survey type	Survival coef, CJS	Survival coef, ORD
CMSA (uses DE, NJ and VT), 2005-2018	Ocean Trawl	0.02 (-0.18 to 0.22)	-0.18 (-0.31 to -0.06)
Virginia Tech (VT), abundance estimate	Ocean Trawl	-0.03 (-0.21 to 0.23)	-0.24 (-0.45 to 0.00)
Virginia Tech, CPUE	Ocean Trawl	0.23 (-0.01 to 0.54)*	-0.19 (-0.37 to -0.02)
Virginia Tech, CPUE adjusted	Ocean Trawl	0.01 (-0.18 to 0.18)	0.08 (-0.12 to 0.26)
DE trawl, CPUE**	Ocean Trawl	0.01 (-0.14 to 0.18)	-0.14 (-0.31 to 0.04)
DE trawl, CPUE adjusted**	Ocean Trawl	-0.02 (-0.17 to 0.17)	-0.10 (-0.27 to 0.09)

NJ trawl, CPUE	Ocean Trawl	-0.05 (-0.31 to 0.15)	-0.16 (-0.32 to 0.10)
NJ trawl, CPUE, adjusted	Ocean Trawl	-0.07 (-0.28 to 0.09)	0.09 (-0.16 to 0.28)
Delaware Bay Spawning Survey	Beach survey	0.02 (-0.18 to 0.27)	-0.09 (-0.26 to 0.12)
NJ Surface Egg Density	Beach survey	-0.08 (-0.24 to 0.06)	-0.09 (-0.29 to 0.09)
NJ Surface Egg Density (NJ REKN data only)***	Beach survey	0.29 (0.07 to 0.52)	0.32 (0.01 to 0.58)

*This relationship has weak statistical support but could be interpreted as evidence for a positive effect of HSC abundance on REKN survival.

*** Data provided on Aug 10, 2023. This work does not represent the opinions of the State of Delaware, Delaware Department of Natural Resources and Environmental Control or Delaware Division of Fish & Wildlife

*** This analysis used REKN capture and recapture records from the NJ side of the bay, since surface egg density was only collected on the NJ side of the bay.

The lack of a demonstrable effect of trawl-based HSC indices on REKN survival (Table 1) likely reflects the weaknesses of these indices rather than the weakness of the underlying biotic interaction. Trawl-based surveys are highly imperfect snapshots of the population of HSCs inhabiting Delaware Bay, obscured by differing survey methodologies and poorly understood aspects of HSC ecology, including seasonal and daily activities, habitat preferences, and degree of clustering on the seafloor. Furthermore, trawl-based surveys ignore that REKNs and other shorebirds do not feed on HSCs directly, but instead use their eggs to fuel their northward migration; therefore, the utility of trawl-based indices may be further eroded as a useful metric by additional, poorly understood processes such as annual variation in the timing of HSC egg deposition and the processes that dislodge eggs from their clusters and thereby render the eggs accessible to shorebirds.

To assess the degree to which the Delaware Bay trawl survey results reflected signal

about true annual fluctuations in HSC abundance versus random noise (likely driven by unmodeled variations in survey conditions, HSC clustering and seasonal movements and other poorly understood aspects of HSC ecology), I tested for pairwise correlations of the raw and adjusted CPUE estimates. Pearson correlations among the raw and adjusted CPUE results ranged from 0 to 0.45 (Fig. 4). The only statistically significant correlation among the three surveys was between the unadjusted CPUE estimates for the NJ and DE trawl surveys. However, this relationship weakened to 0.16 and became inconclusive after controlling for seasonality and site conditions (Fig. 4). Overall, the correlation tests indicated that the results from the three trawl surveys are largely uncorrelated with one another (Fig. 4). Therefore, it is likely that the trawl survey results (and the resulting indices and estimates of HSC abundance) largely reflect factors unrelated to variation in the underlying HSC population.

If annual trawl-based estimates (and estimates derived from these surveys, like the CMSA) are largely uncorrelated with the underlying dynamics of the HSC population, REKN survival could conceivably be strongly correlated with true HSC abundance yet show little correlation with trawl-based HSC indices (Table 1). In this way, the use of trawl-based indices as a proxy for HSC abundance (e.g., in models of REKN survival) may severely misrepresent the true nature of the interaction between these two species. Overall, the results of my reanalysis indicate that trawl-based indices of HSC abundance are a noisy and unreliable indicator of annual fluctuations in the HSC population, and are likely an inadequate metric for quantifying the biotic interaction between REKNs and HSCs in Delaware Bay.
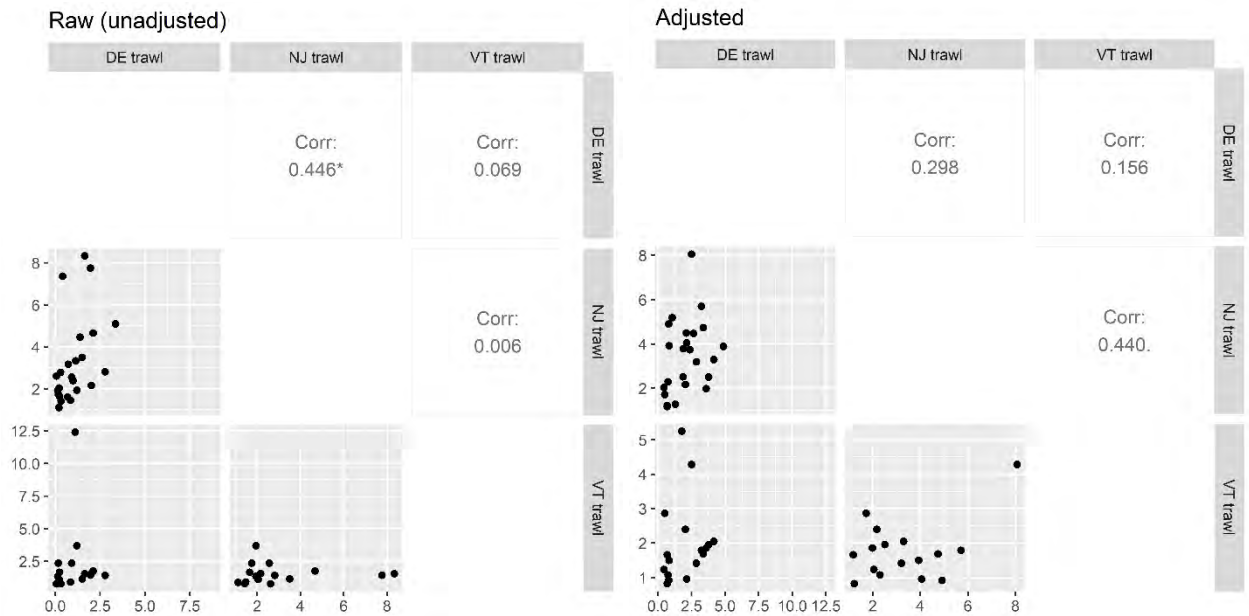
Figure 4. Scatterplot matrices (lower diagonals) and Pearson correlation tests (upper diagonals) for raw (left) and adjusted (right) catch-per-unit-effort (CPUE; HSC abundance indices) from three trawl-based surveys conducted in the Delaware Bay area from 1990 to 2022. Pearson correlations among the different trawl surveys ranged from 0 to 0.45. The only statistically significant correlation among the three surveys was between the unadjusted CPUE estimates for the NJ and DE trawl surveys. This relationship weakened to 0.16 after controlling for seasonality and site conditions. DE trawl data were provided on Aug 10, 2023. This work does not represent the opinions of the State of Delaware, Delaware Department of Natural Resources and Environmental Control or Delaware Division of Fish & Wildlife. Note that fulfillment of data requests does not constitute endorsement by the NJ Marine Resources Administration of any analyses or end products derived from the requested data.

3. REKN survival is strongly sensitive to HSC egg-density, indicating that persistent degradation of the HSC egg resource could have dire consequences for the REKN population

In contrast to trawl-based HSC survey data, surveys of HSC egg densities measured directly on the beaches used by REKNs and other shorebirds are likely to be a far more direct representation of the functional ecological link between these two species. Fortunately, such data are available: researchers have been consistently measuring the shallow-depth (0 to 5 cm)

density of HSC eggs in Delaware Bay (NJ side only) since 2000. While beach surveys (like all ecological data) are subject to sources of error that can obscure underlying signals, there are far fewer intermediate processes that may compromise the signal of the ecological relationship between these species. Although HSC egg surveys and spawning counts have been conducted in Delaware Bay for many years, ASMFC chose to use trawl-based surveys instead of surface egg density surveys to represent the HSC resource available to REKNs in their models (although they also used information on the timing of HSC spawning). To explain this decision, ASMFC has stated (1) that HSC abundance in Delaware Bay (CMSA model and results) has a clearer nexus with their management directive (ASMFC manages the HSC stock rather than the density of eggs deposited on beaches), and (2) that the egg data are highly variable across both space and time (seemingly making a case, without strong evidence, that the surface egg density surveys may be unreliable). Whatever their rationale for ignoring the long-term surveys of HSC surface egg-densities, it is misguided if it misrepresents the true nature of the underlying biotic interaction.

To evaluate the HSC surface egg density data as a proxy for the HSC egg resource available to migrating REKNs, I first reanalyzed the raw data to ensure that comparisons were valid across years for which survey methodologies differed. Overall, three different survey methodologies were used for measuring surface egg density during the period from 2000 to 2023. Although egg densities were always measured in the top 5 cm of the surface, the total area of beach surface measured per sample differed substantially among survey periods. To correct for these differences (effectively putting all samples on an even playing field) I used a modified version of the methods described in Smith et al. (2022) that included an offset term in the linear model formula. Briefly, I used generalized additive mixed models (GAMM) to model the number of eggs counted in each sample as a function of year (fixed effect) and seasonality (Julian day, using a smoothing spline to accommodate a non-linear functional response), with a random intercept term to accommodate for among-site variation, using an offset term (log of surface area sampled) to account for differences in survey effort (surveyed area) among samples. Following Smith et al. (2022), I assumed a negative binomial error distribution and a log link. Also following Smith et al. (2022), models were fitted in R, using the package 'glmmTMB' (Brooks et al. 2017), with goodness-of-fit assessed using the 'DHARMa' package (Hartig and Hartig 2017).

To assess the annual estimates of HSC surface egg density as a proxy for HSC egg resource availability in the REKN survival models, I used the annual adjusted surface egg density estimates as a covariate in the CJS and CMSA models. Since the HSC egg data were only collected on the NJ side of the bay, I only used REKN banding and resighting data from NJ for this analysis. The results of this analysis indicated a strong, positive effect of HSC density on REKN survival (Fig. 5). Years with high HSC egg densities were associated with mean REKN survival rates approaching 85%, whereas survival was reduced to approximately 65% in years with low HSC egg densities. Although these results were based on a standard Cormack-Jolly-Seber model for survival estimation, the open-robust-design model used by ASMFC yielded similar results, although with a wider range of parameter uncertainty (Table 1, section 2).

a (only NJ REKN data used)

b (only NJ REKN data used)
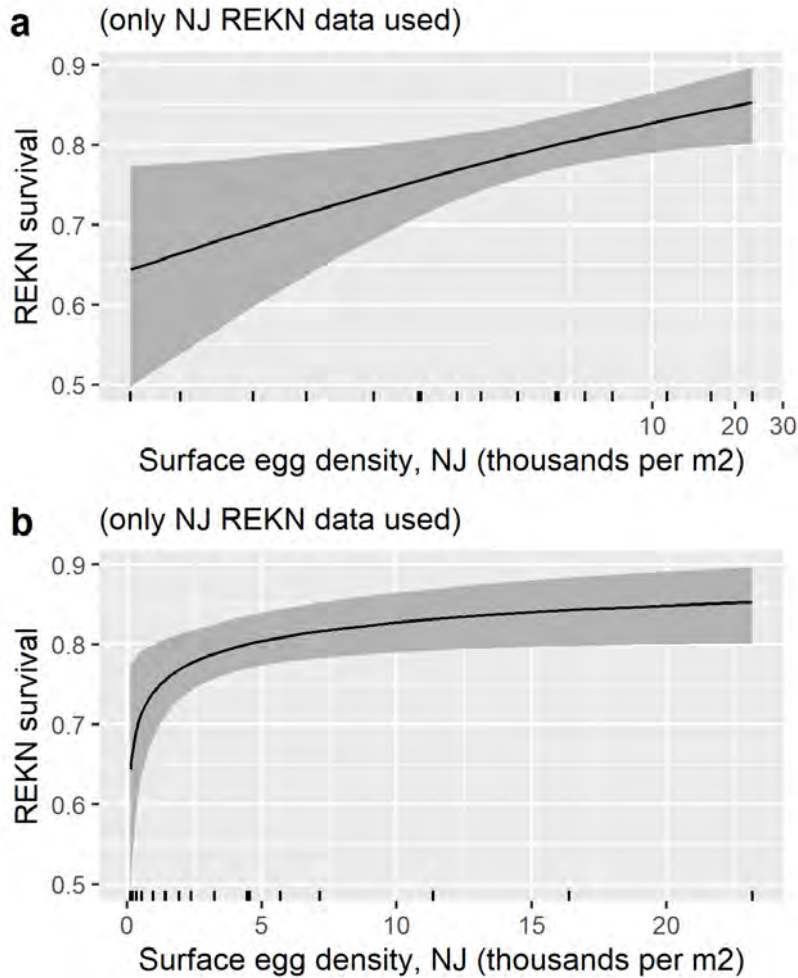
Figure 5. REKN survival as a function of the observed surface density of HSC eggs (thousands of eggs per m2, top 5 cm) on the NJ side of Delaware Bay. The top panel shows this relationship on the log scale (the scale at which the relationship was modeled), and the bottom panel shows the same relationship on the raw, untransformed scale. These results are derived from a Cormack-Jolly-Seber (CJS) model fitted to REKN banding and resighting data from 2003 to 2022. The rug (additional tick-marks along the x-axis) represents the observed egg densities during the study period. Since egg density data was not collected on the DE side of Delaware Bay, only birds resighted in NJ were used for this analysis.

In contrast to the HSC-REKN relationship used by ASMFC, under which the REKN population would be expected to increase even under the complete elimination of the HSC population in Delaware Bay (see attached), the effect of HSC egg density on REKN survival (Fig. 6)

forecasted a steep decline in the REKN population under sustained low densities of HSC eggs (held constant at the lowest observed levels from 2000 to 2022), resulting in near-extinction of the REKN population after 2-3 decades (Fig. 6). The magnitude of this relationship suggests that even 5 years of low HSC egg densities could result in a 50% decline of the REKN population. The strength of the estimated relationship between HSC egg densities with REKN survival is much more consistent (in comparison with the ARM framework) with the observed decline in the REKN population during the late 20th century, which is widely attributed to unregulated HSC harvest. Also in sharp contrast to the ASMFC model, the estimated relationship between REKN survival and HSC egg density indicate that sustained high HSC egg densities (held constant at the highest observed levels) can potentially promote the rapid recovery of the REKN population (Fig. 7).

It is important to recognize that the relationship between HSC abundance and HSC surface egg densities, which is critical for assessing the link between HSC harvest (which affects abundance) and REKN population persistence (which depends upon surface egg densities) remains unclear. Notably, surface egg densities are uncorrelated (in many cases, weakly negatively correlated) with the CMSA results and other trawl-based indices of HSC abundance (Fig. 8). Although knowledge of the link between HSC abundance and egg densities is clearly critical for managing the HSC stock in Delaware Bay, the true HSC abundance in Delaware Bay remains poorly characterized (see part 2, above), and the relationship between HSC abundance and the density of eggs accessible to shorebirds remains poorly understood. Therefore, caution should be used in interpreting any direct comparisons between models using HSC abundance versus egg density as a predictor variable (Figs. 6, 7), as these covariates are not strictly comparable. However, common sense dictates that there is a relationship between HSC egg availability and HSC abundance. Furthermore, the dependability of the egg resource year after year (and ultimately, the recovery of the REKN population) may require a "superabundance" of horseshoe crabs in Delaware Bay, ensuring an adequate supply of eggs available to REKNs even in years where environmental processes may be unfavorable to horseshoe crabs, the timing of their spawning, or the processes that dislodge eggs and make them available to foraging shorebirds. Finally, given the limited state of knowledge about the relationship between surface egg densities and HSC abundance, it is precautionary to assume a strong direct relationship whereby

lower HSC population numbers (e.g., via harvest or other anthropogenic sources of mortality) can reduce the number of HSC eggs available for shorebirds during the critical stopover period.

Finally, the results of this reanalysis strongly argue for continued rigorous monitoring of HSC surface egg densities at multiple beaches across Delaware Bay (on both the DE and NJ sides), as these data are critical for assessing the ecological link between HSCs and REKNs. By ignoring this source of data, ASMFC's revised ARM framework misrepresents the importance of the HSC egg resource to the REKN population and thereby underestimates the risk posed by HSC harvest to the long-term viability of the REKN population. By recommending harvest of female horseshoe crabs each year, the ASMFC's ARM framework has the potential to impede both the survival and the recovery of the REKN population.

Scenario: HSC population collapse
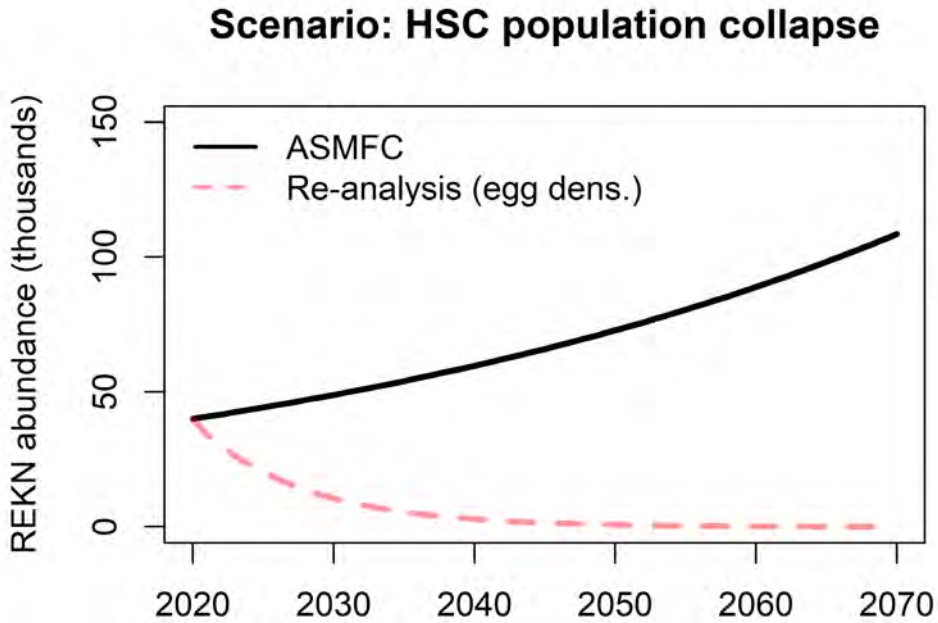
Figure 6. Results from 'back of the envelope' calculations of REKN population growth under scenarios with a depleted HSC population. The solid black line represents REKN abundance from 2020 through 2070 under the HSC-REKN relationship described in the ASMFC ARM framework, which was trained using the CMSA model as a proxy for the HSC egg resource in Delaware Bay. The numbers used for this calculation reflect the mean survival and fecundity values assuming a HSC population of zero. The dashed red line represents REKN abundance from 2020 through 2070 under a reanalysis in which the HSC-REKN relationship was trained using surface egg density data as a proxy for the HSC egg resource in Delaware Bay. In sharp contrast to the ASMFC model, the relationship fitted to the HSC egg density data indicate that collapse of the HSC population (here defined as the lowest observed annual surface egg density values) could easily drive the collapse of the REKN population in Delaware Bay. Note that this figure is based on a simple age-structured population model and does not incorporate a density-dependence mechanism (the revised ARM includes a density ceiling that prevents the REKN population from growing above ~150k).
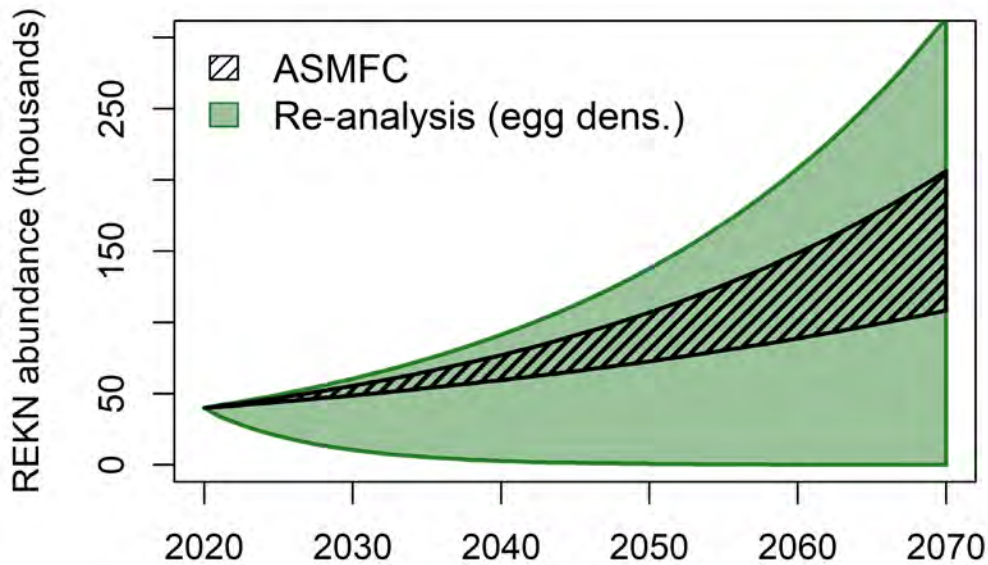
Figure 7. Results from 'back of the envelope' calculations of REKN population growth under scenarios ranging from a worst-case scenario of HSC population depletion (see Fig. 6) to a favorable scenario with constant HSC abundance/egg density at the highest levels observed from the early 2000s to present. The black hashed polygon (with diagonal lines) represents REKN abundance from 2020 through 2070 under the HSC-REKN relationship described in the ASMFC ARM framework, which was trained using the CMSA model as a proxy for the HSC egg resource in Delaware Bay. The light green polygon represents REKN abundance from 2020 through 2070 under a reanalysis in which the HSC-REKN relationship was trained using surface egg density data as a proxy for the HSC egg resource in Delaware Bay. In sharp contrast to the ASMFC model, this reanalysis indicates that HSC egg densities can strongly impact whether the population thrives (under consistently high surface egg densities) or declines to extinction (under consistently low egg densities). Note that this figure is based on a simple age-structured population model and does not incorporate a density-dependence mechanism (the revised ARM includes a density ceiling that prevents the REKN population from growing above ~150k).
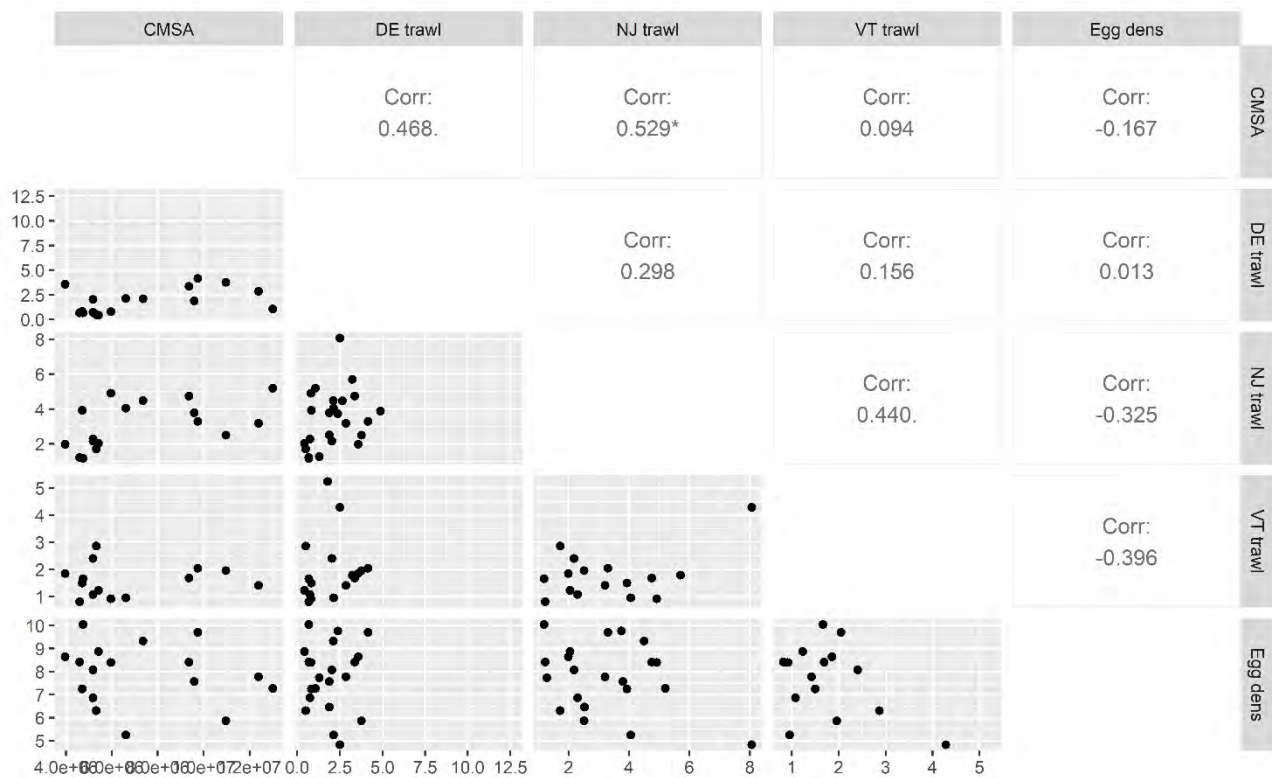
Figure 8. Scatterplot matrices (lower diagonals) and Pearson correlation tests (upper diagonals) for HSC abundance indices derived the CMSA model (used as an estimate of HSC abundance in the ARM framework), three trawl-based surveys conducted in the Delaware Bay area from 1990 to 2022 (used for training the CMSA model; adjusted for seasonality and survey conditions), and surface egg densities (NJ side only). The only statistically significant correlation among these five indices was between the CPUE estimates from the NJ trawl survey and the CMSA results. There was no apparent correlation between surface egg density measurement and any trawl-based index of HSC abundance (including the CMSA results). In fact, surface egg density had a weakly negative relationship with most trawl-based indices of HSC abundance. DE trawl data were provided on Aug 10, 2023. This work does not represent the opinions of the State of Delaware, Delaware Department of Natural Resources and Environmental Control or Delaware Division of Fish & Wildlife. Note that fulfillment of data requests does not constitute endorsement by the NJ Marine Resources Administration of any analyses or end products derived from the requested data.

4. The ARM exaggerates the evidence for an increasing trend in the number of female HSCs in

Delaware Bay

ASMFC used their CMSA model (which used the DE, NJ and VT trawl surveys as primary data sources) to claim that the HSC population in Delaware bay has been undergoing a recovery (population increase) during the period from 2003 to 2018 (ASMFC 2001). Furthermore, Smith et al. (2022) documented evidence for an increase in HSC surface egg densities during the same period. However, after controlling for potentially confounding factors like seasonality, water temperature, and differences in survey effort and methodology, neither the trawl-based surveys used by ASMFC nor the egg-density surveys show strong evidence for increasing abundance of female HSCs in Delaware Bay over the last 20 years.

The NJ trawl data provides an interesting case-in-point. The raw catch-per-unit-effort (CPUE) from New Jersey's ocean trawl survey (one of the major data sources used by ASMFC for documenting a positive trend in HSC abundance) appears to indicate increasing female HSC abundance from 2003 to 2022 (statistically significant at alpha=0.05; Fig. 9). However, when raw CPUE numbers are adjusted for strong effects of seasonality, water temperature, depth and dissolved oxygen on HSC captures in Delaware Bay (Fig. 10) the apparent positive trend in HSC CPUE disappears, becoming statistically inconclusive on the basis of a linear regression weighted by the inverse of sampling variance (Fig. 9). The values used by ASMFC to represent the NJ trawl data values in their CMSA model (which used only trawl data from April and August; ASMFC 2021) match closely with the unadjusted CPUE numbers (Fig. 9; results are similar using all months instead of only April and August trawl results), indicating that ASMFC's estimates of HSC population dynamics failed to control for differences in season and survey conditions. This result was consistent whether or not all NJ trawl results were used for model fitting or whether the data were filtered to include only the months used by ASMFC (April and August). Therefore, the increasing trend in the HSC population reported by ASMFC and used in the ARM may (at least in part) be an artifact of differing survey conditions (e.g., differences in trawl depth or water temperature) rather than evidence of recovery of the HSC population over time.

Similarly, when the egg-density data were adjusted for known differences in survey methodologies (primarily, differences in sampled area), the apparent positive trend in HSC

surface egg densities (Smith et al. 2022) became weak and inconclusive (note that this correction did not impact the estimated relationship between REKN survival and HSC egg-densities) (Fig. 11). Thus, my reanalysis of both the trawl-based surveys (Figs. 9, 10) and the egg-density surveys (Fig. 11) indicates that perceived positive trends in HSC population indices may reflect sampling differences and not trends in the underlying HSC population. Although these findings suggest the trend estimates reported by Smith et al. (2022) may be in error, this finding does not call other findings from Smith et al. (2022) into question, as these findings do not strictly depend upon the comparability of surface egg density samples collected during the study period.

Finally, I tested whether the aggregate evidence from the three trawl-based surveys (both adjusted and unadjusted; see part 2 of this report) showed evidence for HSC population recovery. Specifically, I ran linear regression models combining the CPUE estimates (both adjusted and unadjusted) from all three trawl surveys to assess evidence for an aggregate trend in abundance over time. Neither the raw HSC capture efficiencies (CPUE) from the trawl surveys nor the adjusted CPUE estimates showed conclusive evidence for a trend in HSC abundance over time (Fig. 12). With little correlation in inter-annual variation among trawl surveys (Figs 4, 8), years in which one trawl-based survey tended to indicate a large HSC population were rarely reinforced by the other surveys, resulting in a regression to the mean (Fig. 12).

Overall, the above results suggest that the ARM framework exaggerates the evidence for an increasing trend in female HSC abundance over the first two decades of the 21st century. In so doing, the ARM framework predicts recovery of the HSC population in Delaware Bay under a *status quo* scenario whereby HSC harvest regulations and other protections are maintained at current levels. In contrast, the results from my reanalysis suggest that the recovery of the female HSC population may require additional safeguards – including possibly decreasing harvest and continuing to improve and restore habitat at spawning beaches. Furthermore, by overstating the evidence for recent increases in the HSC population, ASMFC thereby likely underestimates both the vulnerability of the HSC population to harvest pressures in Delaware Bay and the potential carryover impacts on the REKN population.
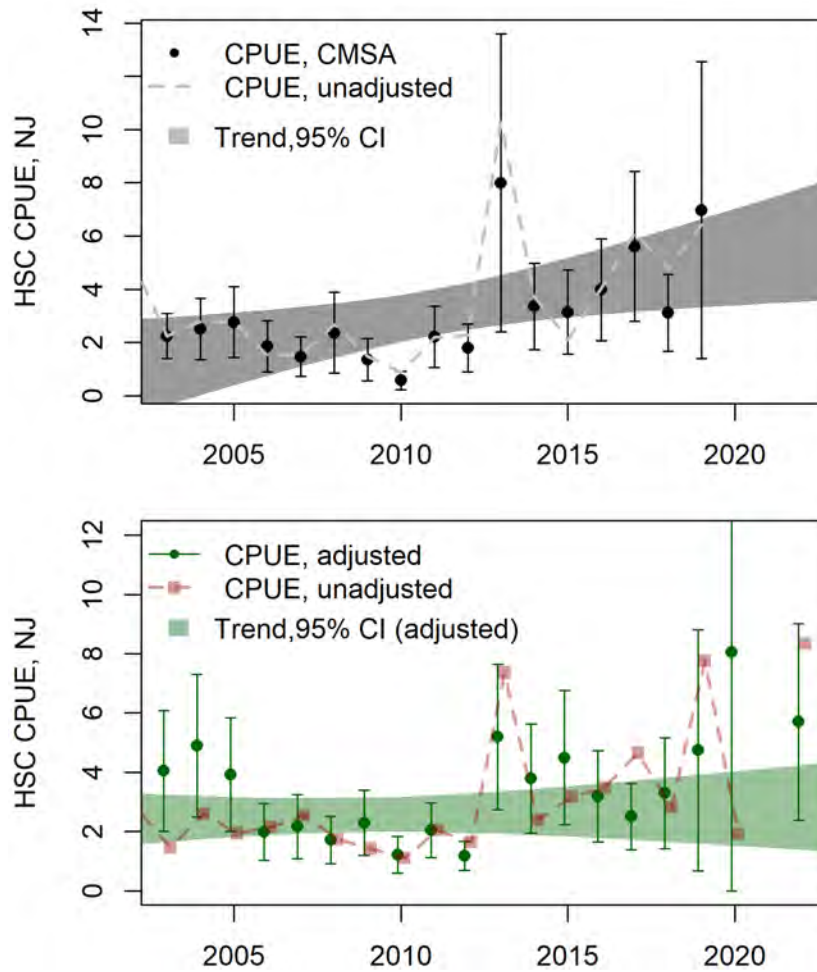
Figure 9. Annual HSC catch-per-unit-effort (CPUE; a type of abundance index) for trawl surveys conducted by the state of New Jersey from 1999 to present. The top figure compares the numbers used by ASMFC for their Catch Multiple Survey Analysis (CMSA) model (black), compared with the unadjusted, raw CPUE computed from the raw data (for comparison, only surveys conducted in April and August were used to compute CPUE; however, results look similar with raw CPUE for all months combined). The gray polygon represents the 95% confidence interval for the linear regression of the unadjusted CPUE against time in years. The bottom panel displays CPUE estimates adjusted for the effects of seasonality, water temperature, depth, and dissolved oxygen, with the dashed gray line and points again representing the (unadjusted, all months combined) CPUE computed from the raw data. Error bars represent 95% credible intervals. The green polygon represents the 95% confidence interval for the linear regression of the adjusted CPUE against time in years, showing no substantive trend over time.
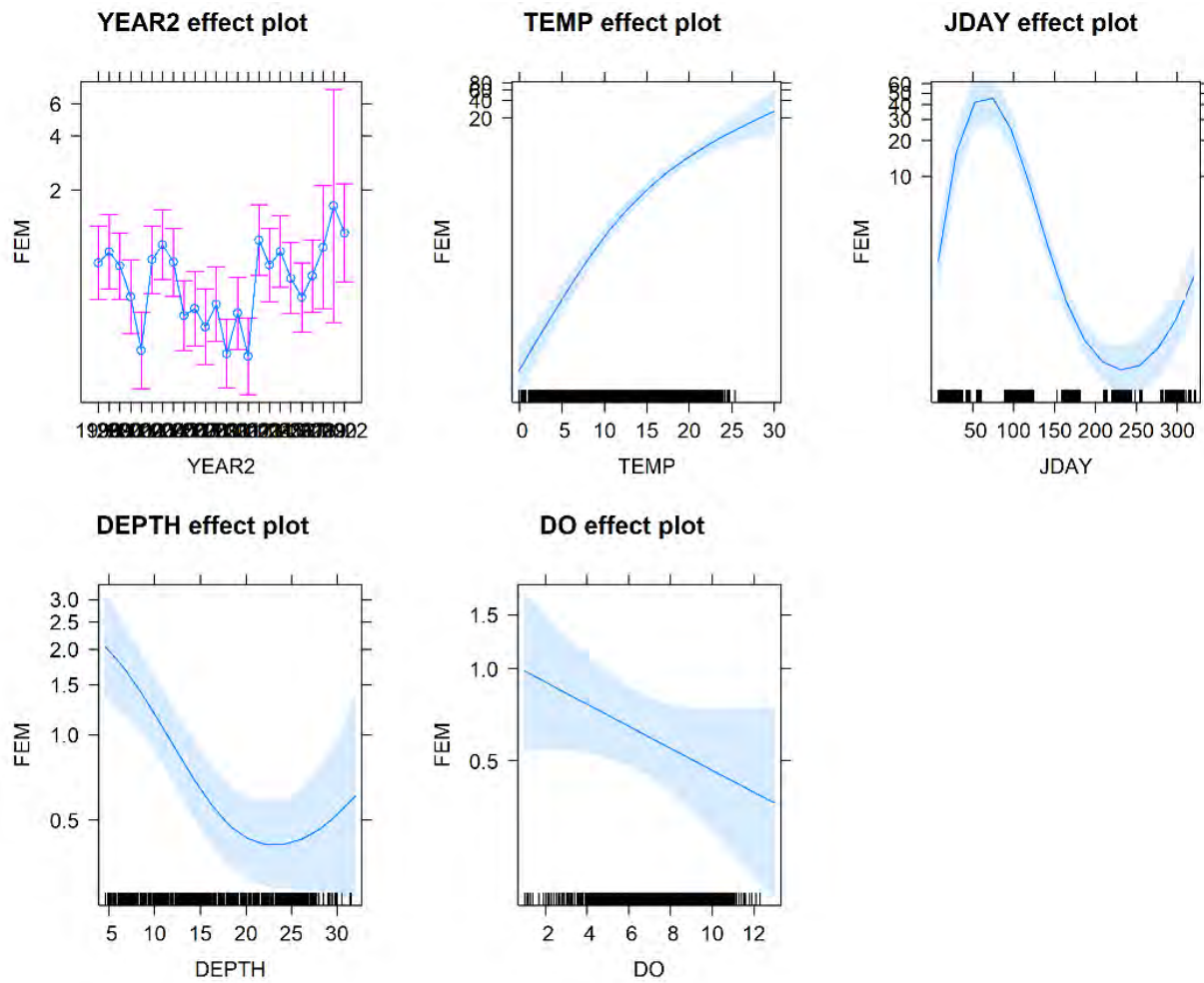
Figure 10. Effects plots illustrating strong, non-linear effects of season and environmental covariates (from top left to bottom right: year, temperature, Julian day, trawl depth, and dissolved oxygen) on the results of the ocean trawl surveys conducted in the Delaware Bay region by the state of NJ. These figures are predictions from a generalized linear model (GLM) using a negative binomial error distribution, quadratic terms to represent non-linear relationships, and an offset term to accommodate differing effort among surveys (amount of seafloor surveyed). The 'rug' on each plot illustrates the distribution of data for each quantitative covariate. Each panel represents the expected effect of a single predictor variable (indicated by the x-axis label), holding all other predictor variables at their mean or most frequent value. Therefore, although temperature and dissolved oxygen (DO) are closely linked, the DO effect plot illustrates the effect of DO after factoring out the effect of temperature.

31
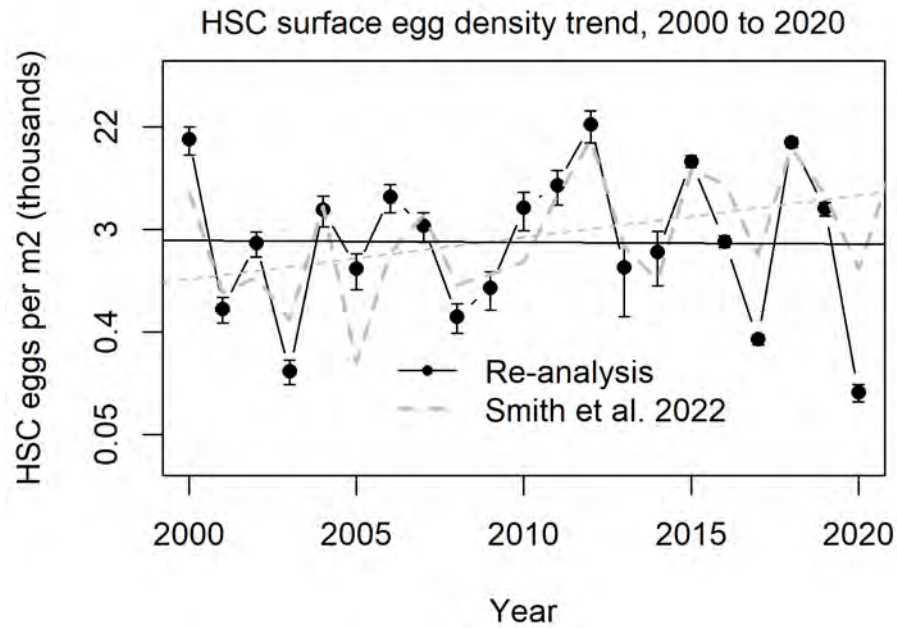
Figure 11. *Reanalysis of the evidence for a temporal trend in long-term surface egg density data from 2000 to 2020. Although the original analysis (Smith et al. 2022) detected a weak but non-negligible positive trend over time (dashed grey line), this regression relationship became inconclusive after accounting for differences in survey methodology across the 20 year study period (area represented by each sampling unit). Therefore, the increasing trend reported in Smith et al. (2022) appears to be an artifact of differing sampling methodologies used during this time frame.*
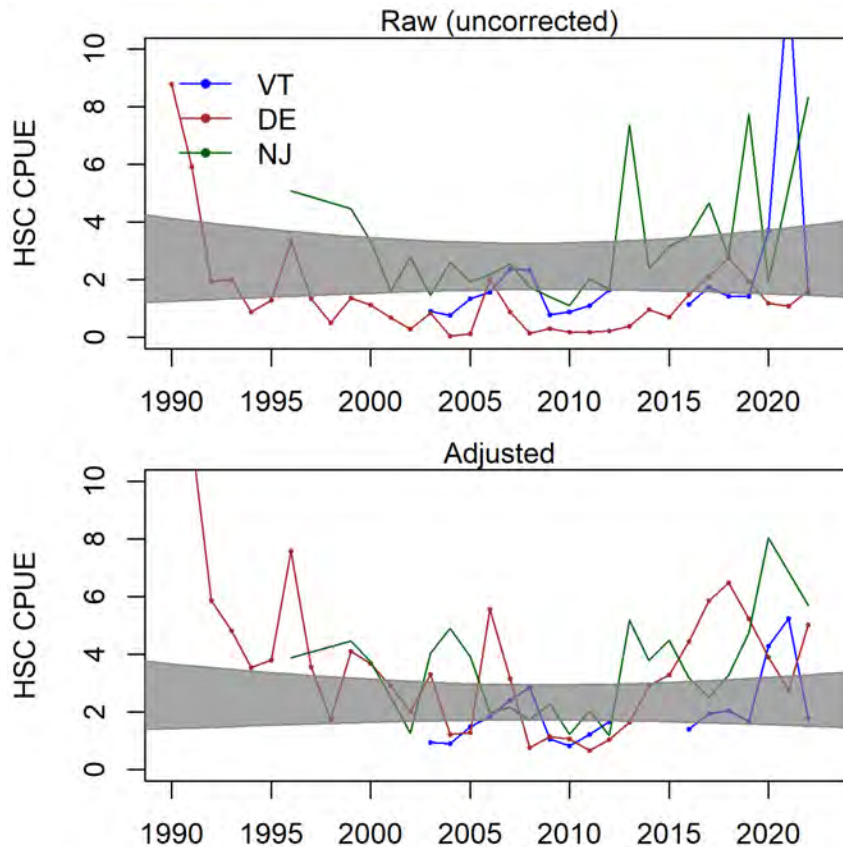
Figure 12. Raw and adjusted HSC catch-per-unit-effort (CPUE, which serves as an index of abundance) from three trawl surveys conducted in the Delaware Bay area from 1990 to 2022. Transparent gray polygons represent the 95% confidence region for a linear regression of CPUE (aggregated across the three surveys) across time. Top panel represents raw CPUE, whereas CPUE values in the lower panel are adjusted for the effects of seasonality, water temperature, salinity, dissolved oxygen, and depth. Taken in aggregate, the trawl data indicate an uncertain and variable population that is neither increasing nor decreasing over time. DE trawl data (Delaware Division of Fish & Wildlife, Delaware Department of Natural Resources and Environmental Control) were provided on Aug 10, 2023. This work does not represent the opinions of the State of Delaware, Delaware Department of Natural Resources and Environmental Control or Delaware Division of Fish & Wildlife. Note that fulfillment of data requests does not constitute endorsement by the NJ Marine Resources Administration of any analyses or end products derived from the requested data.

33

***5. The statistical model (IPM) used for estimating REKN population parameters is over-
parameterized and likely to yield spurious results***

Like many Integrated Population Models (IPMs), ASMFC's Red Knot IPM comprises two sub-
models: (1) a "state-space" model for estimating abundance and recruitment on the basis of
population counts over time, and (2) a model for estimating survival on the basis of capture-
recapture data (history of observation records for all uniquely marked individuals) (Schaub and
Kery 2021). In the REKN IPM, the state-space model is trained using annual 'peak count' data
(total number of REKNs observed during annual aerial and ground surveys), and the capture-
recapture model is trained using REKN banding and resighting records from Delaware Bay.
Whereas adult REKN survival (capture-recapture model) can be estimated directly from available
capture-recapture records (banding and re-sighting data from Delaware Bay), recruitment of
juvenile REKNs into the adult population (state-space model) is not directly estimable from the
peak-count data. Instead, the IPM estimates annual recruitment rates indirectly, as the offsets
required to match the observed dynamics of the peak-count data while accounting for expected
losses to mortality (the latter estimated from the capture-recapture sub-model).

While the data sources for training the capture-recapture model are information-rich
(tens of thousands of banding records and hundreds of thousands of resighting observations),
the peak-count data used by ASMFC to train the state-space model comprised a total of 14 data
points: one for each year from 2005 to 2018. Mathematically, this implies that these data could
be used to assign values to a maximum of 14 unknown parameters. However, with several
sources of 'noise' present in the data (sources of variation that obscure the important underlying
signals), these data are likely to support far fewer than 14 parameters. Some statisticians
informally recommend a rule of thumb of 30 data points per parameter for robust parameter
estimation; however, the optimal ratio differs depending upon many factors, including the signal-
to-noise ratio in the system as well as the risk tolerance of the researcher (Muthen and Muthen
2002). Nonetheless, the REKN IPM treats the peak-count dataset as a much richer source of
information than it actually is. In fact, the number of parameters estimated by the state-space
model exceeded the number of data points, resulting in a highly over-parameterized model that
is inherently prone to generating spurious results. Table 2 (below) enumerates the unknown

parameters estimated in the REKN IPM on the basis of the REKN peak-count dataset.

Table 2. *Free parameters ('unknowns') estimated using the peak-count data (n = 14) in the REKN IPM*

Description	Number of free parameters
Initial abundance	1
Annual recruitment	2 to 12*
Effect of HSC abundance on recruitment	1
Observation error, ground counts	1
Observation error, aerial counts	1
"Availability" parameters	12**
TOTAL	18 to 28

* Random effect
** Strong priors assigned

Taken together, the state-space model used in the REKN IPM estimated between 18 and 28 free parameters on the basis of 14 data points (Table 2). There are two reasons why it is not possible to pinpoint the exact number of free parameters estimated in this model. First of all, the state-space model includes a 'random effect' (representing annual recruitment of new REKNs into the breeding adult population) whereby 12 separate estimates of annual recruitment (12 parameters) are generated on the basis of a two-parameter Gaussian ('normal') distribution (mean and variance; known as 'hyperparameters'). Therefore, the number of free parameters used to estimate annual recruitment could be as high as 12 (number of annual recruitment estimates) or as low as 2 (number of 'hyper-parameters' used for generating the 12 annual estimates); the "truth" lies somewhere between those two extremes. Secondly, several parameters in the state-space model (notably, the 12 'availability' parameters, representing the fraction of the stopover population observable in the aerial and ground counts) were assigned relatively strong priors (in Bayesian inference, parameter estimates combine prior knowledge with additional knowledge inferred from the data). These strong prior distributions were

assigned to the 'availability' parameters on the basis of comparisons between the peak-count data and REKN abundance estimates generated annually as part of the ARM (Lyons 'superpopulation' models). Therefore, one could argue that the 12 'availability' parameters were not strictly 'free parameters' (or 'unknowns') since they were constrained by previous information from the Lyons models. However, my tests indicate that the 'availability' parameters remained sensitive to the peak-count data, and therefore it is more correct to treat these terms as free parameters ('unknowns') rather than as fixed parameters. Nonetheless, even in the most generous interpretation (~8-10 free parameters), the number of unknowns in the state-space model is far greater than the peak-count data (n = 14) could reasonably support, resulting in an over-parameterized model.

Models that fit more parameters than the data can support have a strong tendency to produce spurious results (results that fail to replicate when challenged with new data). Statisticians call such models "over-parameterized", or "over-fitted", and this problem is widely understood by quantitative researchers and statisticians (McNeish 2015). Among the free parameters estimated from this over-fitted model are several terms vital for understanding and simulating REKN population dynamics, including initial abundance, population trends (growth or decline), mean recruitment rate, and the effect of HSC abundance on recruitment. Due to over-fitting, these key parameters in the ARM model are likely to reflect random noise in the peak count data rather than the demographic reality of the REKN population.

To confirm the tendency of the REKN IPM to generate spurious results, I simulated artificial 'peak-count' data under a 'white noise' process (with no underlying trend) and assessed how often the IPM detected a spurious trend. To do this, I ran the IPM 50 times, each time replacing the REKN peak count data with random "white noise" generated using the same mean and variance as the observed peak-count data. Using the REKN abundance estimates from each of the 50 replicates, I ran a linear regression model with log transformed median REKN abundance as the response variable and time (year; 2003 to 2022) as a continuous predictor variable. For each replicate, I recorded whether the temporal trend of abundance over time was "significant" at alpha=0.05, along with the sign and magnitude of the inferred trend. As a second test, I ran 100 80-year projections (one set of projections for each of the 50 replicates) using the

time-varying survival and recruitment estimates from the IPM to project REKN abundance from 2023 to 2100 (propagating uncertainty using standard Bayesian demographic modeling techniques; Goodman 2002). Since the 'peak-count' data in these replicates were simulated with no underlying trend, the final abundance should match the initial abundance on average.

The results demonstrated that the IPM more often than not detected spurious temporal trends in REKN abundance (increases or declines in abundance over time) during the study period (Fig. 13). In fact, linear regressions (n = 50) fitted to the estimated log-median abundance from 2003 to 2022 indicated a non-negligible spurious temporal trend for 84% (42 of 50) of replicates at alpha = 0.05. Consequently, the results from projecting abundance forward to the year 2100 showed a strong tendency to erroneously produce estimates of final REKN abundance either much lower or much higher than the initial abundance (Fig. 14). Surprisingly, spurious negative trends were more common in my analysis than spurious positive trends in my analysis (Fig. 14). However, it is likely that this result is an artifact of the particular data simulation methods, model specification and initial values I used, and I caution against using this result to infer a systemic bias in the REKN IPM results. The apparent biases in my test results may be sensitive to many aspects of model specification, from the distribution and transformations used for simulating the peak-count data, to the prior distributions specified, to the initial values used for Markov-chain Monte Carlo (MCMC) simulations. Lacking access to the full modeling workflow used by ASMFC, I specified many of these parameters somewhat arbitrarily (lacking the bandwidth to complete a full sensitivity analysis). In addition, I modified the capture-recapture data to account for potential misread errors (see above), and this change could have potentially changed or even reversed any apparent biases in the modeling framework used by ASMFC. Therefore, additional sensitivity tests would be necessary to understand the conditions under which systemic biases may manifest in this modeling framework.

Due to over-parameterization, the REKN IPM is unstable and has a strong tendency to produce spurious results. Therefore, the REKN demographic simulations used in the ARM framework are unlikely to accurately capture the dynamics of the real-word population of *Rufa* Red Knots inhabiting Delaware Bay each Spring. Overall, the tendency of the REKN IPM to produce spurious results suggests that this model should not be used for assessing REKN

conservation status, running scenario tests, or guiding recovery efforts for a federally Threatened species.
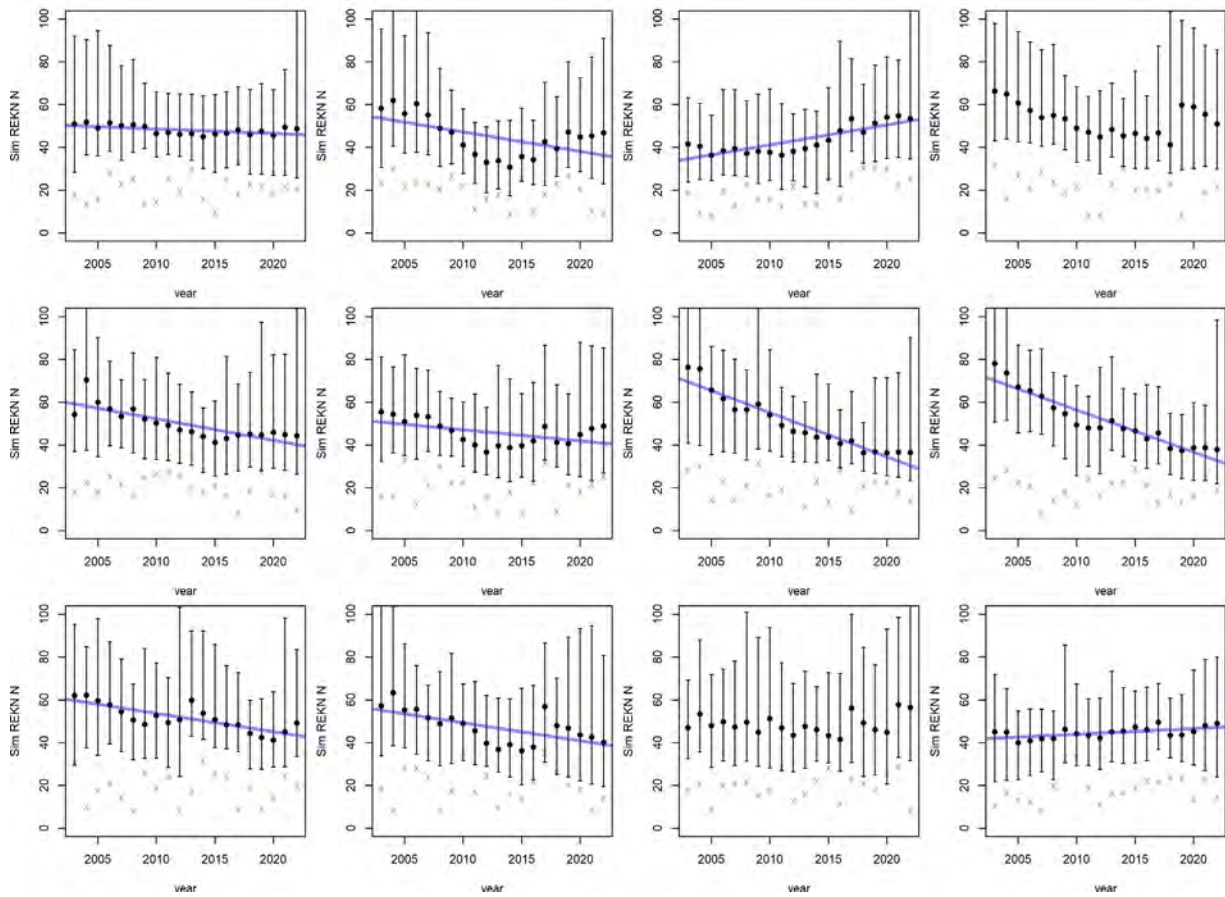
Figure 13. *Simulated REKN abundance (in thousands) over time for 12 replicates (randomly chosen from among 50) of the REKN IPM from 2003 to 2023 in which the peak-count data were replaced with random noise with no underlying trend (simulated data are represented by "X" symbols in the above panels). In many of these replicates, the IPM results detected a spurious trend over time (regression lines in the above panels) despite the lack of a trend in the count data.*
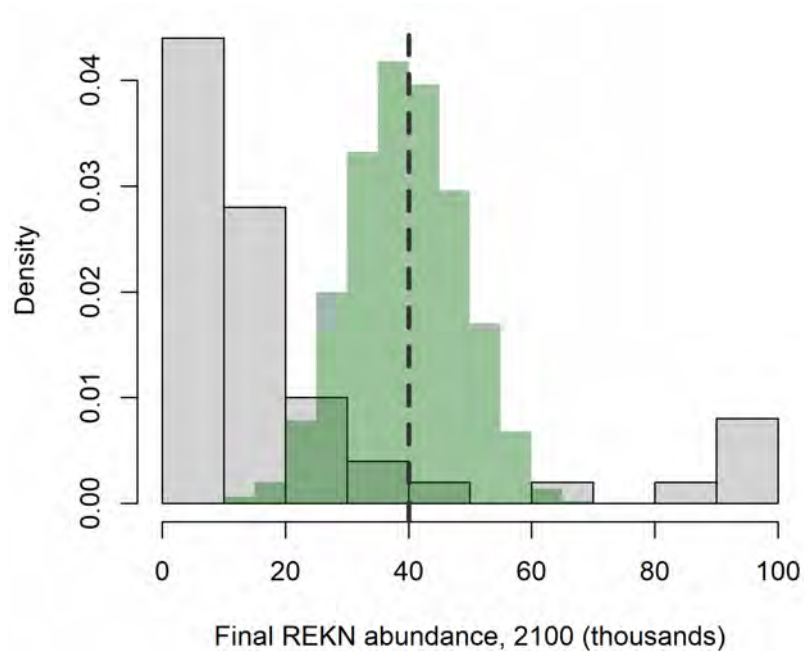
Figure 14. *Histogram of median simulated REKN abundance at year 2100, based on the results from 50 replicates of the REKN IPM whereby the 'peak-count' data were iteratively replaced by randomly generated white noise with the same mean and standard deviation as the observed peak-count data (bars with gray fill). The vertical dashed line represents the initial abundance for the simulations (40,000 REKNs). Since the peak-count data were simulated with no trend, the final simulated abundance should match the initial abundance on average – which in this case would imply a single peak centered on the initial abundance (green histogram). However, the peaks at abundances near zero and 100 indicate that many of these simulations (fitted to white-noise) spuriously projected either near-extinction or a full recovery of the population after 80 years. The fact that more replicates projected spurious declines versus spurious growth is likely to be an artifact of the simulations rather than a systematic bias inherent to the REKN IPM.*

6. The IPM exhibits poor fit to the available data

Goodness-of-fit (GOF) testing is a critical validation step in any model-fitting workflow (e.g., assessing the normality of residuals in linear regression), ensuring that key assumptions made during the modeling process are reasonable and justified (Conn et al. 2018). In the case of IPMs, simulation studies have indicated that indirect estimates of latent parameters (like recruitment rates in the REKN IPM) can be highly sensitive to model assumptions, and can produce biased

and nonsensical results if key assumptions are violated (Riecke et al. 2019; Schaub and Kery 2021). Therefore, it is critical to assess model goodness-of-fit (GOF) to assess whether IPM assumptions are reasonable (Conn et al. 2018). If an IPM fails to exhibits a reasonable fit to the data, key model parameters (like recruitment rates in the REKN IPM) should be used with extreme caution (Riecke et al. 2019).

In the context of hierarchical Bayesian analysis (the paradigm used by the ASMFC modelers), a commonly used approach is to run a Posterior Predictive Check (PPC), in which data are repeatedly simulated under the fitted model and compared to the actual data (Kery and Schaub 2011; Schaub and Kery 2021). If a model is unable to simulate data resembling the real-world observations, the model is determined to be an inadequate representation of the true processes that generated the data. 'Bayesian p-values' are often used to summarize GOF for IPMs, and represent the fraction of simulated datasets whose variance from expected values exceeds that of the true observations (Kery 2010). Whereas statisticians have noted that Bayesian p-values tend to understate a model's lack of fit (Conn et al. 2018), and research on assessing GOF for IPMs is ongoing, Bayesian P-values remain the most commonly used and reported GOF statistic for models like the REKN IPM (Schaub and Kery 2021).

In their ARM report, ASMFC mentions (but does not further document) two PPCs that were performed to assess goodness-of-fit for the IPM. One of these tests – the only test included in the publicly shared IPM code – uses a PPC to assess the degree to which the state-space model adequately represented the peak-count data. However, this test has been previously demonstrated to be an insufficient gauge of model adequacy (Schaub and Kery 2021). Furthermore, the over-parameterization of the state-space model (see above) virtually guarantees that the state-space model will pass this test (over-parameterized models tend to exhibit excellent fit to the observed data, although they tend to perform poorly in other contexts). The second and final goodness-of-fit test mentioned in the ARM report (which suggests mild lack of fit) is not included in the version of the IPM code shared publicly, so it is impossible to assess what test was actually run. However, I ran three additional PPCs to assess the degree to which the IPM (specifically, the open robust design component of the IPM) adequately represented the REKN resighting data from 2003 to 2022. These tests, which were

used and reported in an earlier version of the open robust design (ORD) model for estimating REKN survival and stopover use (Tucker et al. 2021), indicated poor fit to the data (Figure 15), suggesting that the IPM is an inadequate representation of key processes operating in the REKN population – including survival and recruitment. The failure of the IPM to pass rigorous goodness-of-fit tests casts additional doubt on conclusions generated from this model.
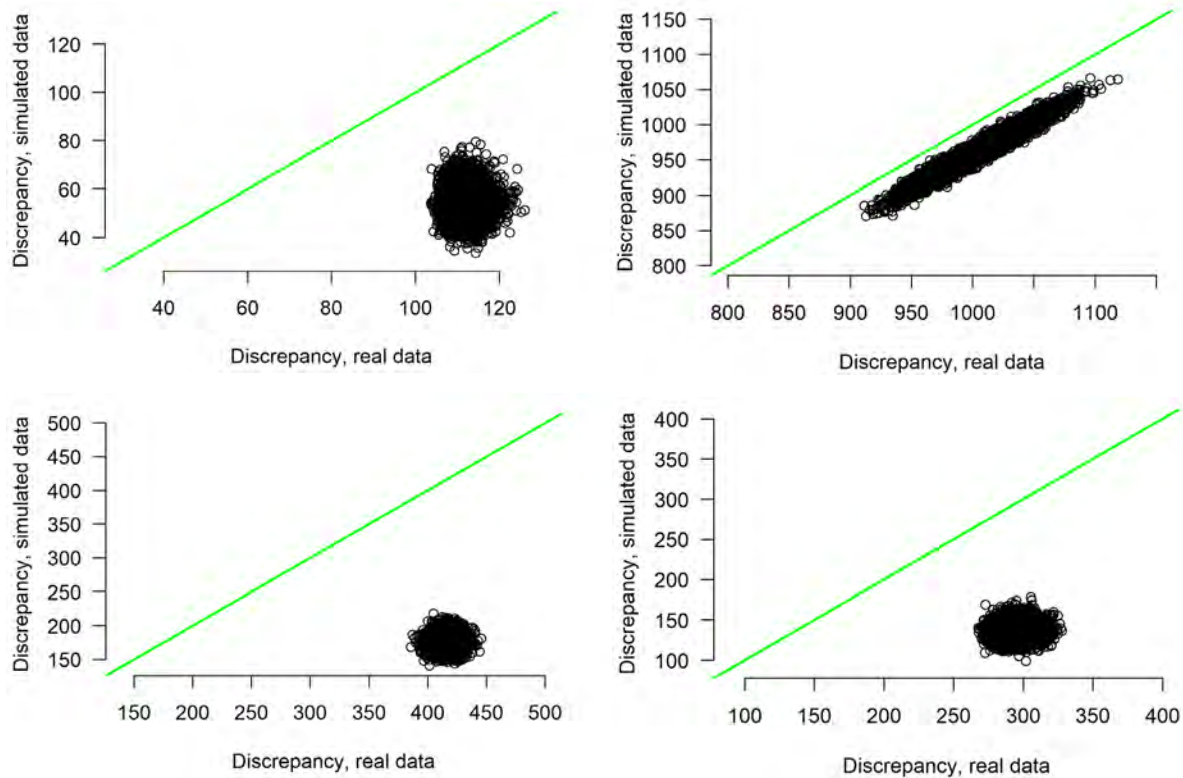
Figure 15. *Four goodness of fit (GOF) tests for the open robust design (ORD) component of the REKN IMP. GOF test #1 (upper left) assesses the adequacy of the survival and temporary emigration parameters, and is therefore the most directly relevant to the REKN population model. The remaining tests assess model fit to the timing of arrival within each year (upper right), numbers of 'transients' observed during each 3-day survey period (lower left), and recaptures of non-transients during each 3-day survey period (lower right). Bayesian p-values for all tests are equal to 1, indicating severe over-dispersion of the data relative to model predictions.*

CONCLUSION

Building on the issues identified in my 2022 review of this ARM framework, I have outlined six additional concerns about the validity of ASMFC's revised ARM framework as a tool for assessing and managing the risks to the *Rufa* Red Knot posed by the horseshoe crab harvest in Delaware Bay. First, I demonstrated that a major component of the Integrated Population Model (used for modeling REKN population dynamics) is severely over-parameterized and prone to generating spurious results. Second, I presented evidence that ASMFC's estimates of REKN survival were biased high due to failure to account for misread errors, thereby artificially inflating the resilience of the REKN population to short-term fluctuations in recruitment. Third, my reanalysis showed that trawl-based indices of HSC abundance – and the CMSA model used by ASMFC for estimating HSC abundance dynamics – have no conclusive relationship with REKN survival. Fourth, I showed that HSC surface egg density has a strong relationship with REKN survival, suggesting that ASMFC is strongly underestimating the strength of the biotic interaction and the dependency of REKNs on HSC eggs for population survival and recovery. Fifth, I show that the ARM exaggerates the evidence for an increasing trend in the number of female HSCs in Delaware Bay, thereby likely over-estimating HSC population resilience to harvest pressure. Finally, I present evidence that ASMFC's model of REKN population dynamics exhibits poor fit to the data, casting additional doubts on the validity of the ARM's model of REKN population dynamics. Based on these concerns, I conclude that this ARM framework is not useful for managing risk to the REKN population due to HSC harvest. Furthermore, my results suggest that the revised ARM misrepresents the importance of HSCs to the REKN population and thereby underestimates both the existential risk to the REKN population posed by female HSC harvest and the potential for promoting REKN recovery through increased regulatory protections and conservation efforts aimed at promoting HSC population increases in the Delaware Bay region.

Referenced literature:

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

Baker, A. J., Gonzalez, P. M., Piersma, T., Niles, L. J., de Lima Serrano do Nascimento, I., Atkinson, P. W., ... & Aarts, G. (2004). Rapid population decline in red knots: fitness consequences of decreased refuelling rates and late arrival in Delaware Bay. Proceedings of the Royal Society of London. Series B: Biological Sciences, 271(1541), 875-882.

Brooks, M. E., Kristensen, K., Van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., ... & Bolker, B. M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. The R journal, 9(2), 378-400.

Conn, P. B., Johnson, D. S., Williams, P. J., Melin, S. R., & Hooten, M. B. (2018). A guide to Bayesian model checking for ecologists. Ecological Monographs, 88(4), 526-542.

Cooch, E. (2008). Program MARK," A gentle introduction". http://www. phidot. org/software/mark/docs/book/.

Dietze, M. C. (2017). Ecological forecasting. Princeton University Press.

Goodman, D. (2002). Predictive Bayesian population viability analysis: a logic for listing criteria, delisting criteria, and recovery plans. Population viability analysis, 447, 454.

Hartig, F., & Hartig, M. F. (2017). Package 'DHARMa'. R package.

Kellner, K., Meredith, M., & Kellner, M. K. (2019). Package 'jagsUI'. A Wrapper Around'rjags' to Streamline'JAGS'Analyses. R Package Version, 1(1).

Kéry, M. (2010). Introduction to WinBUGS for ecologists: Bayesian approach to regression, ANOVA, mixed models and related analyses. Academic Press.

Kéry, M., & Schaub, M. (2011). Bayesian population analysis using WinBUGS: a hierarchical

perspective. Academic Press.

Koons, D. N., Riecke, T. V., Boomer, G. S., Sedinger, B. S., Sedinger, J. S., Williams, P. J., & Arnold, T. W. (2022). A niche for null models in adaptive resource management. Ecology and evolution, 12(1), e8541.

Lovich, J. E., Ennen, J. R., Yackulic, C. B., Meyer-Wilkins, K., Agha, M., Loughran, C., ... & Madrak, S. (2015). Not putting all their eggs in one basket: bet-hedging despite extraordinary annual reproductive output of desert tortoises. Biological Journal of the Linnean Society, 115(2), 399-410.

McNeish, D. M. (2015). Using lasso for predictor selection and to assuage overfitting: A method long overlooked in behavioral sciences. Multivariate behavioral research, 50(5), 471-484.

Méndez, V., Alves, J. A., Gill, J. A., & Gunnarsson, T. G. (2018). Patterns and processes in shorebird survival rates: a global review. Ibis, 160(4), 723-741.

Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. Structural equation modeling, 9(4), 599-620.

Nichols, J. D., Runge, M. C., Johnson, F. A., & Williams, B. K. (2007). Adaptive harvest management of North American waterfowl populations: a brief history and future prospects. Journal of Ornithology, 148(2), 343-349.

Niles, L. J., Bart, J., Sitters, H. P., Dey, A. D., Clark, K. E., Atkinson, P. W., ... & Veitch, C. R. (2009). Effects of horseshoe crab harvest in Delaware Bay on red knots: are harvest restrictions working?. BioScience, 59(2), 153-164.

Piersma, T., Lok, T., Chen, Y., Hassell, C. J., Yang, H. Y., Boyle, A., ... & Ma, Z. (2016). Simultaneous declines in summer survival of three shorebird species signals a flyway at risk. Journal of Applied Ecology, 53(2), 479-490.

Plummer, M. (2012). JAGS Version 3.3. 0 user manual.

Pradel, R., & Sanz-Aguilar, A. (2012). Modeling trap-awareness and related phenomena in capture-recapture studies. PloS one, 7(3), e32666.

Radosavljevic, A., & Anderson, R. P. (2014). Making better Maxent models of species distributions: complexity, overfitting and evaluation. Journal of biogeography, 41(4), 629-643.

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.

Regan, H. M., Colyvan, M., & Burgman, M. A. (2002). A taxonomy and treatment of uncertainty for ecology and conservation biology. Ecological applications, 12(2), 618-628.

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

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

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2016). False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant.

Smith, J. A., Dey, A., Williams, K., Diehl, T., Feigin, S., & Niles, L. J. (2022). Horseshoe crab egg availability for shorebirds in Delaware Bay: Dramatic reduction after unregulated

horseshoe crab harvest and limited recovery after 20 years of management. Aquatic Conservation: Marine and Freshwater Ecosystems, 32(12), 1913-1925.

Tucker, A. (2019). Stopover ecology and population dynamics of migratory shorebirds. Dissertation. Auburn University.

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

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

USFWS 2014. Endangered and Threatened Wildlife and Plants; Threatened Species Status for the Rufa Red Knot. Federal Register / Vol. 79, No. 238

USFWS 2021. Endangered and Threatened Wildlife and Plants; Designation of Critical Habitat for Rufa Red Knot (*Calidris canutus rufa*). Federal Register / Vol. 86, No. 133

Williams, B. K. (2011). Resolving structural uncertainty in natural resources management using POMDP approaches. Ecological Modelling, 222(5), 1092-1102.

Wong, C. Jiao, Y. and Hallerman, E. Report to the Atlantic States Marine Fisheries Commission Horseshoe Crab and Delaware Bay Ecology Technical Committees (ASMFC, Arlington, VA, 2018). 27 p.

SUPPLEMENTAL ANALYSES

Note on the incorrect specification of the "pi" parameter in the REKN IPM

Although not directly related to any of the six primary critiques in this report, it is nonetheless important to note here that the "*pi*" parameter in the REKN IPM, which represents the fraction of the flyway population that is present at the stopover site at any given 3-day time window, is incorrectly specified in the ARM model. This parameter is used internally within the IPM for adjusting the total estimated flyway abundance to reflect the number of REKNs using the stopover at the time of the peak count surveys. Therefore, this parameter provides a critical link between the open robust design model and the state-space model within the IPM, helping to refine estimates of REKN abundance and population trends.

In the REKN IPM, the computation of *pi* follows two steps. First, for each 3-day occasion during the stopover period each year, the probability of being present in the stopover (conditional on using the stopover at least once that year) is computed using the *delta* (arrival) terms, the *tau* (stopover residency) and the *psi* (stopover retention/persistence) parameters. This derived term, which appears to be performed correctly, is called *alpha* in ASMFC's IPM code. To

45

compute *pi* (proportion of the entire flyway present at the stopover site during each period) from *alpha* (the proportion of stopover users present at the stopover site each period) we just need to multiply alpha by a factor representing the fraction of flyway individuals using the stopover each year (we will call this factor "*z*"). ASMFC computed '*z*' as the sum of two parameters from the open-robust-design (ORD) model: '*gammaII*' and '*gammaOI*', which represent the probability of returning to the stopover (conditional on having been there last year), and the probability of returning to the stopover (conditional on having NOT been in the stopover last year), respectively. Importantly, the '*gammaII*' and '*gammaOI*' parameters are conditioned on two distinct segments of the flyway population; these parameters have no meaning when added together. For 'gammaII' and 'gammaOI' to have meaning at the level of the flyway population, we would need to know the fraction of the flyway population that used the stopover last year, which we call '*f*'. With this information, we could compute *z* and *pi* as:

z = gammaII(f) + gammaOI*(1-f)*   (Correct formulation)

*pi = z * alpha*

Multiplying this term (*z*) by *alpha* would yield the appropriate estimate of *pi*. However, ASMFC computed the *z* parameter as:

z_i = *gammaII + gammaOI*   (Incorrect formulation)

pi_i = z_i * *alpha*

Since z_i does not have meaning as a probability (this quantity can theoretically exceed 1), the resulting estimate of *pi* has no discernible meaning. Since *pi* is used to make the link between the annual peak counts and true flyway abundance, this error may introduce another source of bias in the estimates of REKN abundance and growth rate derived from the IPM. Although this is an important error, likely to have implications for the IPM results and the ARM framework, I consider this issue secondary in importance to the over-parameterization of the state-space model.

Note on over-parameterization of ASMFC's REKN survival model

In contrast to the ORD model, the 'classical' Cormack-Jolly-Seber (CJS) framework yielded estimates of the REKN-HSC relationship that were neither positive nor negative (inconclusive; Table 1). The increased tendency of the ORD model to yield conclusive (but negative)

46

relationships may be a consequence of the increased complexity of the ORD model versus the CJS models, as more complex models have a greater tendency to generate spurious results. Furthermore, there is reason to suspect that the ASMFC model of REKN survival tried to estimate more parameters than the data could support. With 14 years of data used for training the ASMFC model (2005 to 2018), there are 13 years for which survival is theoretically estimable (one fewer than the years of data; Cooch 2008). In the IPM, these 13 estimable rates represent the degrees of freedom (independent information used for parameter estimation) needed for modeling annual variation in REKN survival. In ASMFC's IPM, these 13 data points are used to estimate no fewer than five parameters: (1) the effect of HSC abundance on REKN survival, (2) the effect of spawn timing on REKN survival, (3) the effect of arctic snow cover on REKN survival, (4) an interaction between HSC abundance and spawn timing, and (5) a temporal process variance that allows survival to vary among years. Fitting five parameters using 13 degrees of freedom (a ratio of 2.6 data points per free parameter) suggests that this model (like the model of REKN recruitment; see above) is prone to over-fitting and thereby producing spurious results (see above).

The model instability that is characteristic of over-fitted models is apparent in the estimation of the effect of trawl-based HSC indices on REKN survival. Notably, when I specified the ORD model with the full set of time-varying covariates used by ASMFC – including HSC abundance derived from the CMSA model, the fraction of HSCs spawning in May, arctic snow cover, and an interaction between HSC and HSC spawn timing -- the previously significant (and nonsensical) negative relationship between HSC abundance and REKN survival disappeared (B= -0.04, 95% CI: -0.20 to 0.08). This relationship remained inconclusive regardless of whether potential misread errors were included in the model training set.


Potential biases due to over-representation of Mispillion harbor in the REKN resighting dataset
Tabular summaries of the number of observations by site and by state exposed a strong over-representation of a single study site (Mispillion harbor, in DE) in the REKN resighting dataset, raising concerns that patterns in the REKN survival results used for the ARM framework may represent the idiosyncrasies of a single site rather than general patterns across Delaware Bay

(Fig. S1). In fact, some Delaware Bay shorebird experts indicated to me that Mispillion harbor likely has a greater concentration of HSC eggs than many other sites and tends to support rapid weight gain in REKNs, which could induce lower mortality rates. To test this, I ran multiple models of REKN survival – including the ORD formulation used by ASMFC in addition to simpler Cormack-Jolly-Seber (CJS) models – using data sets excluding Mispillion harbor and including only data from Mispillion harbor. Overall, I found that mean REKN survival estimates were very similar for birds captured inside and outside of Mispillion harbor. However, patterns in survival among years showed some marked differences that could potentially indicate different drivers of survival inside and outside of Mispillion harbor (Fig. S2). In particular, survival for birds captured and resighted in Mispillion harbor was more stable across years, yet showing a slight declining trend. In contrast, survival for birds captured and resighted outside of Mispillion harbor was more variable (showing a strong reduction in 2010 and 2017), with no apparent trend over time.
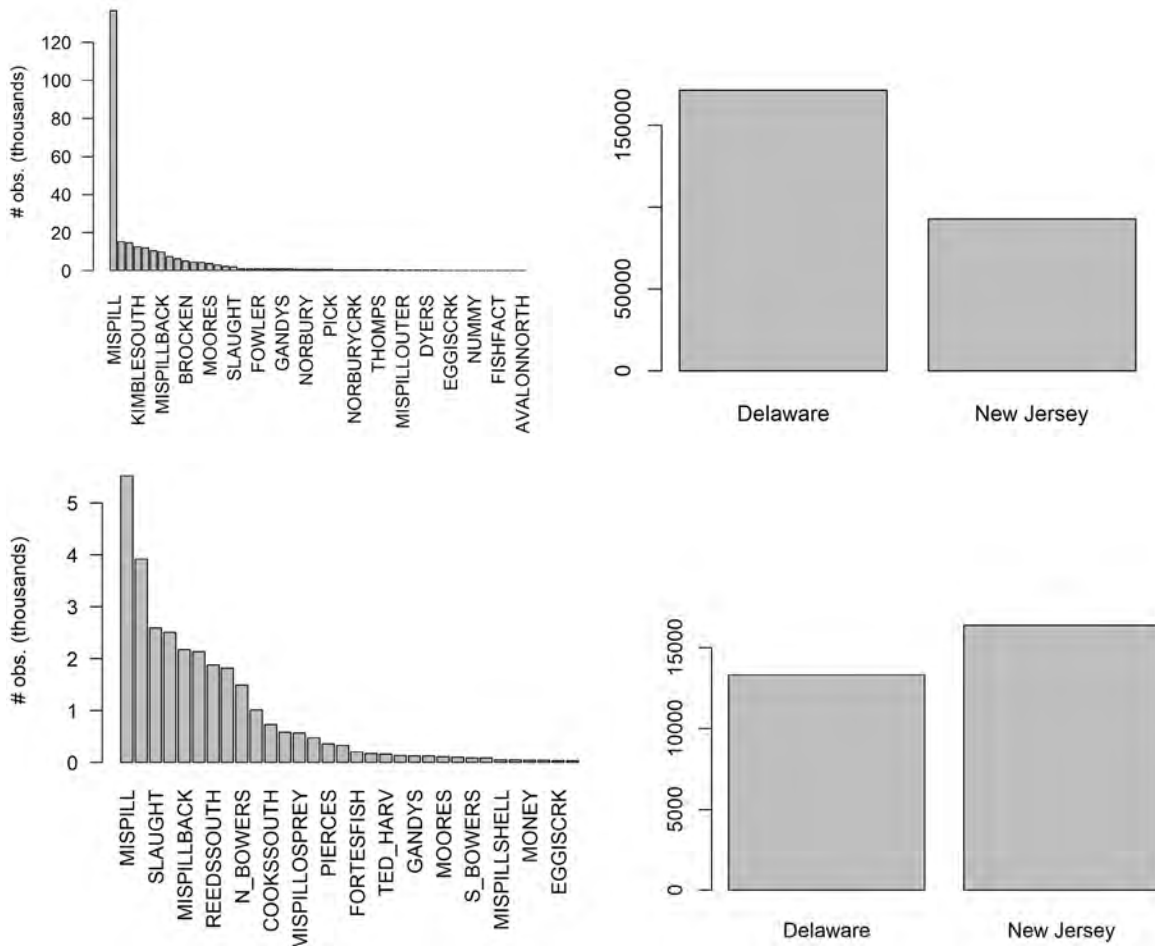
Figure S1. Top panels: number of resighting observations per site and by state. Note that resighting observations within Mispillion harbor ("MISPILL" in the above figures) far outweigh all other sites, leading to some concern that analysis results may be biased if this site differs from other sites. Delaware (which is dominated by Mispillion harbor data) has about 2x the number of resighting observations than NJ. Bottom panels: banding data summary by site and by state. As opposed to the resighting data, there are more banding records from New Jersey, and Mispillion harbor does not dominate the banding records to the same degree as it does the resighting data.
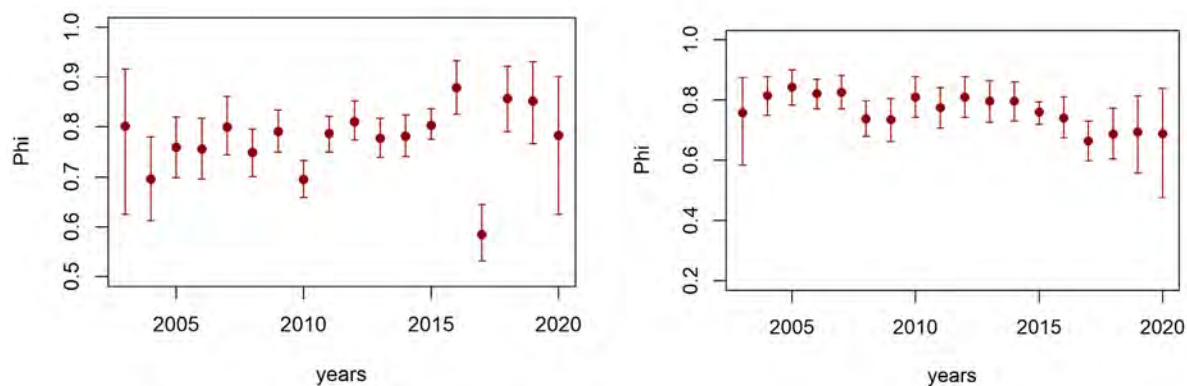
Fig. S2. *Comparison of annual REKN apparent survival estimates using (left) only birds marked and re-sighted outside of Mispillion harbor, and (right) only birds marked and re-sighted inside Mispillion harbor. Only birds first captured in Delaware Bay were included in the analysis, following ASMFC's stated data protocols. This figure illustrates different temporal patterns in survival, with REKN survival showing little trend outside of Mispillion harbor and decreasing slightly for birds captured and resighted inside Mispillion harbor. Outside of Mispillion harbor, estimated apparent survival was particularly low for two years: 2010 and 2017. Both models indicated reasonable goodness of fit.*

Trawl-based indices of HSC abundance

The figures below are a supplement to section 5 of this report, which documents that the evidence for a recent increase in the HSC population in Delaware Bay may be overstated. The figures below illustrate my efforts to generate adjusted indices of HSC abundance from trawl surveys to control for factors known to influence HSC capture rates: seasonality, trawl depth, salinity, and temperature (note that dissolved oxygen also emerged as an important factor in the New Jersey trawl surveys; Fig. 10).
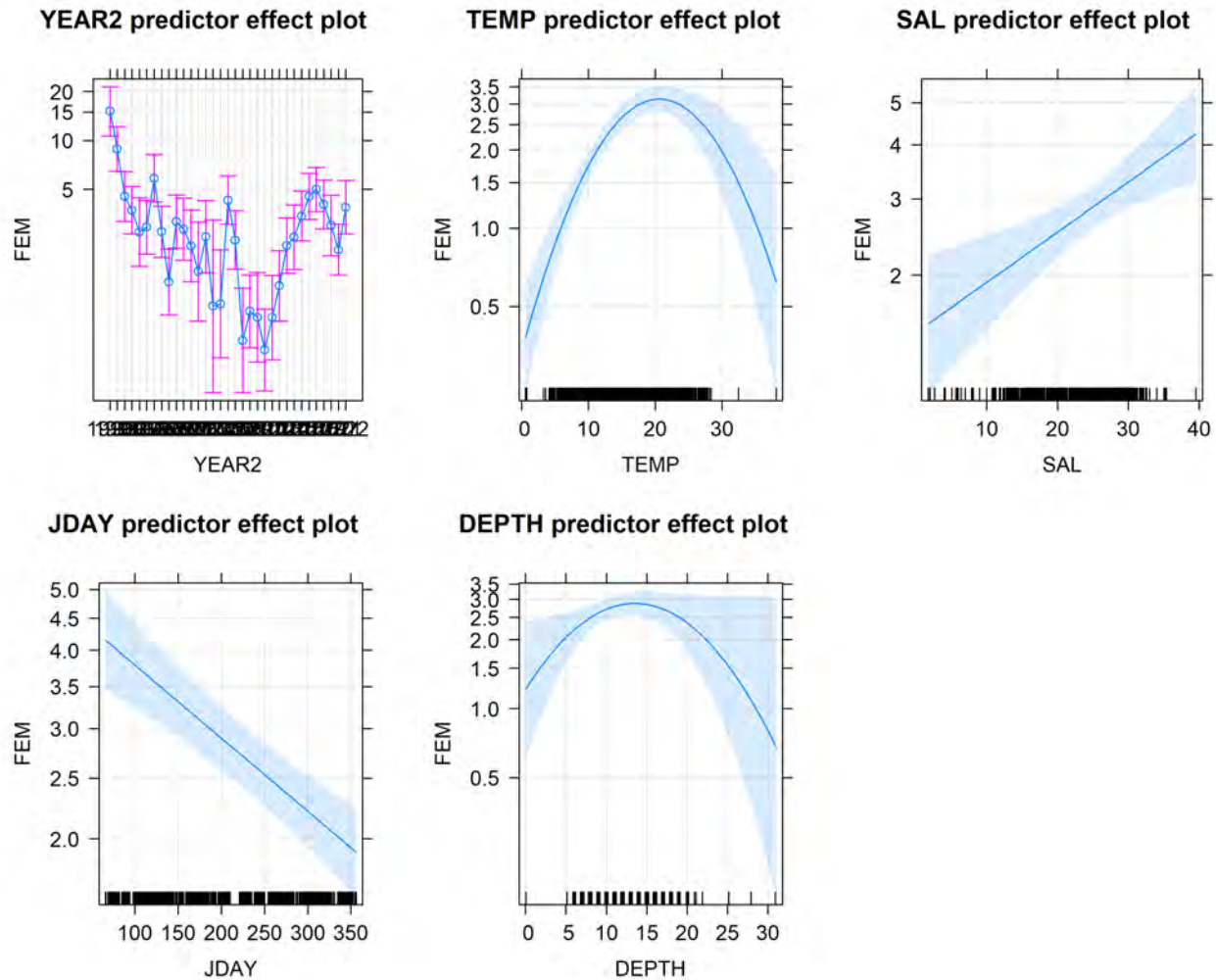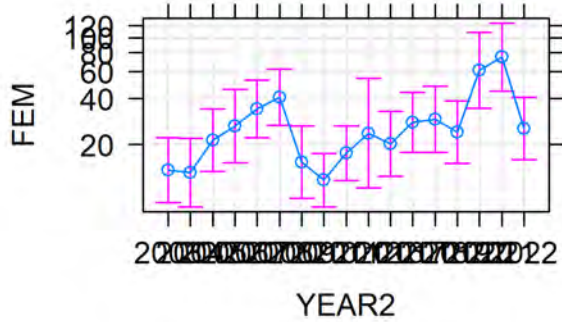
Figure S3. Effects plots illustrating strong linear and non-linear effects of season and environmental covariates (from top left to bottom right: year, temperature, salinity, Julian day, and trawl depth) on the results of the Delaware Bay trawl surveys conducted by the state of DE. These figures are predictions from a generalized linear model (GLM) using a negative binomial error distribution, quadratic terms to represent non-linear relationships, and an offset term to accommodate differing effort among surveys (amount of seafloor surveyed). The 'rug' on each plot illustrates the distribution of data for each quantitative covariate. DE trawl data were provided on Aug 10, 2023 by Delaware Division of Fish & Wildlife, Delaware Department of Natural Resources and Environmental Control. This work does not represent the opinions of the State of Delaware, Delaware Department of Natural Resources and Environmental Control or Delaware Division of Fish & Wildlife
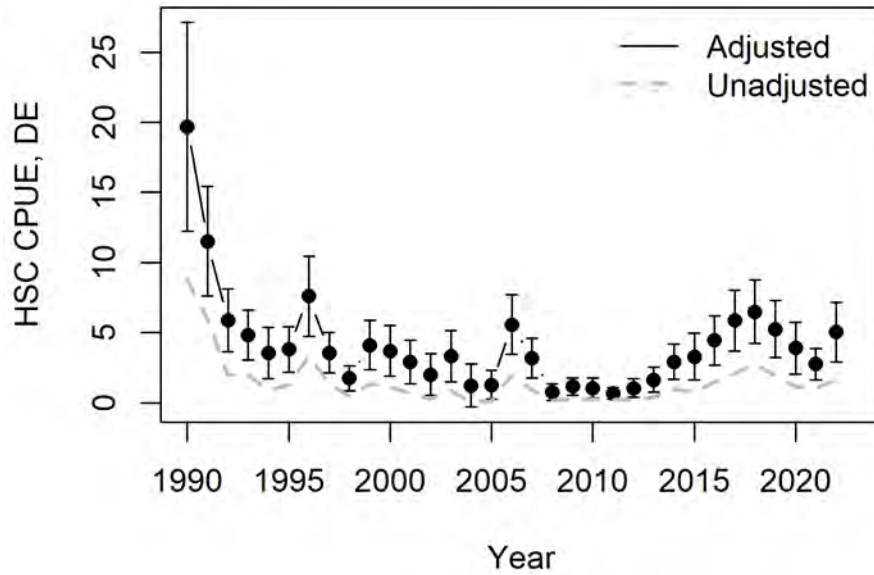
Figure S4. *Annual HSC catch-per-unit-effort (CPUE; a type of abundance index) for trawl surveys conducted by the state of Delaware from 1990 to present. Solid black dots are adjusted for the effects of seasonality, water temperature, depth, and salinity, while dashed gray line represents the unadjusted CPUE. Error bars represent one standard error on either side of the adjusted CPUE estimate. Unlike for the NJ data, the correction does not alter the general pattern of HSC abundance versus the unadjusted CPUE. DE trawl data were provided on Aug 10, 2023 by Delaware Division of Fish & Wildlife, Delaware Department of Natural Resources and Environmental Control. This work does not represent the opinions of the State of Delaware, Delaware Department of Natural Resources and Environmental Control or Delaware Division of Fish & Wildlife*

YEAR2 predictor effect plot

TEMP predictor effect plot
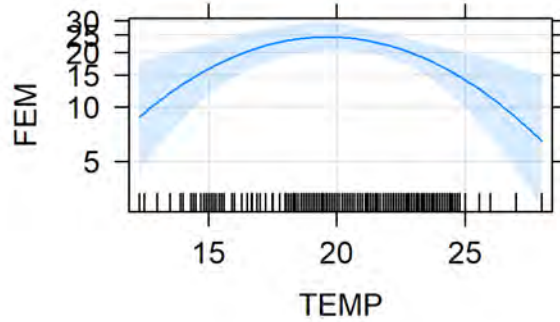
DEPTH predictor effect plot
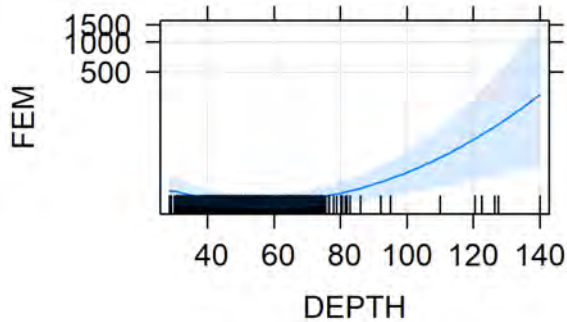
Figure S5. *Effects plots illustrating strong effects of year and environmental covariates (temperature and trawl depth) on the results of the Delaware Bay trawl surveys conducted by Virginia Tech (VT). These figures are predictions from a generalized linear model (GLM) using a negative binomial error distribution, quadratic terms to represent non-linear relationships, and an offset term to accommodate differing effort among surveys (amount of seafloor surveyed). The 'rug' on each plot illustrates the distribution of data for each quantitative covariate.*
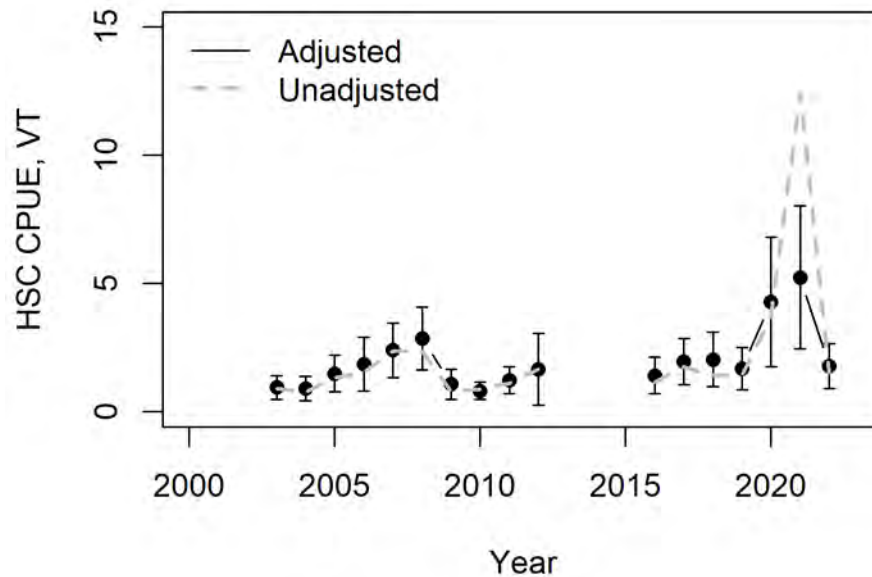
Figure S6. *Annual HSC catch-per-unit-effort (CPUE; a type of abundance index) for trawl surveys conducted by Virginia Tech from 2003 to present. Solid black points are adjusted for the effects of seasonality, water temperature, and depth, while dashed gray line represents the unadjusted CPUE. Error bars represent one standard error on either side of the adjusted CPUE estimate. Unlike for the NJ data, the correction does not generally alter the pattern of HSC abundance versus the unadjusted CPUE.*

Open robust design (ORD) validation tests

In this section, I report validation tests for assessing the ability of the open robust design (ORD) model to estimate the known values of key parameters (like survival) from simulated band-resighting data. In general, the ORD model successfully recovered the true parameters used to simulate the data, indicating that this model was correctly specified and capable of estimating parameters correctly. Overall, the ORD model was able to estimate many parameters related to survival, temporary emigration, the timing of stopover arrivals and departures, and detection probability (Figs S7-10). However, while the ORD model appears to perform well in simulation tests, recall that goodness of fit (GOF) tests showed that this model was not an adequate representation of the observed REKN data from Delaware Bay (see section 6, above). In addition, issues with potential misread errors further compromised the validity of the results (see above).
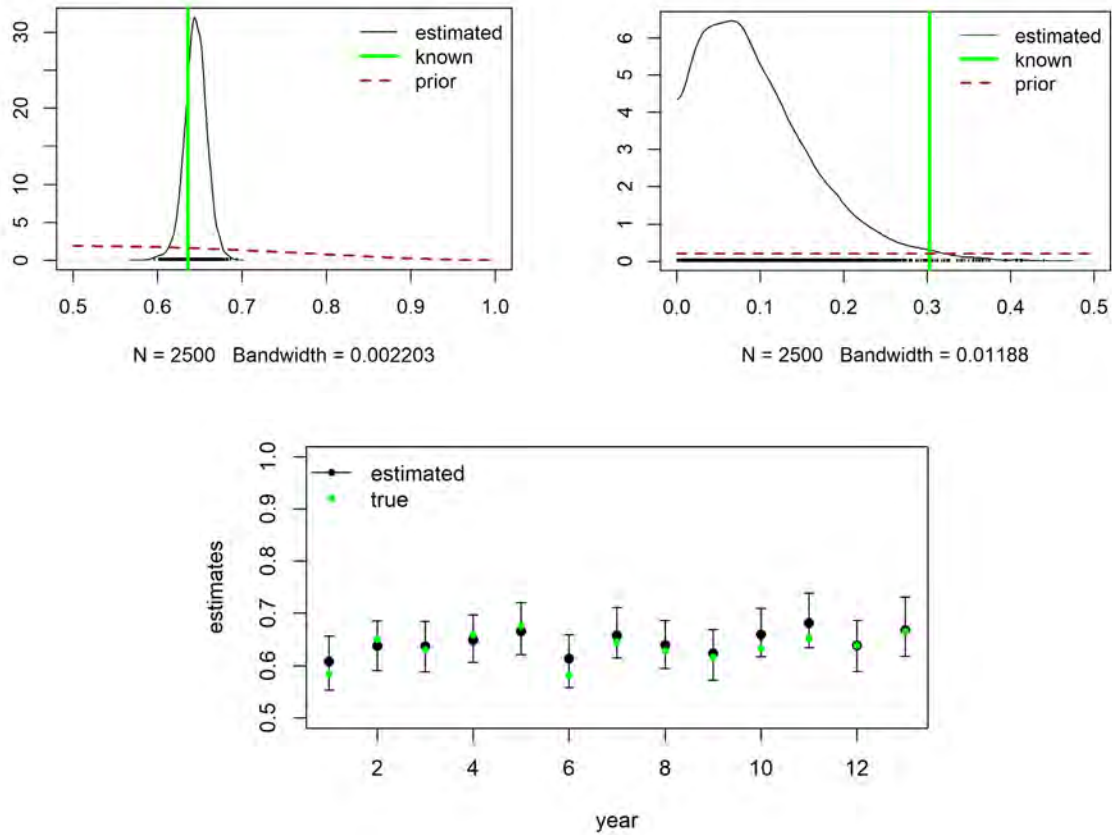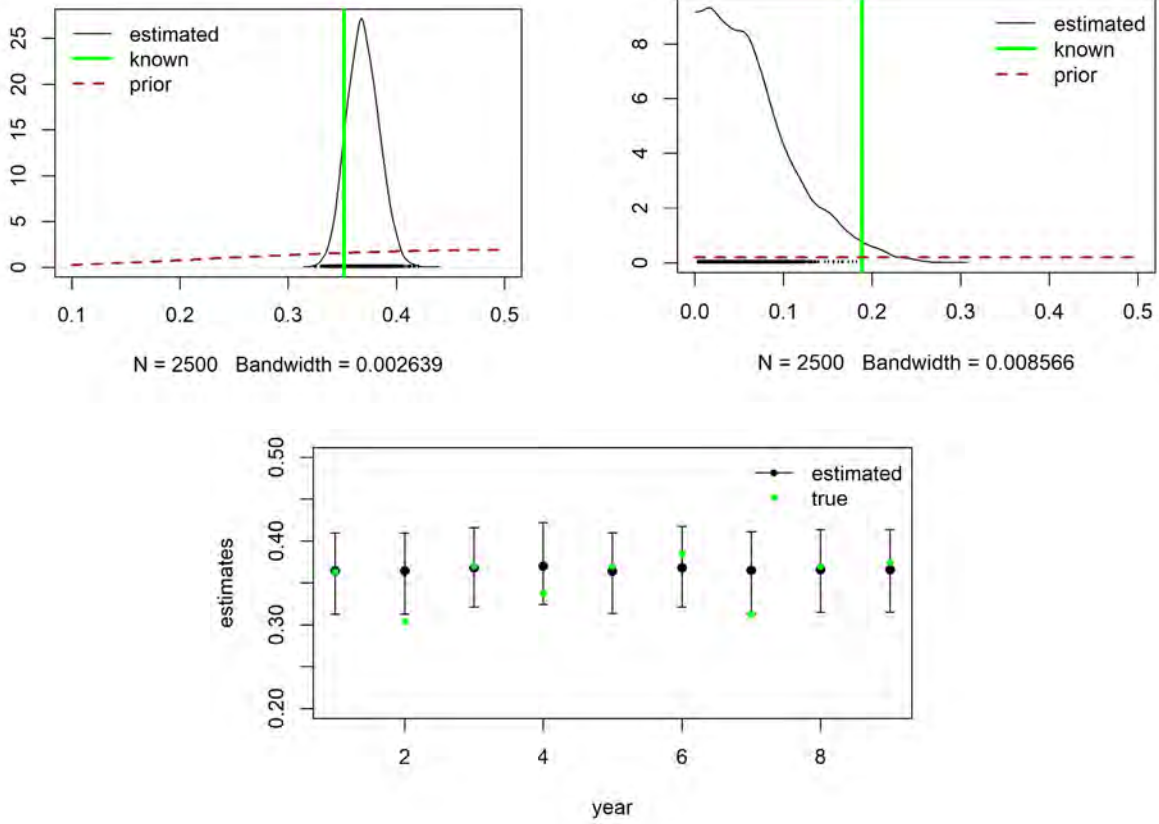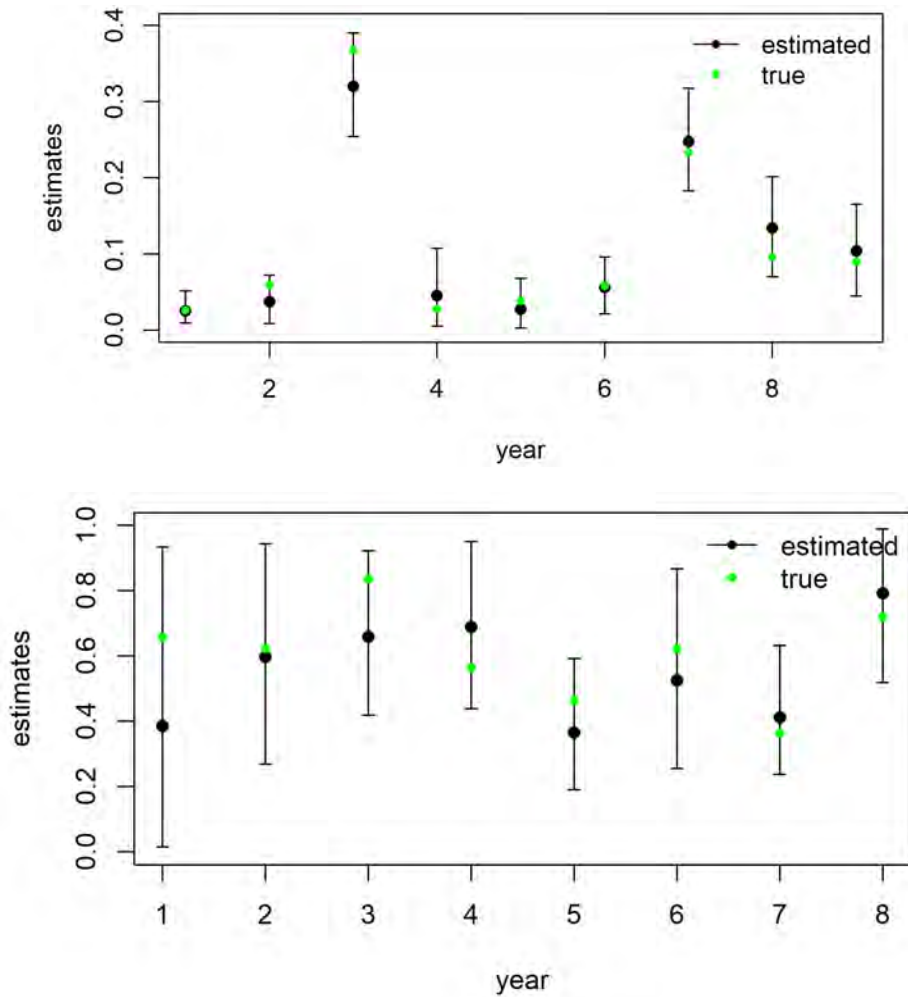
Figure S7. *In simulation tests, the ORD model seems to do a good job of recovering true mean survival (top left) from simulated data. The ORD model frequently fails to capture the true variance in survival (top right), leading to some concern about its ability to model annual variation in survival. However, the model performs well in capturing true annual survival values (bottom). Green dots and vertical lines represent the true values used in simulations, black curves, points, and confidence intervals represent parameter estimates from the ORD model, and dashed brown curves represent the prior probability distributions used for Bayesian model fitting.*
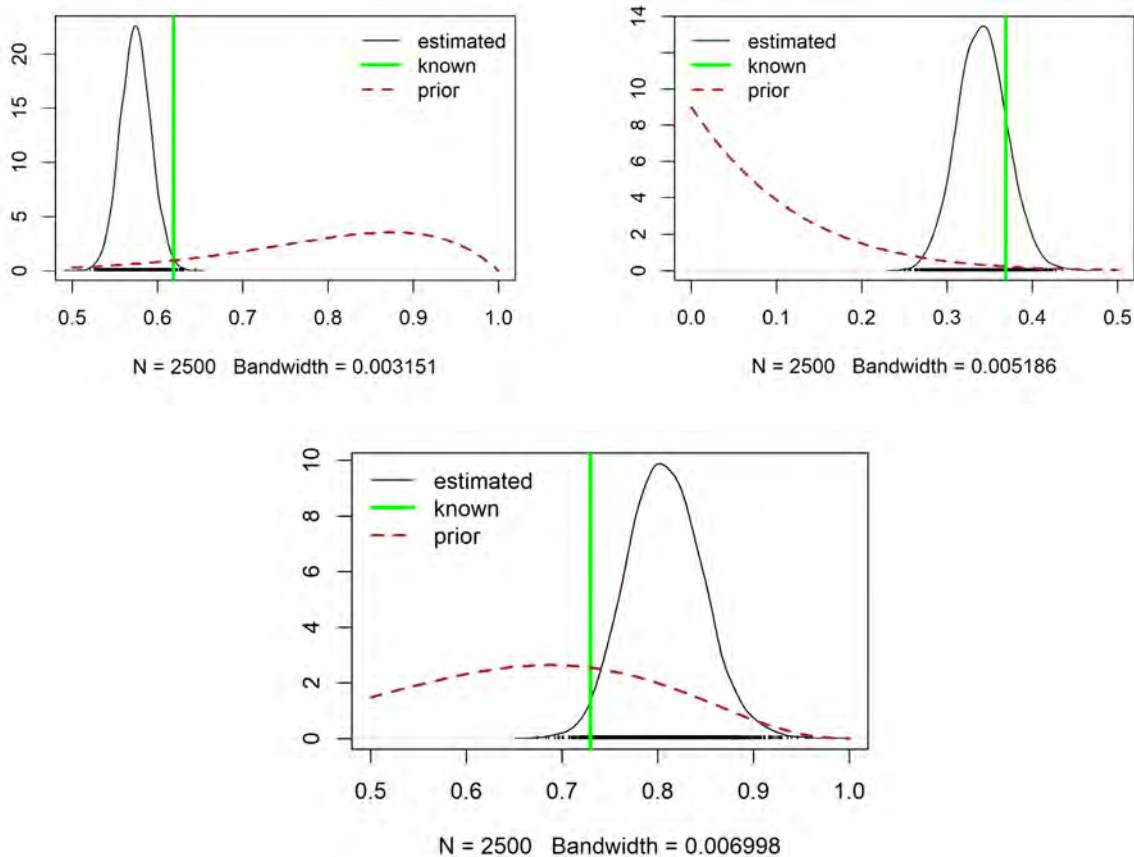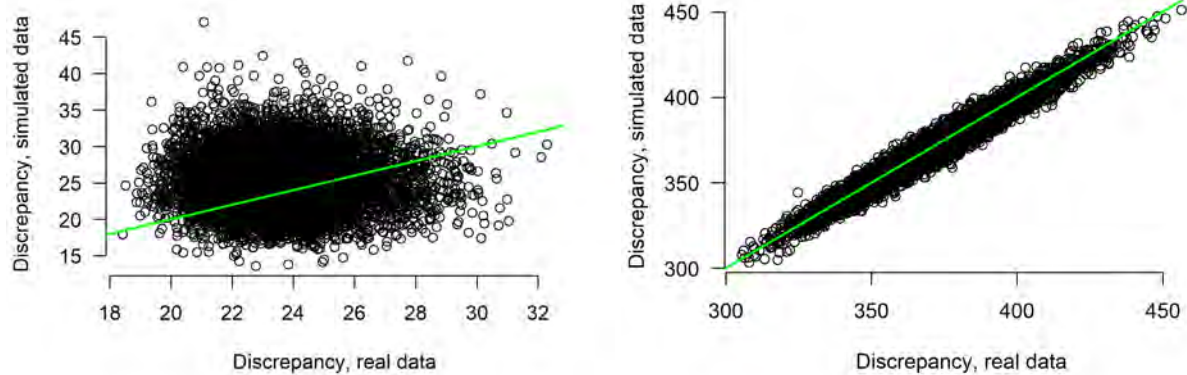
Figure S8. In simulation tests, the ORD model tended to perform moderately well at recovering the true gammaII term (temporary emigration- prob of returning to the stopover after using it last year) from simulated data (top left), GammaOI term (temporary emigration- prob of returning to the stopover after skipping last year) (top right) and Tau (stopover residency probability) (bottom panel). Green dots and vertical lines represent the true values used in simulations, black curves, points, and confidence intervals represent parameter estimates from the ORD model, and dashed brown curves represent the prior probability distributions used for Bayesian model fitting.

Fig. S9. *In simulation tests, the ORD model tended to perform moderately well at recovering the true temporal mean detection probability (top left). However, the ORD model performed somewhat poorly at recovering the temporal process variation in p (variation across both primary and secondary occasions); this parameter doesn't seem to fit well, and the chains exhibited very slow mixing. The bottom panel indicates detection probability per 3-day sampling occasion; the model appears to be underestimating variation among secondary occasions. Green dots and vertical lines represent the true values used in simulations, black curves, points, and confidence intervals represent parameter estimates from the ORD model, and dashed brown curves represent the prior probability distributions used for Bayesian model fitting.*

Fig. S10. *In simulation tests, the ORD model tended to perform moderately well at recovering the 'Delta' parameter (entrance probabilities) (top panel)— here, estimated from simulated data for year 8 (selected randomly from among years). The ORD model also performed well in recovering information about the 'Psi' parameter (probability of stopover persistence) (bottom panel). The green dots and vertical lines represent the true values used in simulations, while black curves, points, and confidence intervals represent parameter estimates from the ORD model.*

Fig. S11. *In simulation tests, the open-robust-design (ORD) model exhibited adequate goodness of fit, demonstrated here through posterior predictive checks (PPCs) involving the among-year survival process (L1; left panel) and the timing of first entry to the stopover each year (L2; right panel). The ORD model passed all four GOF tests when data were simulated using the same model assumptions used for model fitting (two tests not shown). In contrast, when the real REKN mark-resight data were used for model fitting, these tests indicated poor model fit (see section 6, above).*

ATTACHMENT

The following materials were submitted in September 2022 to inform the ASMFC Horseshoe Crab Management Board's consideration of Addendum VIII to the Horseshoe Crab Fishery Management Plan. They are included here for reference.

EARTHJUSTICE

September 30, 2022

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: Draft Addendum VIII to the Horseshoe Crab Fishery Management Plan for
        Public Comment

Dear Members of the Horseshoe Crab Management Board:

I write on behalf of New Jersey Audubon and Defenders of Wildlife to urge you to reject
Addendum VIII to the Horseshoe Crab Fishery Management Plan. Since the Board instituted the
Adaptive Resource Management ("ARM") Framework in 2012, red knot[1] abundance at
Delaware Bay has fallen to historically low levels, and the U.S. Fish & Wildlife Service
("FWS") has listed the species as "threatened" under the Endangered Species Act ("ESA").
Horseshoe crabs, too, remain severely depleted compared to historical benchmarks. These
circumstances demand greater protections and a precautionary strategy. But Addendum VIII
would instead weaken the protections currently in place. Among other harmful outcomes, the
Addendum almost certainly would reinitiate the female horseshoe crab bait harvest. Recognizing
that neither red knots nor horseshoe crabs have recovered, the ARM Framework, until this
proposal, has prohibited female harvest to protect the eggs on which the red knots rely.

Horseshoe crab eggs are critical to the red knot's ability to survive its 9,000-mile migration from
as far south as Tierra del Fuego and to breed successfully in the Arctic Circle. The importance of
horseshoe crab eggs to red knot success has long been recognized by scientists, government
agencies, and the Atlantic States Marine Fisheries Commission ("ASMFC" or "Commission"),
and the overharvest of horseshoe crabs has been a primary cause of the red knots' decline over
the past three decades.

Nevertheless, despite the well-established link between horseshoe crab eggs and red knot
survival and reproduction, Draft Addendum VIII proposes a starkly different version of reality.
Through a combination of modeling defects and risk-prone decision-making, the revised ARM
Framework now determines that the relationship between these species is scarcely perceptible,
and that red knots would be virtually indifferent to the renewed harvest of female horseshoe
crabs.

---

[1] In this document, "red knot" refers to the *rufa* subspecies.

As detailed in these comments and the attached expert reports by Dr. Kevin Shoemaker and Dr. Romuald Lipcius, this depiction of the relationship between horseshoe crab eggs and red knot demography is deeply flawed. Contrary to the conclusions represented in Draft Addendum VIII, adopting a new management approach that would enable resumption of the harvest of female horseshoe crabs at this juncture, when both red knots and horseshoe crabs are depleted, would harm red knots and present risks to the horseshoe crab population itself. Accordingly, the revised ARM Framework is not suitable for recommending horseshoe crab bait harvest quotas.

More specifically, the Board should reject Addendum VIII for reasons including but not limited to:

- **The revised ARM Framework errs in concluding that red knots are not highly dependent on horseshoe crabs at Delaware Bay.**
  - After flying thousands of miles, red knots arrive at Delaware Bay to renourish on horseshoe crab eggs. Under ideal conditions, red knots can double their body weight in less than two weeks. In the late 20th century, the peak count of red knots at Delaware Bay usually exceeded 40,000 and sometimes exceeded 90,000.
  - Horseshoe crabs were overharvested in the 1990s. In 2015, FWS listed red knots as "threatened" under the ESA and called horseshoe crab overharvest and corresponding egg depletion a "primary causal factor" in red knot decline. The peak red knot count has stayed below 13,000 for each of the past two years.
  - Despite this strong evidence of the importance of horseshoe crab eggs to red knots, the revised ARM Framework posits a weak link between the two species. By so doing, the revised ARM Framework subverts the premise of ASMFC's management regime for the horseshoe crab fishery, which is to manage the horseshoe crab harvest for red knot recovery.

- **New analysis reveals significant technical flaws that make the revised ARM Framework unsuitable for managing the horseshoe crab harvest.**
  - The revised ARM Framework abandons the well-established understanding of the importance of horseshoe crab eggs to red knots in favor of an extreme, contrary reconstruction of the ecosystem that defies history and reality. Even if horseshoe crabs vanished entirely today, the revised ARM Framework's computer model predicts that red knot abundance would remain stable on average or even increase over the next 50 years. The model clearly would not have predicted the decline of red knots that resulted from horseshoe crab overharvest in the 1990s, which discredits its usefulness in making projections that could help both species recover.
  - The revised ARM Framework also undermines sustainable management of horseshoe crabs. By miscalculating uncertainty, the horseshoe crab projection model generates artificially stable horseshoe crab population projections, when there actually exists a significant threat of decline.
  - The horseshoe crab population projections are significantly influenced by nonsensically high recruitment rates that were plugged in for years when recruitment was not measured empirically, thus further undermining the reliability of its projections.

ii

- o The horseshoe crab population model bears very little correlation even to the data that the model is based upon, raising significant additional doubt about its predictive power and usefulness.

- **The revised ARM Framework's risk-prone assumptions and decisions are inappropriate, especially when a threatened species is at stake.**
  - o Horseshoe crab demographic information, including size and sex ratio, strongly suggests that the species is not recovering and that a risk-averse management approach is required.
  - o The Framework does not consider the availability of horseshoe crab eggs, which is the most direct measure of food resources for red knots. Analysis of horseshoe crab demographic trends indicates that egg production may be declining more than abundance estimates suggest.
  - o The model finds a weak relationship between horseshoe crabs and red knots partly because it is based on data from years when both species had already declined rather than when the ecosystem was flourishing. Modeled projections of a depleted ecosystem offer no guidance on managing to achieve recovery of either red knots or horseshoe crabs.
  - o The Framework does not assess whether Delaware Bay provides adequate food for Southern wintering red knots, which are especially dependent on horseshoe crab eggs.
  - o The Framework would eliminate protective population thresholds that must be met prior to any female harvest, creating risks to red knots and horseshoe crabs and contravening stakeholders' precautionary intent.
  - o For population estimates, the model equally weights three surveys, despite stakeholders' express preference—and ASMFC's practice until now—to rely exclusively upon the model that is purpose-designed for counting horseshoe crabs. This results in artificially inflated horseshoe crab population estimates.

- **ASMFC has repeatedly excluded input from stakeholders and the broader public.**
  - o In addition to its other flaws, the revised ARM Framework is based on a model that has never been released to the public. Analysis of even the limited information made available to the public to date indicates significant problems with the model, as discussed above. If the Board approves Addendum VIII now and the model is subject to public evaluation, new concerns and critiques will inevitably arise after the revised ARM Framework is already in use.
  - o The ARM Subcommittee failed to solicit formal stakeholder input in this proceeding, in violation of its own procedures and past practice.
  - o By designating Addendum VI the "No Action" alternative, the Board artificially narrowed its options to two addenda that would reinitiate the female horseshoe crab harvest, thus deciding the most important issue before the public comment period even began.

- **The flaws in the revised ARM Framework must be addressed now.**
  - o The authority of ASMFC to deviate from the ARM Framework's harvest quotas in the future is not a rationale for approving Addendum VIII based on a flawed modeling framework now. Prematurely approving Addendum VIII would set the stage for contentious and arbitrary decisions about annual quotas for years to come.
  - o The authority of states to set lower quotas than ASMFC provides does not lessen the Board's obligation to ensure that the revised ARM Framework is fully vetted and reflects stakeholder values.
  - o Updating the revised ARM Framework's model as new data become available will not correct its fundamental flaws, many of which—as explained in these comments—are apparent from expert reviews of even the limited data made publicly available to date.

- **Approving Addendum VIII would likely lead to a violation of the Endangered Species Act by ASMFC.**
  - o The ESA requires a precautionary approach to protecting threatened species.
  - o By reinitiating the bait harvest of female horseshoe crabs, ASMFC would commit "take" of red knots. ASMFC is responsible under the ESA for harvests conducted pursuant to the quotas it sets.
  - o FWS's purported "evaluation" of the revised ARM Framework merely repackages ASMFC's modeling, with all of its flaws, and uses it to generate an unreliable conclusion regarding the impact of red knots. It therefore sheds no new light on the Board's stewardship responsibilities or the Commission's legal obligations.

The objections listed above are elaborated in the comments and expert reports that follow. Each objection is an independently sufficient reason to reject Addendum VIII. Collectively, they demonstrate that Addendum VIII is incompatible with the Board's mandate to maintain the ecosystem integrity of Delaware Bay and to comply with the Endangered Species Act.

Respectfully submitted,

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

Table of Contents

I. THE REVISED ARM FRAMEWORK ERRS IN CONCLUDING THAT RED KNOTS ARE NOT HIGHLY DEPENDENT ON HORSESHOE CRABS AT DELAWARE BAY.

Each year, a population of red knots completes one of the most epic migrations in the animal kingdom. Starting from Tierra del Fuego at the southern tip of South America, the red knots fly more than 9,000 miles to their breeding grounds in the Arctic Circle. For most red knots, the final staging area before the Arctic Circle is Delaware Bay, where their stopover coincides with another ecological marvel: the spawning of millions of horseshoe crabs that emerge from the water and lay clusters of approximately 4,000 eggs, with the potential for an individual to lay more than 100,000 eggs over the course of several nights.[2] For red knots that have already flown thousands of miles at enormous physiological expense, the eggs provide essential replenishment, enabling a doubling of body mass in fewer than 14 days, versus 21 to 28 days at comparable stopovers where they eat clams and mussels.[3] This unique resource fuels the duration of their journey and enhances breeding success in the Arctic.[4]

The abundance of red knots and horseshoe crabs at Delaware Bay as recently as the 1990s is almost unimaginable today. From 1981 to 2002, the peak red knot count in Delaware Bay usually exceeded 40,000 and twice surpassed 90,000.[5] One participant in an aerial survey of shorebirds during that period described "lines of deposited horseshoe crab eggs set like mineral veins in smooth white marble, virtually an unlimited food supply."[6] In a single day, his survey tallied 62,000 red knots and 318,000 total shorebirds on just the New Jersey side of Delaware Bay.[7]

In the 1990s, increasing and unregulated horseshoe crab harvest by the bait and biomedical industries crashed the population of horseshoe crabs.[8] Red knots, no longer able to rely on the irreplaceable horseshoe crab eggs, declined in tandem. ASMFC adopted a fishery management plan for horseshoe crabs in 1998 and instituted adaptive management in 2012. Since then, the female bait harvest has been prohibited. But the fate of horseshoe crabs remains highly uncertain, and red knots have continued to decline. Red knot peak counts that previously topped 90,000 have, for the past two years, languished below 13,000, including a record low of 6,800 in 2021. Twenty years have passed since the population topped a modest 33,000.[9] Instead of these peak

---

[2] NOAA Fisheries. *Horseshoe Crabs: Managing a Resource for Birds, Bait, and Blood* (July 31, 2018), https://www.fisheries.noaa.gov/feature-story/horseshoe-crabs-managing-resource-birds-bait-and-blood.

[3] Lawrence Niles et al., *Effects of Horseshoe Crab Harvest in Delaware Bay on Red Knots: Are Harvest Restrictions Working?*, 59 BioScience 153, 154 (2009); New Jersey Department of Environmental Protection, *Wildlife Populations: Red Knot* 1-2 (2020), https://www.nj.gov/dep/dsr/trends/wildlife-redknot.pdf.

[4] Sjoerd Duijns et al., *Body Condition Explains Migratory Performance of a Long-Distance Migrant*, 284 Proceedings of the Royal Society of London B 20171374, at 4-6 (2017).

[5] FWS, *Rufa Red Knot Background Information and Threats Assessment* 100 tbl. 12 (2014) (excluding 1984-1985, when the survey was not conducted).

[6] Pete Dunne, *Tales of a Low-Rent Birder* 10 (1986).

[7] *Id.* at 13-14.

[8] FWS, *Rufa Red Knot Background Information and Threats Assessment* 232 ("Evidence that commercial harvests caused horseshoe crab population declines in recent decades comes primarily from a strong temporal correlation between harvest levels . . . and population levels.").

[9] *Id.* at 100 tbl. 12 (for years 1981-2014); 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)* 155 tbl.

counts, the revised ARM Framework uses modeled estimates of the total number of red knots passing through Delaware Bay. While these modeled estimates face criticism for overrepresenting red knots' use of Delaware Bay, they have fallen as well, from as high as 152,900 in 1989, to an average of 77,000 per year for 1998-2001, to numbers in the 40,000s over the past several years.[10]

In 2015, FWS formally listed the red knot as a threatened species under the Endangered Species Act.[11] At the time of the listing, FWS cited several studies indicating that red knot abundance had declined, "probably sharply," since the 1980s.[12] FWS found that "[r]educed food availability in Delaware Bay due to commercial harvest of the horseshoe crab . . . is considered a primary causal factor in red knot population declines in the 2000s."[13] Reduced food availability is a particular threat for the Southern wintering population of red knots, which is disproportionately reliant on the Delaware Bay staging area and which FWS views as "a bellwether for the subspecies as a whole."[14] According to FWS, "[R]educed food availability at just one key migration stopover area (Delaware Bay) is considered the driving factor behind the sharp decline in the Southern wintering population in the 2000s."[15]

As FWS has stated, "Studies have shown red knot survival rates are influenced by the condition (weight) of birds leaving the Delaware Bay staging area in spring."[16] Research has also shown that, while red knots arriving relatively late to Delaware Bay were able to compensate by gaining weight at a higher rate, that was not the case in years with low horseshoe crab egg availability.[17]

Until now, the well-established link between horseshoe crabs and red knots has been the cornerstone of ASMFC's management of the horseshoe crab fishery at Delaware Bay. Addendum VIII would subvert that regime. While the proposed model nominally bases harvest quotas on red knot and horseshoe crab abundance estimates, it assigns an extremely weak correlation between the abundance of the two species. It thereby concludes that red knots would be essentially unaffected by the resumption of the female horseshoe crab bait harvest.

As explained below, Addendum VIII's baseline assumption—that increasing the horseshoe crab harvest would only marginally impact red knots at Delaware Bay—is unsupported. It relies on evaluating a limited dataset that omits years when the ecosystem flourished. (For example, its dataset about horseshoe crab abundance is drawn entirely from the last 20 years, after the crash

---

12 (2021) ("ARM Report") (for years 2011-2020); Larry Niles, "2022 Delaware Bay Stopover Project Final Update-5 June 2, 2022," *A Rube with a View* (June 15, 2022), https://www.arubewithaview.com/2022/06/15/2022-delaware-bay-stopover-project-final-update-5-june-22022/ (for years 2021-2022).

[10] FWS, *Rufa Red Knot Background Information and Threats Assessment* 101 tbl. 13; ASMFC, *ARM Report* 155 tbl. 12.

[11] FWS, "Endangered and Threatened Wildlife and Plants; Final Threatened Status for the Rufa Red Knot," 79 Fed. Reg. 73,706 (Dec. 11, 2014). The listing became effective on January 12, 2015. *Id.* at 73,706.

[12] FWS, *Rufa Red Knot Background Information and Threats Assessment* 85. While FWS primarily analyzed red knot population trends within individual regions, it "note[d] a temporal correlation between declines at Tierra del Fuego and Delaware Bay." *Id.* at 84.

[13] 79 Fed. Reg. at 73,707.

[14] FWS, *Draft Recovery Plan for the Rufa Red Knot* 13 (May 2021).

[15] *Id.* at 14.

[16] *Id.* at 25; FWS, *Rufa Red Knot Background Information and Threats Assessment* 254.

[17] FWS, *Rufa Red Knot Background Information and Threats Assessment* 253.

of the horseshoe crab population and during a period when red knot abundance has been comparatively low.) And it suffers from modeling defects that, among other things, erroneously overstate the size and stability of the horseshoe crab population.

For these reasons and others detailed below, Addendum VIII is not a pathway for sustaining red knots, much less restoring a thriving ecosystem, nor does it honor the precautionary approach required when a threatened species is at stake. Instead, it risks a violation of ASMFC's legal obligations, including its obligation to avoid "take" of red knots under the ESA. The Board therefore should reject Addendum VIII and instead adopt adequate protections for horseshoe crabs and red knots at Delaware Bay.

II. NEW ANALYSIS REVEALS SIGNIFICANT TECHNICAL FLAWS THAT MAKE THE REVISED ARM FRAMEWORK UNSUITABLE FOR MANAGING THE HORSESHOE CRAB HARVEST.

As detailed in the following sections, the parties to this letter solicited independent expert reviews of the revised ARM Framework. These reviews reveal significant technical and methodological flaws that render the Framework unreliable for ASMFC management decisions.

For the first expert review, Dr. Kevin Shoemaker conducted an independent analysis of the horseshoe crab abundance and projection model that informs the revised ARM Framework. Dr. Shoemaker demonstrates that the Framework contains significant flaws that make it unsuitable for managing the horseshoe crab harvest. These flaws are especially alarming given the implications of the Framework for a threatened species such as the red knot. This section details many of Dr. Shoemaker's key findings, all of which are explained in more detail in the attached expert report.

At the outset, it is important to note that most of the components of the revised ARM Framework's model still have not been made available to the public. As a result, Dr. Shoemaker was unable to evaluate the components that link horseshoe crab abundance to red knot abundance or generate horseshoe crab harvest recommendations. Although Dr. Shoemaker was able to draw some conclusions about those aspects of the model, most of the analysis below necessarily focuses on the horseshoe crab model. As these comments proceed to discuss, the analysis that Dr. Shoemaker was able to conduct reveals severe issues concerning the reliability of the modeling. Nevertheless, Dr. Shoemaker's focus on the publicly available modeling information should not be interpreted to suggest that the unreleased components do not also contain significant flaws. To the contrary, given the flaws that are apparent in the information released to date, it is vital that *all* components of the model be subject to public evaluation before the Board takes any action to approve Addendum VIII.

A. The revised ARM Framework Is an Inappropriate Tool for Helping to Reverse the Decline and Promote the Recovery of Red Knots.

Considering that adaptive management is premised on the link between horseshoe crabs and red knots, the weakness of that link in the revised ARM Framework is breathtaking. By way of illustration:

- Dr. Shoemaker shows that, even if the horseshoe crab population in Delaware Bay completely collapsed to zero, the revised ARM Framework would predict that red knot abundance would remain stable or even increase over the next 50 years on average.[18]
  - Furthermore, "This simulation exercise makes it very clear that the REKN model used in the revised ARM would not be able to predict or explain the decline in the REKN population observed during the 1990s."[19] In other words, the model could not even have diagnosed the problem that it is supposed to solve.
- The data informing the revised ARM Framework actually show a negative correlation between female horseshoe crab abundance and red knot recruitment.[20] That is, according to the model, as female horseshoe crab abundance *increases*, red knot recruitment *decreases* on average.
- Due to the weak relationship between red knot and horseshoe crab abundance, it is not implausible that, with future updates to the revised ARM Framework, the relationship will disappear entirely or even become negative. Dr. Shoemaker observes that "[t]his outcome would pose an existential problem for the ARM framework . . . . There does not appear to be a contingency plan for this outcome."[21]
- Whatever weak signal the model has detected in historical data appears to be overwhelmed by random noise. As Dr. Shoemaker explains, it is highly likely that the model's "information about the HSC/REKN relationship would explain little if any of the variation in independent validation data."[22]

Due to the weak relationship between red knots and horseshoe crabs represented in the revised ARM Framework, it is unlikely that the model would outperform—much less significantly improve upon—a "null" model that entirely omits any effect of horseshoe crab abundance.[23] Yet it was impossible for Dr. Shoemaker to explore this key issue further because of the limitations on the materials made publicly available to date. Nevertheless, the concerns raised by the analysis that Dr. Shoemaker was able to perform are profound and call into question the revised ARM Framework's utility to guide any decision-making about the status or management of the affected species.

In sum, while the revised ARM Framework nominally recommends harvest quotas based on the relationship between horseshoe crabs and red knots, it effectively decouples the fates of the two species, unjustifiably transforming the methodology and philosophy that underlie the management of this fishery. This is an independently sufficient reason for the Board to reject Addendum VIII.

---

[18] Kevin Shoemaker, *Review of 2021 ASMFC ARM Revision* 6-9 & fig. 1 (Sept. 2022) ("*Shoemaker Expert Report*").
[19] *Id.* at 8.
[20] *Id.* at 9 fig. 2.
[21] *Id.* at 10.
[22] *Id.* at 26.
[23] *Id.* at 25-26.

B. The Horseshoe Crab Population Simulation Model Does Not Properly Account for Uncertainty, Resulting in Artificially Stable Abundance Projections.

The revised ARM Framework profoundly underestimates uncertainty in the horseshoe crab recruitment rate, thereby calling into question its projections concerning the impact of harvest. As Dr. Shoemaker explains, the rate at which new recruits join the reproductive population "is the most consequential empirically fitted component of the HSC simulation model."[24] Other components of the model, such as natural and biomedical mortality, are fixed values, but the recruitment rate is calculated based on data.

Dr. Shoemaker shows[25] that the model errs by conflating two distinct types of uncertainty: (i) natural, year-over-year variation and (ii) the potential that the model incorporates incorrect parameters (most importantly, the mean horseshoe crab recruitment rate). The model treats both types of uncertainty as natural, year-over-year variation, with the consequence that the abundance estimates regress to a mean. In other words, the variations cancel each other out, making the projected population appear highly stable. But if evaluated properly, parameter uncertainty would likely compound over time, yielding a very different picture of the population. For example, if average recruitment is actually lower than the rate used in the model, that uncertainty would *not* cancel out over time. Instead, the horseshoe crab population could be headed for a one-way decline. Notably, the revised ARM Framework accounts for the two types of uncertainty separately in the *red knot* projection model, suggesting that the modelers recognized the importance of that approach, but nevertheless they did not implement it when projecting horseshoe crab abundance.

The consequences of this error are significant for estimates of the population's trajectory. Properly accounting for uncertainty, Dr. Shoemaker found that the horseshoe crab population faces a very real threat of declining well below levels acknowledged by the revised ARM Framework's projection model. Notably, he used the same estimates of uncertainty as the revised ARM Framework (as well as the same values for natural mortality, biomedical mortality, etc.). All that changed in his analysis was the method of evaluating uncertainty. Dr. Shoemaker's analysis[26] reveals that:

- Even under a scenario with *no* bait harvest, *no* biomedical mortality, and *no* discard mortality, the female horseshoe crab population has a 17.4% probability of declining below 4 million, and a 3.8% probability of declining below 3 million, over the next 50 years.
    - o For comparison, 4 million is the lowest female abundance estimated for any year from 2003 to 2019 (the years upon which the model was based).
    - o In contrast, by incorrectly accounting for uncertainty, the revised ARM Framework's model does not project female abundance values below 4 million within the 95% confidence interval under optimal harvest scenarios, *including* bait harvest, biomedical mortality, and discard mortality.[27]

---

[24] *Id.* at 12.

[25] The information in this paragraph is drawn from *Shoemaker Expert Report* 12-18 & figs. 3-4.

[26] Except where noted, these findings are presented in greater detail at *Shoemaker Expert Report* 15, 18 fig. 4.

[27] ASMFC, *Supplemental ARM Report* 35 fig. 15.

- Under a scenario in which horseshoe crabs are harvested for bait under the maximum quotas of 500,000 males and 210,000 females but are still *not* subject to biomedical or discard mortality, the female population has a 33% probability of declining below 4 million, an 11% probability of declining below 3 million, and a 2% probability of declining below 2 million, over the next 50 years.

Dr. Shoemaker concludes that, "if sources of error in the recruitment process are properly accounted for, the outlook for the HSC population in Delaware Bay is uncertain even in the absence of any harvest pressures."[28] If the Board approves Addendum VIII, it would increase harvest pressure through a model that fails to properly account for the risk of a declining horseshoe crab population.

      C.   The Horseshoe Crab Projection Model's Recruitment Estimates Are Strongly Influenced by Nonsensical, Unverified Estimates from the Virginia Tech Gap Years.

The revised ARM Framework's conclusions are further undermined by its reliance on fantastical recruitment projections to fill in a key gap in actual population-monitoring data for horseshoe crabs. Of the three trawl surveys that inform the catch multiple survey analysis ("CMSA") component of the framework, only the Virginia Tech survey measures primiparous (i.e., newly mature) females to provide an empirically based estimate of recruitment. Thus, the CMSA does not incorporate any direct measurement of recruitment during the 2013-2016 period when the Virginia Tech survey was not conducted. Instead, it indirectly estimates annual recruitment rates, but two of these estimates are many times higher than any estimate from years with direct observations. Since the average recruitment rate in the population projection model treats all of the estimates as equally valid—whether or not they were based on empirical observations or hypothetical estimates—the model's estimated annual recruitment rate is heavily influenced by the nonsensical estimates from the Virginia Tech gap years.

To understand the impact of the nonsensical gap year estimates, first consider the years with empirically derived recruitment estimates. The average annual estimated recruitment for 2003-2012 was 1.2 million primiparous females. The average annual estimated recruitment for 2017-2019 was 1.9 million. Now consider the non-empirically derived gap year estimates. In 2013, the estimate was *9.6 million*—roughly eight times larger than the average over the previous ten years, and four times larger than the maximum annual estimate from that period.[29] In 2014, the estimate dropped to only two primiparous females across all of Delaware Bay, but the estimate is so uncertain that the upper limit of the confidence interval approaches infinity.[30] All told, the *average* estimate for the four Virginia Tech gap years was 4.2 million primiparous females, which is nearly *2 million* higher than the *maximum* ever estimated for any year with empirical observations.[31]

---

[28] *Shoemaker Expert Report* 17.

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

[30] *Id.* at 25 fig. 5.

[31] *Id.* at 16 tbl. 3.

The nonsensical estimates from the Virginia Tech gap years compromise the horseshoe crab projection model because they significantly affect its recruitment estimate. As Dr. Shoemaker shows,[32] in the original ARM report, the ARM Subcommittee based the recruitment rate exclusively on data from 2013 to 2019, which relied overwhelmingly on estimates from the gap years and generated an annual recruitment estimate of 3.1 million primiparous females. Following criticism from the Peer Review Panel, the Subcommittee expanded the dataset to include 2003-2019, which reduced the recruitment estimate to 1.67 million. But if the nonsensical data from the gap years were excluded, this estimate would fall to 1.26 million. Dr. Shoemaker illustrates how the difference in these estimates has huge implications for the model's projection of future horseshoe crab abundance.

Dr. Shoemaker concludes that "the inflated estimates of recruitment during the VT gap years are likely to be an artifact of the CMSA model specification (and the lack of data on recruitment for those years) and are unlikely to be reflective of true HSC recruitment rates. . . . [A] conservative (precautionary) approach would be to exclude the VT gap years when computing recruitment for the HSC population simulations."[33] Doing so would yield a substantially lower recruitment estimate with a commensurately lower capacity to withstand a resumption of female harvest.

D. The Horseshoe Crab Population Model Has a Poor Correlation to Existing Data.

The CMSA's usefulness is cast further into doubt by its failure to correlate with any source of data about horseshoe crab abundance. As Dr. Shoemaker shows from an analysis of female horseshoe crab abundance estimates, the model does not correlate even with the data sources upon which it was based, much less any independent validation data.

For the years 2003-2019, the CMSA's correlation with the Delaware Adult Trawl Survey is extremely weak, and any correlation that exists is entirely attributable to the model's apparent ability to predict that horseshoe crab populations rose during 2013-2016, when the Virginia Tech survey was not conducted.[34] For the years before and after the Virginia Tech gap—that is, for the vast majority of years evaluated—the coefficient of determination (R^2) between the CMSA model and the Delaware Survey was *negative*, meaning that the model performed worse than a null model. The CMSA performs almost as poorly against data from the New Jersey Ocean Trawl Survey, with a weak positive correlation for the years prior to the Virginia Tech gap and a negative R^2 for the years after. The CMSA's worst performance comes when measured against the Virginia Tech survey, with a negative R^2 across the full time series for which data are available. To test the CMSA against independent validation data, Dr. Shoemaker compared it to the results of Delaware Bay spawning surveys and found no detectable relationship whatsoever between the results.

As this summary makes clear, the CMSA's modeled outcomes bear little relationship to actual data on the Delaware Bay horseshoe crab population. For this reason, Dr. Shoemaker recommends comparing the CMSA's horseshoe crab estimates to a null model that omits all information about horseshoe crab harvest from the model fitting process. Given its poor fit to

---

[32] The data discussed in this paragraph can be found at *Shoemaker Expert Report* 22-24 & fig. 7.
[33] *Id.* at 23.
[34] The findings in this paragraph are presented in greater detail at *Shoemaker Expert Report* 19-22 & figs. 5-6.

existing data, the CMSA's horseshoe crab projection model is "unlikely to outperform" even a relatively simple null model.[35] Dr. Shoemaker concludes, "If the HSC simulation model fails to outperform a model in which population dynamics are driven by noise instead of harvest, it should prompt managers to acknowledge that our current understanding of the effects of harvest on HSC populations remains insufficient for robust forecasting."[36] Absent a sound basis for robust forecasting, adoption of Addendum VIII and its attendant resumption of the female harvest cannot be justified.

III. THE REVISED ARM FRAMEWORK'S RISK-PRONE ASSUMPTIONS AND DECISIONS ARE INAPPROPRIATE, ESPECIALLY WHEN A THREATENED SPECIES IS AT STAKE.

In addition to its technical flaws, the revised ARM Framework incorporates risk-prone assumptions and decisions that further render it unsuitable as a management tool. It neglects important variables related to horseshoe crab demography and egg density that cast doubt upon the recovery of horseshoe crabs and their ability to provide adequate food resources for red knots. It draws conclusions from data collected when both red knots and horseshoe crabs were already depleted and therefore does not understand how the species would interact in a healthy ecosystem. It also reverses precautionary decisions made by stakeholders in the original ARM Framework—without soliciting renewed stakeholder input—in order to eliminate protections against the female horseshoe crab harvest and utilize previously-rejected surveys that inflate horseshoe crab abundance estimates.

The findings in this section draw heavily from an independent analysis of the revised ARM Framework and related materials conducted by Dr. Romuald Lipcius, as well as the analysis of Dr. Shoemaker. Both expert reports are attached.

A. Demographic Trends Indicate that the Horseshoe Crab Population Is Not Recovering.

Despite the Subcommittee's assertion that horseshoe crab abundance is increasing in Delaware Bay, Dr. Lipcius has identified troubling indicators that are inconsistent with a recovering population. The revised ARM Framework ignores these trends and treats abundance estimates as a comprehensive indication of population health. That would be a risk-prone approach even if the abundance estimates were fully reliable (which they are not).

As shown in Dr. Lipcius's report, the mean size (prosomal width) of female horseshoe crabs has recently declined. In the most recent three years of available data (2018-2020), adult female horseshoe crabs recorded the lowest mean sizes of any year since data collection began in 2002.[37] The same is true for newly mature females over the most recent two years of available data.[38]

---

[35] *Id.* at 25.
[36] *Id.*
[37] Romuald Lipcius, *Expert Report* 6 (Sept. 2022) ("*Lipcius Expert Report*").
[38] *Id.*

Dr. Lipcius explains that, given constant recruitment, a prohibition on female harvest would typically lead to an increase in size due to reduced harvest pressure on older, larger females.[39] The declining size of female horseshoe crabs is inconsistent with the premise that the female segment of the population has recovered.[40] It is further evidence that the revised ARM Framework does not properly account for the population dynamics of horseshoe crabs.

A female harvest prohibition would also be expected to decrease the ratio of males to females in the population. But the data indicate that the male-to-female ratio increased between 1999 and 2019, suggesting *fewer* females for every male.[41] This is another warning sign that the population has not recovered, and the harvest of female horseshoe crabs should not resume.[42] Resuming such harvest would only further deplete a critical component of the population that has failed to show expected signs of recovery even under the female harvest prohibition.

Abundance data for immature and newly mature females raise additional concerns about the recovery of the female population. In 2019 and 2020, the Virginia Tech survey estimated the lowest abundance of newly mature female horseshoe crabs since data collection began in 2002, "indicating low influx of young mature females into the spawning stock."[43] Moreover, abundances of immature females and males for 2016-2020 were similar to those before 2013, when there was no female harvest prohibition in place. That is again contrary to expectations, since a prohibition on harvesting females should correlate to an increase in younger individuals.[44]

Dr. Lipcius explains that estimates of abundance can be less sensitive to serious problems in a population than variables including female size, female size structure, spawning stock biomass, and sex ratio. But the revised ARM Framework relies on abundance estimates to the exclusion of these other important variables. That is a risk-prone strategy and is not suitable for protecting horseshoe crabs or the threatened red knots.

> B. The Revised ARM Framework Fails to Consider Horseshoe Crab Egg Density, the Most Direct Measure of Food Availability for Red Knots.

Another critical omission in the revised ARM Framework is its exclusion of data about the most direct measure of the adequacy of food resources for red knots: the availability of horseshoe crab eggs on the beach. As explained above, for red knots arriving at Delaware Bay after flying thousands of miles, horseshoe crab eggs provide energy-rich, easily digestible nutrition as the birds prepare to complete their journey northward and breed in the Arctic Circle. Red knots flying from South America shrink their digestive organs for the journey, and no other food source can replace easily digestible horseshoe crab eggs in enabling red knots to quickly rebuild their organs and muscles.[45] When conditions permit, a red knot at Delaware Bay can double its

---

[39] *Id.*
[40] *Id.*
[41] *Id.* at 10.
[42] *Id.*
[43] *Id.* at 6, 7 fig. 1.
[44] *Id.*
[45] Niles et al., *Effects of Horseshoe Crab Harvest* 154.

body mass in as little as 12 days by feasting on horseshoe crab eggs.[46] Research indicates that the red knots that have flown the farthest, from Tierra del Fuego, are particularly dependent on the density of horseshoe crab eggs (i.e., the number of eggs per square meter of beach).[47] Nevertheless, the revised ARM Framework has failed to consider actual data on egg density in the Delaware Bay region. Whatever concerns may have existed about such data at the time the original ARM Framework was developed, egg density should now be considered in light of new scholarship (discussed below) and the importance of horseshoe crab eggs for red knots. The revised ARM Framework's failure to do so represents another key flaw.

       1.   Egg density is the most direct measure of food availability for red knots.

Scientific studies link food availability at Delaware Bay to red knot survival and fecundity. Under favorable conditions including abundant horseshoe crab eggs, red knots at Delaware Bay roughly double their body mass from 90-120 grams to 180-220 grams before departing for the Arctic.[48] Individual red knots can gain up to 15 grams per day, "probably when horseshoe crab eggs are superabundantly available," allowing even late-arriving red knots to gain adequate mass in a brief period.[49] Researchers have observed that red knots experience "striking fitness consequences . . . correlated with the amount of nutrient stores accumulated in Delaware Bay."[50] Specifically, research has found a positive correlation between the mass of birds leaving Delaware Bay in the spring and the speed at which they complete their migration to the Arctic, reproductive success, and survival to the autumn.[51]

A superabundance of horseshoe crab eggs is required to meet the nutrition needs of red knots, other shorebirds, and the many other species that rely on this unique resource. Horseshoe crabs lay eggs too deep in the sand for red knots to access. But as more horseshoe crabs spawn on the beach, they disturb the sand, churning some of the eggs closer to the surface.[52] It is this churning, as well as wave action, that makes horseshoe crab eggs accessible to red knots.[53] The system depends on the successive spawning of large numbers of horseshoe crabs.[54]

       2.   Egg Density Has Declined Dramatically in Recent Decades, Correlating with the Decline in Red Knots.

Research strongly demonstrates that the abundance of horseshoe crab eggs near the beach surface (where the eggs are accessible to red knots) used to be at least ten times greater than the

---

[46] New Jersey Department of Environmental Protection, *Wildlife Populations: Red Knot* 1-2.

[47] FWS, *Species Status Assessment Report for the Rufa Red Knot* (Version 1.1) 9 (Sept. 2020) ("*Species Status Assessment Report*").

[48] Allan J. Baker et al., *Rapid Population Decline in Red Knots: Fitness Consequences of Decreased Refuelling Rates and Late Arrival in Delaware Bay*, 271 Proceedings of the Royal Society of London B 875, 876 (2004).

[49] *Id.* at 876.

[50] *Id.* at 881.

[51] Duijns et al., *Body Condition Explains Migratory Performance* 5-6.

[52] Niles et al., *Effects of Horseshoe Crab Harvest* 155.

[53] *Id.*

[54] *Id.*

abundance in recent years.[55] Measurements from 1985 to 1987 conservatively indicate that egg density averaged 156,000 eggs per square meter of beach. In recent years, egg density averaged only around 10,000 eggs per square meter of beach.[56]

This decline in egg density correlates with the dramatic decline of migratory shorebirds, especially red knots. The trends mirror each other over decades but also converge on smaller timescales. Among years when measurements were taken, the nadir for horseshoe crab egg density appears to have been the early 2000s, shortly after the unregulated overexploitation of horseshoe crabs in the 1990s.[57] This corresponds to a "changepoint" for red knots when the peak count dropped from more than 43,000 to fewer than 16,000.[58]

> 3. Horseshoe Crab Abundance Is Not an Adequate Proxy for Egg Availability.

Notwithstanding the research documenting a dramatic decline in the availability of horseshoe crab eggs, the revised ARM Framework posits that the abundance of female horseshoe crabs is increasing. That is a dubious claim, as explained in section III.A of these comments. But even assuming for the sake of argument that it were correct, it would not necessarily result in more eggs for horseshoe crabs. To the contrary, demographic trends suggest that the production of eggs per horseshoe crab is likely decreasing.

Dr. Lipcius describes how egg production is directly proportional to the weight of horseshoe crabs, such that heavier crabs produce more eggs.[59] Data from the Virginia Tech Horseshoe Crab Trawl Survey indicate that the average prosomal width of female horseshoe crabs has fallen considerably, with an especially marked drop in the largest crabs over the past few years (2018-2020). Weight is an exponential function of prosomal width, meaning that even a modest decline in crab width could signify a very significant decline in weight and therefore in egg production. The trend toward smaller female horseshoe crabs may partially explain the low egg density numbers in recent years. Dr. Lipcius concludes that "total reproductive (egg) output has likely not improved, which hampers recovery of the HSC and RK populations."[60]

> 4. The ARM Report Presents No Compelling Reason to Ignore Egg Density.

There is no defensible rationale for completely excluding from the revised ARM Framework any direct measure of the most direct indicator of the adequacy of the red knot food supply: egg density. None of the ARM Subcommittee's reasons for excluding data about food availability withstands scrutiny.

---

[55] Joseph A.M. Smith et al., *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) (in press) ("*Horseshoe Crab Egg Availability*").
[56] *Id.*
[57] *Id.*
[58] *Id.*
[59] The information in this paragraph is drawn from *Lipcius Expert Report* 7-10 & figs. 2-6.
[60] *Id.* at 10.

First, the Subcommittee asserted that the protocol for measuring egg density over the years was too variable to provide reliable comparisons.[61] Even if that was previously a legitimate concern, scientists have now demonstrated a long-term reduction in the surface availability of horseshoe crab eggs based on multiple studies using similar methods and sampling from comparable or even identical locations.[62] More fundamentally, in the context of a threatened species, major warning signs should not be disregarded on the basis of uncertainty in the data, especially when the data that exist point strongly in the same troubling direction. As Dr. Lipcius explains, "Lack of use of HSC egg density data, as a proxy for RK food availability, amounts to a failure to incorporate all available scientific information into the analysis to guide management decisions in a risk-averse manner."[63]

The Subcommittee next asserted that habitat loss had not been "adequately rule[d] out" as the cause of declining egg density. This argument is equally misplaced. Recent research demonstrates that egg density has declined even where habitat continues to be suitable, such as where sand depth exceeds 40 centimeters.[64] Moreover, habitat loss does not provide a basis for disregarding the availability of horseshoe crab eggs for red knots. As Dr. Lipcius explains, while the Board does not have control over all sources of stress on horseshoe crabs, the existence of multiple stressors demands a *more* risk-averse approach with respect to factors such as harvest quotas that are fully within the Board's control.[65]

In addition, the Subcommittee denied the ability to link horseshoe crab egg abundance with red knot nutrition or survival.[66] However, as shown above, there is a strong correlation between declining egg density and declining red knot abundance.

Regardless of the Subcommittee's concerns that egg density data are not sufficiently conclusive, or that habitat loss is a contributing factor, multiple studies over several decades uniformly point in the same direction: egg density has declined to an alarming degree, as have the red knots that consume the eggs. At a minimum, the Commission must recognize that plentiful eggs are a necessary and critical element of red knot recovery and solicit formal stakeholder input on incorporating that principle into harvest decisions in light of recent research.

> ### C. The Revised ARM Framework Finds a Weak Relationship Largely Because It Relies on Data from Years When Both Red Knots and Horseshoe Crabs Were Already Depleted.

In contrast to all of the scientific information discussed above demonstrating a critical connection between horseshoe crabs and red knots, the revised ARM Framework finds a weak link between these species partly because it is based entirely on data from after the ecosystem

---

[61] ARM Subcommittee, *Majority Response to Niles and Justification for Why Opinion Not Adopted* (in *ASMFC, ARM Report*) 105-06.
[62] Smith et al., *Horseshoe Crab Egg Availability*.
[63] *Lipcius Expert Report* 12.
[64] Smith et al., *Horseshoe Crab Egg Availability*.
[65] *Lipcius Expert Report* 13.
[66] ARM Subcommittee, *Majority Response to Niles* 104.

crashed in the late 1990s.[67] The most the model can do is interpret the interaction between two perilously depleted species, without any concept of how a healthy ecosystem would function. In defiance of historical and scientific evidence, the revised ARM Framework seems to assume that a supposedly minimal correlation between horseshoe crabs and red knots when both species are degraded is indicative of how the ecosystem would operate when both species are plentiful. Rather than viewing its finding of a weak link appropriately as a symptom of an ailing ecosystem, the revised ARM Framework leverages it to justify greater exploitation.

As one example of why recent data may not represent the historic relationship between the two species, consider the population of red knots migrating from southern South America. These birds travel the farthest to reach Delaware Bay and need to rebuild their digestive organs upon arrival, making them particularly dependent upon easily digestible horseshoe crab eggs.[68] Even more than other red knots, this Southern wintering population has suffered "sharp and well-documented declines" in recent decades due to reduced food availability at Delaware Bay.[69] As a result, the relatively small number of red knots that pass through Delaware Bay may be increasingly skewed toward birds that winter farther north, with fewer of the birds that most heavily depend upon horseshoe crab eggs. The revised ARM Framework would interpret these conditions to mean that red knot abundance is less affected by horseshoe crab abundance and that greater exploitation is acceptable. It would thus ignore the impact of egg scarcity on the most vulnerable population of red knots.

While the revised ARM Framework may necessarily be limited by the years from which data are available, it should not draw overbroad conclusions from a constrained dataset. As Dr. Shoemaker explains, these constraints give the model a "limited scope of historical variation . . . . Using these models to forecast system dynamics under conditions outside the range of values used to fit the model (e.g., lower HSC abundances, higher REKN abundances) therefore requires extrapolation, which can be highly uncertain (and often inaccurate)."[70] Based on Dr. Shoemaker's expert judgment, "[I]t does not seem prudent to implement management 'experiments' that could potentially imperil a threatened or endangered species (TES), even under the rubric of adaptive management."[71]

D. The Revised ARM Framework Would Arbitrarily and Unjustifiably Remove Abundance Thresholds Below Which the Harvest of Female Horseshoe Crabs Is Prohibited.

The revised ARM Framework would arbitrarily lift the protective abundance thresholds intended to preserve the availability of food for red knots. Specifically, under the existing Framework, the female harvest quota is zero until the estimated abundance of female horseshoe crabs exceeds 11.2 million or the estimated abundance of red knots exceeds 81,900 in Delaware Bay.[72] These

---

[67] *E.g.*, *ARM Report* 156 tbl. 13 (illustrating that the catch multiple survey analysis for horseshoe crabs uses data starting from 2003). Compounding the chronological limitations on the data informing the model, the revised ARM Framework also imposes geographic constraints by including only data from Delaware Bay.

[68] FWS, *Species Status Assessment Report* 9.

[69] *Id.* at 28; FWS, *Draft Recovery Plan for the Rufa Red Knot* 14.

[70] *Shoemaker Expert Report* 11.

[71] *Id.*

[72] ASMFC, *ARM Report* 21.

thresholds reflect stakeholders' desire to take a precautionary approach to managing the delicate relationship between horseshoe crabs and red knots. Because neither species has reached its threshold since the original ARM Framework was implemented, the model has never recommended a female harvest. Under the revised ARM Framework, the model could (and likely would) recommend a significant female harvest even when neither red knot nor female horseshoe crab abundance has exceeded its protective threshold. Indeed, the Subcommittee's calculations show that the model would have recommended a female harvest of approximately 150,000 for 2017-2019, years when the original ARM Framework recommended a female harvest of zero.[73]

> 1. ASMFC Has Provided No Defensible Rationale for Removing the Protective Thresholds.

Removal of the protective thresholds received significant criticism in the minority opinions submitted by ARM Subcommittee members.[74] In rejecting these critiques, the Subcommittee relied on two primary arguments, neither of which is defensible.

First, the Subcommittee stated, "The presence of these threshold constraints in the utility function was criticized during this revision for not being consistent with adaptive management and optimization procedures and therefore they were removed from the utility functions."[75] But the Subcommittee's argument assumes that stakeholder values have no role in adaptive management, and that adaptive management is inconsistent with any constraint that arises from something other than an optimization model. This view squarely defies the adaptive management process as described in Addendum VII, which highly values stakeholder input, as explained in section IV.B of these comments. Moreover, the Subcommittee's view is internally inconsistent, as the revised ARM Framework appropriately maintains precautionary limits on the maximum harvest of male and female horseshoe crabs,[76] which represents a constraint on the model in deference to precautionary values. Thus, the revised ARM Framework is arbitrarily selective about its willingness to consider precautionary constraints.

Second, the Subcommittee described the thresholds as a "knife-edge utility function[]" and stated that, once the thresholds were exceeded, the existing ARM Framework would immediately recommend the maximum harvest package, with its female quota of 210,000.[77] According to the Subcommittee's calculations, the model is unlikely to ever select the interim harvest package, with a female quota of 140,000.[78]

The Subcommittee's argument misses the mark. The immediate issue is whether female harvest is allowed *below* the thresholds. The Subcommittee may have concerns about what

---

[73] ASMFC, *Supplemental ARM Report* 21 tbl. 11.

[74] *E.g.*, Wendy Walsh, *Walsh Minority Opinion* (in ASMFC, *ARM Report*) 113-14.

[75] ARM Subcommittee, *Majority Response to Niles* 107.

[76] ASMFC, *ARM Report* 81 ("[O]ne feature from the packages used in the original ARM version was retained: the maximum harvest for females was set to 210,000 and for males 500,000."). The Subcommittee pointed to these limits as an example of maintaining an "earlier decision[] made by stakeholders." ARM Subcommittee, *Majority Response to Walsh and Justification for Why Opinion Not Adopted* (in ASMFC, *ARM Report*) 125.

[77] ARM Subcommittee, *Majority Response to Walsh* 124.

[78] *Id.*

recommendations the current model would make in the unprecedented event that the thresholds were exceeded, but that is a separate question. In addition, if the current model would catapult over the interim harvest package and immediately recommend the maximum harvest package in the event that red knots or female horseshoe crabs met their abundance threshold, that would seem to indicate a defect in the existing model. A more reasonable correction would be to adjust the existing model to facilitate a gradual increase in female harvest recommendations once an abundance threshold is met. It is not at all clear why removing the thresholds altogether is a necessary or logical solution. Regardless, a potential defect in the current model's response to the achievement of protective thresholds for horseshoe crabs or red knots cannot offer any justification for eliminating the thresholds well before they are met. At the very least, the Subcommittee should have made its decision in consultation with stakeholders, not unilaterally.

2. The Elimination of the Protective Thresholds Illustrates the Improper Exclusion of Stakeholder Input.

In section IV.B, these comments detail why the exclusion of formal stakeholder input from the development of the revised ARM Framework was inappropriate and violated the requirements for adaptive management. This section explains why excluding stakeholders from decisions about the protective thresholds was particularly improper and contravened the views of the Commission's own experts and peer review panel.

During the Board's early consideration of developing Addendum VIII, the ARM Subcommittee Chair explained what process would be required to change (much less eliminate) the protective thresholds:

> [M]oving forward with this new Population Dynamics Model, where that threshold is at 11.2 million, you know that could change. It is a possibility to have a different utility function. *That is something that would have to be discussed amongst stakeholders* and among the ARM Workgroup members.[79]

Despite the Chair's acknowledgement that changing the female horseshoe crab threshold would require stakeholder input, the revised ARM Framework would eliminate the threshold even in the absence of stakeholder input.

The exclusion of stakeholders and elimination of the thresholds was criticized in the minority opinion of Subcommittee member (and Chair of the Delaware Bay Ecosystem Technical Committee) Dr. Wendy Walsh, the national lead for red knot recovery at FWS. Dr. Walsh meticulously detailed the role of stakeholder input in adaptive resource management and observed that the ARM Subcommittee had "failed to consult a broad array of stakeholders in the reinterpretation of previously agreed-upon objectives."[80] With respect to the abundance thresholds, Dr. Walsh explained:

---

[79] Comments of John Sweka, ARM Subcommittee Chair, *Proceedings of the Atlantic States Marine Fisheries Commission Horseshoe Crab Management Board* 5 (Oct. 29, 2019) (emphasis added), https://www.asmfc.org/uploads/file/5fb2ea02HorseshoeCrabBoardProceedingsOct2019.pdf.
[80] *Walsh Minority Opinion* 113.

These threshold values act as a constraint on female harvest, which was the express intent of the stakeholders. . . . [T]he formulation of these values as a constraint was an explicit and clear choice in the development of the existing framework. . . . [T]he high risk-aversion to female crab harvest by the stakeholders is clear, and thus it can be presumed that the new utility function . . . would be of considerable concern to those same stakeholders.[81]

The ASMFC-convened Peer Review Panel echoed these concerns. Recognizing that the Subcommittee had not convened stakeholders for this proceeding, the Panel tentatively stated that it "does not disagree" with the revised modeling functions, "as long as they truly reflect the objectives related to HSC harvest and REKN recovery and the risk associated with the HSC harvest."[82] The Panel reiterated its concern in its list of recommendations:

The new utility and harvest functions are a representation of values, and the Panel understands that convening a group of stakeholders for this revision was not possible. Therefore, the Panel recommends the WG fully consider whether the new utility and harvest functions represent stakeholder values as articulated in 2009.[83]

The rejection of Dr. Walsh's minority opinion indicated a troubling misunderstanding of the Subcommittee's assignment. The Subcommittee wrote that retaining the threshold values "is more consistent with a simple harvest control rule" and "would not be adaptive management and would not require the Framework developed in this assessment."[84] By this statement, the Subcommittee revealed that it viewed stakeholder input as an impediment to adaptive management—an obstacle to the Framework the Subcommittee had already devised. But as explained in more detail below in section IV.B, stakeholder input has consistently been recognized as the foundational step of adaptive management. There is no adaptive management without stakeholder input, and the revised ARM Framework is therefore not an exercise in adaptive management.

E. The Horseshoe Crab Population Estimates Are Improperly Based, in Large Part, on Two Surveys that Stakeholders Have Rejected.

The omission of stakeholder input was particularly harmful because it obscured stakeholder objections to new survey data upon which the revised ARM Framework extensively relies. Since its inception, the ARM Framework has based horseshoe crab abundance estimates entirely on data from the Virginia Tech Horseshoe Crab Trawl Survey, which reflected the original stakeholders' greater confidence in that survey compared to other surveys of horseshoe crabs in Delaware Bay. The Virginia Tech survey is purpose-designed to count horseshoe crabs, as opposed to general surveys that count horseshoe crabs just incidentally, and FWS has called it

[81] *Id.* at 113-14.

[82] ASMFC, *Horseshoe Crab Adaptive Resource Management Revision Peer Review Report* (in ASMFC, *ARM* Report) 10 (277 of PDF) ("*Peer Review Report*"). Significantly, the Peer Review Panel's tentative approval of the revised ARM Framework was uninformed by independent expert reviews such as those offered by Drs. Shoemaker and Lipcius in this comment process.

[83] *Id.* at 12.

[84] ARM Subcommittee, *Majority Response to Walsh* 122.

"the best benthic trawl survey to support the ARM."[85] Yet the revised ARM Framework would drastically downgrade the model's reliance on the Virginia Tech survey, rendering it one of three equally weighted surveys.[86] The two additional surveys that would comprise the abundance estimates—the New Jersey Ocean Trawl Survey and the Delaware Adult Trawl Survey—are general trawl surveys and not purpose-designed to count horseshoe crabs.

In her minority opinion, Dr. Walsh explained (as the Subcommittee acknowledged) that the revised approach would generate significantly higher abundance estimates,[87] which will lead to higher harvest recommendations for female horseshoe crabs. Dr. Walsh urged that, if the Subcommittee determined to rely upon all three surveys, it should at least accord greater weight to the Virginia Tech survey based on its "technical rigor and deliberate design" and "the high level of confidence that stakeholders have expressed in" it, among other reasons.[88] As Dr. Walsh noted, using all three surveys generates such high estimates that it would sometimes have resulted in female harvest recommendations even under the existing ARM Framework.[89]

The original decision to rely exclusively on the Virginia Tech survey reflected explicit stakeholder input. By introducing two additional surveys that stakeholders previously disfavored, and weighting all three surveys equally, the revised ARM Framework alters yet another stakeholder-driven component of the model without soliciting formal stakeholder input.

IV.   ASMFC HAS REPEATEDLY EXCLUDED INPUT FROM STAKEHOLDERS AND THE BROADER PUBLIC.

The development of Draft Addendum VIII omitted input from stakeholders and the public throughout the process. The Atlantic Coastal Fisheries Cooperative Management Act of 1993 requires the Commission to "provide[] adequate opportunity for public participation in the [fishery management] plan preparation process."[90] ASMFC has violated legal requirements and its own guidelines by severely limiting public participation in this proceeding. Specifically, the Commission held a public comment period before essential information was publicly available, failed to solicit formal stakeholder input, and decided to artificially limit its range of options to adopting Addendum VIII or reverting to Addendum VI—both of which would lead to resuming the female horseshoe crab harvest—without any public input whatsoever.

---

[85] FWS, *Rufa Red Knot Background Information and Threats Assessment* 247.
[86] ASMFC, *ARM Report* 55.
[87] Walsh Minority Opinion 111; ARM Subcommittee, *Majority Response to Walsh* 123 ("[I]t was noted in the 2019 assessment that equally weighting the surveys resulted in higher population estimates and that characterization by Walsh is accurate."); ASMFC, *Supplemental ARM Report* 21 tbl. 11 (for a comparison of abundance estimates under the current and proposed methodologies).
[88] *Walsh Minority Opinion* 111.
[89] *Id.* at 111-12.
[90] 16 U.S.C. § 5104(a)(2)(B).

A. ASMFC Held the Public Comment Period Before the Revised ARM Framework's
Core Model Was Publicly Available.

The public comment period for Addendum VIII occurred while crucial, material information was
being withheld from the public. Specifically, the public still has not been allowed to see the
model that generates bait harvest recommendations for horseshoe crabs in Delaware Bay.

New Jersey Audubon and Defenders of Wildlife requested the model on February 23, 2022, in
FOIA requests submitted to the U.S. Geological Survey ("USGS") and FWS, as well as a record
request submitted to ASMFC. While ASMFC provided certain components related to the
horseshoe crab estimates, USGS controls the core component that links horseshoe crabs and red
knots to generate harvest recommendations. In a letter prior to the Board's August 2022 meeting,
New Jersey Audubon and Defenders of Wildlife explained that USGS had not yet released the
model and urged the Board not to initiate the public comment period on Draft Addendum VIII
until the public could access the model that underlies the revised ARM Framework.[91] At the
Board meeting, several members expressed concern about the unavailability of the model, noted
USGS's stated intent to release the model following internal review,[92] and asked to be kept
apprised of developments in the public's access to the model.

As of September 30, 2022—the close of the public comment period on Draft Addendum VIII—
USGS has still not released the model. As a result, the public's ability to submit substantive
technical comments has been severely constrained. As this comment letter demonstrates, public
evaluation is essential for identifying significant issues for the Board's consideration. Indeed,
many of Dr. Shoemaker's critiques were enabled by the limited model components released by
ASMFC. But the preponderance of the model underlying the revised ARM Framework still has
not been subject to public evaluation. Dr. Shoemaker listed several questions that he could have
investigated more thoroughly if that model were available,[93] including:

- Does the red knot projection model outperform a null model that excludes any effect of
  horseshoe crab abundance?
- How much variation in apparent survival in the red knot IPM model is explained by the
  horseshoe crab effect compared to random among-year variation?
- Would an index of horseshoe crab egg density explain more variation in red knot survival
  and fecundity than the CMSA-derived estimate of horseshoe crab abundance?

While the Board should resolve the issues that have already been raised before further
considering Addendum VIII, it is impossible to anticipate all of the additional questions that will

---

[91] Letter from Benjamin Levitan, Earthjustice, to ASMFC Commissioners re *Consideration of Draft Addendum VIII
on the Implementation of Recommended Changes from 2021 Adaptive Resource Management Revision and Peer
Review Report for Public Comment* (July 26, 2022).

[92] In an email accompanying its denial of a Freedom of Information Act Request for the model, a U.S. Geological
Survey representative wrote, "We have withheld the two USGS models, but they and their associated use
publications will be published following the required USGS Fundamental Science Practices reviews." Email from
Janis Wilson, USGS, to Benjamin Levitan, Earthjustice, re: *FOIA: DOI-USGS-2022-002312 – Response* (July 28,
2022). On August 15, 2022, New Jersey Audubon and Defenders of Wildlife administratively appealed the denial of
access to the model, but USGS has not yet responded.

[93] *Shoemaker Expert Report* 26-27.

be identified once the model is released. New issues will inevitably arise. The proper time to address those questions is before the Board approves Addendum VIII. Enabling the public to identify additional questions only after the revised ARM Framework has been approved would subject red knots and horseshoe crabs to unacceptable risk and raise difficult administrative questions about how to limit the harm even as the Framework is in place.

<div style="text-align:center"></div>

       B.   The Subcommittee Violated ASMFC's Procedures by Failing to Solicit Formal Stakeholder Input.

The ARM Subcommittee's failure to solicit formal stakeholder input in this proceeding violated the principles and process of adaptive management. When the Board first approved the ARM Framework in Addendum VII more than a decade ago, stakeholder input was integral to the process. The *first sentence* of the "ARM Framework" section of Addendum VII was, "A goal of the ARM Framework is to transparently incorporate the views of stakeholders along with predictive modeling to assess the potential consequences of multiple, alternative management actions in the Delaware Bay Region."[94] The ARM Subcommittee expressed the same sentiment about the "ARM approach" in the current proceeding: "First, there is a great emphasis on complete elicitation of objectives and management actions from a full range of stakeholders."[95] The Subcommittee took that sentence verbatim from the Commission's Framework for Adaptive Management from 2009,[96] demonstrating how consistently stakeholder input has been acknowledged as the cornerstone of adaptive management.

The Board formalized the role of stakeholder input when it approved Addendum VII, which implemented an adaptive management framework for the Delaware Bay horseshoe crab fishery. Addendum VII required that the ARM Framework's "[i]mplementation *shall* be comprised of two cycles."[97] The *first step* of the "Longer Term Cycle," which was to occur "every 3 or 4 years," was to "[s]olicit formal stakeholder input on ARM Framework to be provided to the relevant technical committees."[98]

The ARM Subcommittee's failure to convene stakeholders in preparing Addendum VIII violated the Board's express requirements, as well as the principles underlying the adoption of adaptive management. And if the Board approves Addendum VIII, the exclusion of stakeholders is unlikely to be rectified anytime soon. Addendum VIII sets forth a default period of "every 9 or 10 years" for revising the ARM Framework, which "should incorporate" soliciting "formal stakeholder input."[99] Pursuant to that schedule, if the Board approves Addendum VIII in 2022—which it should not do—the ARM Framework will be due for a revision in the early 2030s. Assuming that stakeholders are formally consulted at that time (unlike this time), roughly *20*

---

[94] ASMFC, *Addendum VII to the Interstate Fishery Management Plan for Horseshoe Crabs for Public Comment: Adaptive Resource Management Framework* 2 (2012), https://www.asmfc.org/uploads/file/hscAddendumVII_Feb2012.pdf ("*Addendum VIII*").
[95] ASMFC, *ARM Report* 21.
[96] ASMFC, *Stock Assessment Report No. 09-02 (Supplement B): A Framework for Adaptive Management of Horseshoe Crab Harvest in the Delaware Bay Constrained by Red Knot Conservation* 1 (2009), https://www.asmfc.org/uploads/file/2009DelawareBayARMReport.pdf.
[97] ASMFC, *Addendum VII* at 4 (emphasis added).
[98] *Id.*
[99] ASMFC, *Horseshoe Crab Draft Addendum VIII for Public Comment* 8 (Aug. 2022).

years will have elapsed between such consultations, a striking contrast to the "3 or 4 year[]" interval required by Addendum VII. That would also mean that stakeholders would not be formally consulted for roughly *17 years* after FWS's 2015 determination to list red knots under the Endangered Species Act. While it is impossible to know all the ways that soliciting stakeholder input would have affected the current proceeding, the revised ARM Framework's elimination of the protective abundance thresholds (described above in section III.D.2) demonstrates that this concern is not merely theoretical.

It bears repeating how significantly the revised ARM Framework departs from the paradigm that the stakeholders accepted in preparation for Addendum VII, which instituted harvest recommendations based on the relationship between horseshoe crabs and red knots. The revised Framework would weaken that relationship almost to the point of nonexistence and recommend quotas accordingly. While presented as a technical update, the revised ARM Framework cannot plausibly be considered a reflection of the stakeholders' articulated values. At the very least, stakeholders should have been involved in designing a revised approach. Failure to involve them represents another reason for rejecting the current proposal.

C. Even Before the Public Comment Period, ASMFC Purported to Limit Its Options to Those that Would Reinitiate the Female Horseshoe Crab Harvest.

In addition to the inaccessibility of crucial information and the exclusion of stakeholder input, there was no public notice or comment for arguably the most critical decision presented by Draft Addendum VIII, which ASMFC now presents as a foregone conclusion: designating a reversion to Addendum VI as the "No Action" alternative if the Board does not approve Addendum VIII.[100] Addendum VI would increase the Bay-wide horseshoe crab harvest quota and *allow for the resumption of the female harvest* in Maryland and Virginia. Thus, the Board has effectively foreclosed public comment on the pressing question of *whether* to resume female harvest for this fishery. Under the terms of draft Addendum VIII, whichever option the Board selects—and regardless of any information that might surface during the public comment period—that decision is preordained.

On the merits, selecting Addendum VI as the "No Action" alternative was arbitrary, unnecessary, and misleading. Addendum VI would completely transform the management framework. The transition from Addendum VI to Addendum VII was arguably the most significant event in ASMFC's management of the horseshoe crab fishery, and reverting to Addendum VI would be equally significant.

To justify the selection of Addendum VI, Draft Addendum VIII indicates that Addendum VII is unavailable as the "No Action" alternative because the model underlying it was built on obsolete software and can no longer be utilized.[101] Even if the software is obsolete, that does not back the Board into a corner with no option but to adopt an addendum with a female harvest. The current ARM Framework has generated the same harvest quota for ten consecutive years, and the legitimate "No Action" alternative would be to apply the same quota to the 2023 fishing season. In fact, Addendum VII contains two "fallback option[s]" for when the data required to run the

---

[100] *Id.* at 5.
[101] *Id.*

ARM model are not available: use the quotas from Addendum VI *or* use the same quotas as the previous year.[102] It is unclear why the Board would have fewer options when the Addendum VII model cannot be run. The natural understanding of "No Action" would be to maintain the current status quo—i.e., the current addendum and current quotas—not to revert to an addendum and quotas that mark a major departure from the status quo.

At the August 2022 Board meeting, ASMFC staff explained that simply reusing last year's quotas is not appropriate because that would not qualify as "adaptive resource management."[103] Even if that were so, the solution should not be to reinstate the 12-year-old static quotas from Addendum VI. If the Board has authority to impose such a drastic change, then surely it has authority to continue relying on the most recent outputs of the current ARM Framework. It may be that neither option offers a satisfactory long-term solution, but the question now is what to do while questions about the revised ARM Framework are being addressed. The Board is not required to rush through a new (or old) addendum. It can temporarily maintain the current Framework to allow for thorough consideration of the appropriate next step, which clearly does not include accepting Addendum VIII as currently proposed.

V. THE FLAWS IN THE REVISED ARM FRAMEWORK MUST BE ADDRESSED NOW.

The Board's decision on Addendum VIII is highly consequential and could determine the course of the horseshoe crab fishery for many years to come. It is vital that the revised ARM Framework be subject to full vetting, and that foreseeable flaws be identified, prior to implementation by the Board. There will not be realistic opportunities to remedy defects in the revised ARM Framework in the future—at least not without imposing large burdens on both the Board and the public.

A. Flaws in the Revised ARM Framework Cannot Realistically Be Remedied at the Quota-Setting Stage.

At the Board's meeting in August 2022, some speakers observed that Addendum VIII will not, in itself, set binding quotas because the Board will retain discretion to deviate from the ARM Framework's harvest recommendations, and states will retain discretion to set quotas below those set by the Board.[104] But that is not a valid rationale for approving an addendum that has not been fully vetted and has been demonstrated to be flawed based on even the limited amount of information that has been made publicly available.

The purpose of the ARM process is to generate harvest recommendations based on rigorous science and sound policy.[105] As these comments detail, the revised ARM Framework incorporates many substantive and procedural flaws, and additional flaws are likely to emerge

---

[102] ASMFC, *Addendum VII* at 6.

[103] ASMFC, *Horseshoe Crab Management Board Proceedings Aug2022*, at 5:11, https://www.youtube.com/watch?v=OZvpdTTPj8c.

[104] *E.g.*, *id.* at 28:00, 1:12:57.

[105] 16 U.S.C. § 5104(a)(2)(B) (requirement in the Atlantic Coastal Fisheries Cooperative Management Act of 1993 for ASMFC to follow "standards and procedures to ensure that . . . [fishery management] plans promote the conservation of fish stocks throughout their ranges and are based on the best scientific information available.").

when the underlying model is released to, and evaluated by, the public. Regardless of the Board's or states' ability to deviate from those recommendations, the Board must ensure that the Framework represents the best available—and properly vetted—science and policy. To do otherwise would call into question the purpose of the ARM process and the harvest recommendations.

It would also not be practical for the Board or states to resolve the flaws in the revised ARM Framework at the quota-setting stage. If Addendum VIII were approved and the Board were unable to rely upon the Framework's flawed harvest recommendations, there would be no clear criteria or guidelines for establishing quotas, leading to a confusing, burdensome, and arbitrary quota-setting process. Similarly, if the Board approved Addendum VIII and adopted the revised ARM Framework's flawed harvest recommendations, states would need to determine the proper course in the absence of reliable information or direction from ASMFC. That would undermine the Horseshoe Crab Fishery Management Plan's purpose of creating "[a] coordinated and consistent management strategy."[106]

> B. <u>Flaws in the Revised ARM Framework Cannot Be Addressed Through Updates to the Model.</u>

While the revised ARM Framework can be "updated based on the annual routine data collected in the region,"[107] updates will not remedy its flaws. Many of the defects identified in these comments cannot be addressed by new data but rather demand a deeper restructuring of the model. For example, the model's miscalculation of the uncertainty in horseshoe crab abundance projections will persist despite new data. The same is true for all of the variables that are omitted from the model but indicate an unstable horseshoe crab population: egg density, prosomal width, sex ratio, etc.

Other defects would theoretically be alleviated by new data, but not on any relevant timescale. For example, the effect of the nonsensical horseshoe crab recruitment rates from the Virginia Tech gap years will gradually be diluted as new data are added, but they will continue to have perilously high influence for many years—realistically, for as long as Addendum VIII will be in effect. And even if, for the sake of argument, the estimated recruitment rate will slowly become more accurate over the years, that does not justify neglecting to fix a clear defect before implementing the revised ARM Framework.

Finally, some defects may be compounded by the addition of more data. As explained above in section III.C, the model is based entirely on data from when both horseshoe crabs and red knots had already crashed. It does not reflect the dynamics of a properly functioning ecosystem. As more data from the post-crash years are added, the model may only grow more confident that the current state of the ecosystem represents the norm. As Dr. Shoemaker observes, additional data may even yield a negative relationship between the abundance of horseshoe crabs and red knots, which would pose an existential problem for the Framework.[108]

---

[106] ASMFC, *Fishery Management Report No. 32 of the Atlantic States Marine Fisheries Commission: Interstate Fishery Management Plan for Horseshoe Crab* 1 (1998).
[107] ASMFC, *Draft Addendum VIII* at 8.
[108] *Shoemaker Expert Report* 10.

VI.   APPROVING ADDENDUM VIII WOULD LIKELY LEAD TO A VIOLATION OF THE ENDANGERED SPECIES ACT BY ASMFC.

In addition to the other bases for rejecting Addendum VIII discussed above, the Endangered Species Act provides a powerful further reason: adopting Addendum VIII would threaten to violate the federal prohibition against "taking" a threatened species. The ESA prohibits any person from "tak[ing] any [endangered] species within the United States or the territorial sea of the United States."[109] Such prohibited "taking" includes actions that "harm" listed species, including "significant habitat modification or degradation where it actually kills or injures wildlife by significantly impairing essential behavioral patterns, including breeding, feeding or sheltering."[110] The ESA's "taking" prohibition extends to governmental authorization to take protected species that facilitates such harm by "solicit[ing]" or "caus[ing]" an offense.[111] By regulation, that prohibition extends to the taking of most threatened species, including the red knot.[112]

A.   The Endangered Species Act Requires a Precautionary Approach.

In the Endangered Species Act, Congress adopted a precautionary approach. As the Supreme Court has stated, "Congress has spoken in the plainest of words, making it abundantly clear that the balance has been struck in favor of affording endangered species the highest of priorities, thereby adopting a policy which it described as 'institutionalized caution.'"[113] This principle is echoed in the ARM Framework's objective statement, which calls for "*ensur[ing]* that the abundance of horseshoe crabs is not limiting the red knot stopover population or slowing recovery."[114] Within the context of the ESA's legal framework, to ensure against such harms means taking a precautionary approach of "giv[ing] the benefit of the doubt to the species."[115] By setting ASMFC on a path to harm a threatened species whose population shows no sign of recovery, the revised ARM Framework would fall far short of ESA requirements and ASMFC's own objective.

As shown above, in many instances, Addendum VIII would enshrine a risk-prone approach instead of the risk-averse, precautionary approach required under the ESA. Even as it would allow the renewed harvest of female horseshoe crabs, Addendum VIII would utilize a model that, among other risky decisions:

- rejects the significant connection between horseshoe crabs and red knots,

---

[109] 16 U.S.C. § 1538(a)(1)(B).
[110] 50 C.F.R. § 17.3.
[111] *Strahan v. Coxe*, 127 F.3d 155, 163 (1st Cir. 1997); 16 U.S.C. § 1538(g).
[112] 50 C.F.R. § 17.31(a) (applying the provisions of § 17.21 (addressing endangered species) to threatened species); *id.* § 17.21(a), (c) ("[I]t is unlawful . . . to solicit another to commit or to cause to be committed" the taking of an endangered species.).
[113] *Tenn. Valley Auth. v. Hill*, 437 U.S. 153, 194 (1978).
[114] ASMFC, *ARM Report* 25 (emphasis added).
[115] *See, e.g., Roosevelt Campobello Int'l Park Comm'n v. U.S. Envtl. Prot. Agency*, 684 F.2d 1041 (1st Cir. 1982) (quotations and citation omitted); *see also Defs. of Wildlife v. U.S. Dep't of the Interior*, 931 F.3d 339, 351 (4th Cir. 2019) (same regarding scientific determinations).

- neglects egg-density data, which provide the most direct measure of the adequacy of food for red knots,
- rejects protective populations thresholds that were essential to the only group of stakeholders that ASMFC ever formally consulted about this matter,
- assumes that horseshoe crabs are recovering despite negative demographic trends, and
- uses horseshoe crab projections that fail to account for uncertainty and are scarcely more accurate than a null model.

The exclusion of public input at multiple stages of this proceeding exacerbates the risk of an ESA violation because ASMFC has evaded the public scrutiny that would be appropriate for such a consequential proceeding. A risk-averse approach would be to welcome public input in order to identify and address weaknesses that create unacceptable risk for the red knot. But the Board has taken a different, risk-prone approach: hastening a vote on Addendum VIII even as the underlying model continues to be withheld, despite record requests submitted more than seven months ago. The Board will therefore make a decision without the benefit of crucial public input and the important considerations such input would raise.

Both ASMFC and FWS suggest that the model will be improved by future updates.[116] As shown above in section V.B, updates cannot remedy the flaws in the revised ARM Framework. But even if they could, relying on future updates is not appropriate when an ecosystem is dangerously degraded and a threatened species hangs in balance. Future updates are likely to come too late.

B. By Utilizing the Revised ARM Framework, ASMFC Would Harm Red Knots.

Like any other association or governmental entity, ASMFC is subject to the ESA taking prohibition.[117] Under the Atlantic Coast Fisheries Cooperative Management Act of 1993,[118] ASMFC's fishery management plans are legally binding upon affected states. Once the Commission issues a plan, states "shall implement and enforce the measures of such plan within the timeframe established in the plan."[119] Because ASMFC's quotas cannot be exceeded, states have been prohibited from authorizing female horseshoe crab bait harvest in Delaware Bay under the existing framework. States may authorize a female bait harvest only if ASMFC sets a non-zero female harvest quota.[120]

---

[116] ASMFC, *Draft Addendum VIII* at 8; FWS, *U.S. Fish and Wildlife Service Evaluation of the Atlantic States Marine Fisheries Commission Horseshoe Crab/Red Knot Adaptive Resource Management Revision* at 3 of PDF (2022) ("*Evaluation*"), https://www.fws.gov/sites/default/files/documents/service-evaluation-of-atlantic-states-marine-fisheries-commission-horseshoe-crab-red-knot-adaptive-resource-management-revision.pdf.

[117] The ESA applies to any "person," which is broadly defined. 16 U.S.C. § 1532(13) ("The term 'person' means an individual, corporation, partnership, trust, association, or any other private entity; or any officer, employee, agent, department, or instrumentality of the Federal Government, of any State, municipality, or political subdivision of a State, or of any foreign government; any State, municipality, or political subdivision of a State; or any other entity subject to the jurisdiction of the United States.").

[118] Atlantic Coastal Fisheries Cooperative Management Act of 1993, Pub. L. 103-206, 107 Stat. 2419, Tit. VIII (codified at 16 U.S.C. § 5101 *et seq.*).

[119] *Id.* § 5104(b)(1).

[120] *Cf. Defs. of Wildlife v. U.S. Envtl. Prot. Agency*, 882 F.2d 1294, 1301 (8th Cir. 1989) (EPA's registration of pesticide effected a taking because the pesticide could not be used without such registration).

ASMFC's fishery management decisions therefore have a direct causal connection to the ultimate bait-harvesting actions that impact horseshoe crabs and red knots.[121] Indeed, the connection between the Board's management decisions and red knot demographics is the premise and intent of the ARM Framework's objective statement:

> 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.[122]

Draft Addendum VIII shows that, if the revised ARM Framework had been utilized in 2017-2019, it would have allowed for the harvest of around 150,000 female horseshoe crabs each year,[123] compared to the actual quota of zero for each of those years. Going forward, allowing such an increase in the harvest of female horseshoe crabs, upon which egg abundance depends, threatens significant degradation and modification of red knot habitat at Delaware Bay that would kill or injure red knots by significantly impairing breeding and feeding activities that are essential to the continued existence of the species.[124]

As explained above, the revised ARM Framework raises serious questions that the Board has not answered or publicly considered. After 24 years of ASMFC management, including 10 years under an ARM Framework, neither red knots nor horseshoe crabs are on a trajectory to recover. There are serious reasons to doubt even the modest increase in the horseshoe crab population that ASMFC reports. ASMFC's red knot abundance estimates are essentially flat at low numbers, while other estimates based on direct counting have shown a dangerous decline in recent years.

Now, in the Board's first addendum since red knots were listed as threatened, Addendum VIII would result in the increased harvest of horseshoe crabs, including the resumed harvest of females, thus magnifying the factors imperiling red knots. This poses an enormous risk to the ecosystem, which is precisely the wrong response to a species being listed under the ESA.

C. FWS's "Evaluation" Does Not Offer Independent Support for Addendum VIII.

Recent statements from FWS do not bolster the credibility of the revised ARM Framework. When FWS listed red knots as threatened under the ESA, it stated, "[A]s long as the ARM is in place and functioning as intended, ongoing HSC bait harvests should not be a threat to the red knot."[125] In her minority opinion raising concerns about the revised ARM Framework, Dr. Walsh

---

[121] *E.g.*, *Sierra Club v. Yeutter*, 926 F.2d 429, 438-39 (5th Cir. 1991) (holding that government agency violated ESA taking prohibition by authorizing logging that destroyed habitat and thereby impaired essential behavioral patterns of listed woodpecker species); *Loggerhead Turtle v. County Council of Volusia County*, 896 F. Supp. 1170, 1181-82 (M.D. Fla. 1995) (holding that county that regulates vehicular access to beaches is liable under ESA for taking of sea turtles caused by nighttime beach driving).

[122] ASMFC, *ARM Report* 25.

[123] ASMFC, *Draft Addendum VIII* at 12 app'x A tbl. 1 (showing annual female harvest quotas ranging from 144,803 to 154,483).

[124] 50 C.F.R. § 17.3 (defining "[h]arm").

[125] 79 Fed. Reg. at 73,709.

wrote that "[i]mmediate resumption of female harvest by the means described in the draft report may prompt the USFWS to reconsider if the ARM is functioning as intended."[126]

In contrast to Dr. Walsh's minority opinion, the document that FWS released on August 16, 2022, styled as an "evaluation" of the revised ARM Framework, did not offer any independent assessment of the revised ARM Framework. Rather, it repackaged the revised ARM Framework's modeling with all of its flaws detailed above, at times appearing to copy and paste figures directly from the Subcommittee's materials, and stated that the revision "poses negligible risk to red knot recovery and negligible risk of take under the Endangered Species Act."[127] Nowhere did FWS question the validity of the revised ARM Framework or any of the underlying assumptions or decisions, including on any of the bases discussed in these comments and accompanying expert reports.

With its complete deference to ASMFC's flawed modeling, assumptions, and conclusions, FWS unsurprisingly reached the same flawed result but did not bolster its validity. As these comments have shown, the revised ARM Framework incorporates numerous erroneous methodologies and assumptions. In its document, FWS propagated the same errors and replicated the same flaws as ASMFC. Moreover, since FWS relied on ASMFC's non-public model, its assertions are effectively unverifiable. The revised ARM Framework is unreliable for the reasons demonstrated in these comments. The Framework also still needs a legitimate, thorough, independent review based on all underlying information—not just the information released publicly to date. FWS's imprimatur does not resolve the defects of Addendum VIII.

VII. CONCLUSION

The window to save red knots is closing rapidly, especially for Southern wintering birds that fly the farthest and are most reliant upon horseshoe crab eggs at Delaware Bay. The revised ARM Framework would increase the pressure on this species, which is already vastly diminished on the beaches that once hosted its extraordinary migration. The Framework does not appreciate the importance of horseshoe crabs to red knots or the fragility of the horseshoe crab population itself. The weak relationship that it perceives between red knots and horseshoe crabs may well become a self-fulfilling prophecy, as the computer model continues to run while the ecosystem around it fades away.

The Horseshoe Crab Management Board has an obligation to restore red knots and horseshoe crabs at Delaware Bay. Just as importantly, it has a real—and maybe a final—opportunity to do so. For the reasons described above and in the attached expert reports, the Board should reject Addendum VIII.

---

[126] *Walsh Minority Opinion* 117.
[127] FWS, *Evaluation* at 3 of PDF. While the document is dated January 18, 2022, it was not released to the public until August 16. For an example of a copied figure, compare ASMFC, *Supplemental ARM Report* 30-31 figs. 10-11, with FWS, *Evaluation* at 5 of PDF fig. 1.

Review of 2021 ASMFC ARM revision

Kevin Shoemaker, Ph.D.,

Associate Professor, University of Nevada, Reno

September 2022

This is an expert review of the Adaptive Resource Management plan (ARM) proposed by ASMFC to guide management of the Horseshoe Crab fishery in Delaware Bay, performed by Kevin Shoemaker, Ph.D.

Dr. Shoemaker has a Ph.D. in Conservation Biology, a Master of Science degree in Conservation Biology, both from SUNY-ESF in Syracuse, NY, and a Bachelor of Science degree in Biology from Haverford College. He is a former Postdoctoral Fellow in the Department of Ecology and Evolution at Stony Brook University and a former Senior Scientist at Applied Biomathematics, an ecological research and development company located in Setauket, NY. Dr. Shoemaker is currently employed as an Associate Professor of Population Ecology at the University of Nevada, Reno. He has over 15 years of experience as a wildlife conservation scientist and has authored over 45 peer-reviewed scientific articles and book chapters on topics in wildlife ecology and conservation. He has expertise in Bayesian inference, machine learning, population ecology, and ecological modeling.

OVERVIEW

This report presents my review of the Adaptive Resource Management plan (ARM) proposed for use by the Atlantic States Marine Fisheries Commission (ASMFC) as a tool for guiding management of the horseshoe crab (HSC) fishery in Delaware Bay and protecting the Federally Threatened *Rufa* Red Knot (*Calidris canutus rufa*; REKN). Delaware Bay is a critical stopover site for REKN in their spring migration to breeding grounds in the high arctic from wintering grounds as far south as Tierra del Fuego (USFWS 2021). Specifically, HSC eggs deposited on coastal beaches provide a necessary high-calorie food resource for REKNs and other migrating shorebird

species as they replenish fat reserves depleted from their long migration and prepare for breeding. At the heart of the proposed ARM framework is an optimization model that provides harvest recommendations for female and male HSC, conditional on current estimates of HSC and REKN abundance. These recommendations are calibrated to maximize HSC harvest while causing minimal risk to the REKN population. The optimization model is based on a linked two-species simulation model (comprising a HSC and a REKN simulation model) that incorporates a one-way biotic interaction in which annual REKN survival and recruitment depend on female HSC abundance in Delaware Bay (among other covariates). While the stated objectives of the revised ARM are sensible, my review identified several concerns that suggest the revised ARM framework is not an appropriate tool for managing risk to HSC or REKN populations. Specifically, this report identifies six main areas of concern:

(1) **The fitted relationship between HSC abundance and REKN vital rates (survival and fecundity) is of insufficient magnitude to forecast a decline in mean projected REKN population growth even under a total collapse of the HSC population**. The extremely weak REKN/HSC relationship used in the revised ARM is inconsistent with previous research documenting HSC eggs as a critical food resource for migrating REKN and with the documented decline of the REKN population over recent decades, which experts have linked to increases in HSC bait harvest during the 1990s (Niles et al. 2009; USFWS 2014). If the REKN population model is inconsistent with what has been observed in the recent past, it seems unlikely to yield robust forecasts of future risk to the REKN population (or recovery of this population) from which to base management decisions. The inclusion of a REKN population model within the ARM framework (both the initial and revised versions) presupposes that HSC harvest could put REKN populations at risk, at least under some scenarios. As it stands, the apparent inability of the revised ARM model to predict a decline of the REKN population even under a total collapse of the HSC population seems to violate this premise, and practically guarantees that the REKN population model will play an insignificant role in setting optimal HSC harvest rates.

(2) **The HSC population simulation model fails to correctly propagate uncertainty about mean recruitment rates**. In specifying the bivariate normal distribution used to generate

annual male and female HSC recruitment rates (the most consequential empirically fitted parameters of the HSC simulation model), the proposed ARM framework treats incertitude about annual recruitment rates as representative of temporal process variance (natural year-to-year fluctuations) rather than as a mixture of parameter uncertainty and process variance (Link and Nichols 1994; Regan et al. 2002; McGowan et al. 2011). This subtle but significant shortcoming will tend to manifest in simulation replicates that closely resemble one another, since key sources of uncertainty "regress to the mean" (good years cancel out bad years) instead of propagating over time. The importance of this distinction is magnified for long-lived iteroparous species like HSC, since these populations tend to be resilient to short-term fluctuations in reproduction or recruitment (Lovich et al. 2015). When this issue is corrected (using the same Bayesian approach used to treat process variation and uncertainty in the REKN simulation models in the revised ARM framework), preliminary simulation results suggest a highly uncertain outlook for the HSC population in Delaware Bay, especially when faced with harvest pressures. In sharp contrast to the ARM report and supplement, the population of HSCs in Delaware Bay appears to have a substantial (17.5%) probability of falling below the lowest previously estimated levels even in the absence of all direct anthropogenic sources of mortality (bait harvest, biomedical bleeding and discard mortality) over the next 50 years. Furthermore, a scenario in which HSCs are harvested annually at the current maximum allowable rates is accompanied by a severe risk of decline (33.45%) and disruption to the population age structure (lower multiparous/primiparous ratios than previously observed). Finally, an extreme harvest scenario in which two million male and female HSCs are harvested each year results in near-certain catastrophic population collapse over the 50-year time horizon, in contrast to the (original) ARM report, which suggests a relatively stable HSC population even under this extreme scenario (which greatly exceeds current maximum allowable rates).

(3) **The Catch Multiple Survey Analysis (CMSA) exhibits poor fit to training and independent data, raising concerns about its use in projecting future HSC abundance**. Aside from being able to explain the apparent difference in mean HSC abundance before and after the "VT gap years" (see below; higher HSC abundance is both predicted and observed after the

period 2013-2016), the CMSA model explains very little, if any, of the observed variation in the primary data sources (three trawl surveys conducted in and around Delaware Bay). The CMSA results exhibit relatively good fit ($R^2 > 0.5$) to the recruitment data (primiparous abundance); however, this is unsurprising since there is only one source of data (VT swept area surveys) for estimating annual primiparous abundance versus three sources for estimating adult (multiparous) and total abundance. Given the overall lack of fit to training data, the HSC simulation model is unlikely to perform well for predicting independent validation data (data not used to fit the model). Indeed, when the CMSA results are challenged against the HSC spawning surveys – an independent estimate of HSC abundance for this region – there is no detectable relationship between these two independent estimates of HSC abundance. This lack of fit to both training and validation data raises concerns about the utility of the CMSA model, which informs all aspects of the proposed ARM, including the REKN IPM (where it represents the abundance of female HSC each year), the HSC projection model, and the annual harvest recommendation.

(4) **The "gap years" in the VT trawl survey data raise concerns about HSC recruitment estimates from the Catch Multiple Survey Analysis (CMSA).** As noted above, the CMSA is fundamental to all aspects of the proposed ARM framework. For the HSC population simulation models, the primary role of the CMSA is to parameterize HSC recruitment rates (which are the most consequential empirically derived inputs for the HSC simulation model). Unfortunately, of the three trawl surveys used to fit the CMSA models, the only survey that provides information for estimating recruitment – the Virginia Tech (VT) trawl surveys – was not conducted during a critical four-year period from 2013 to 2016 (hereafter referred to as the "VT gap", during which no direct information was available for estimating annual HSC recruitment rates). The CMSA results suggest that the HSC population underwent a substantial state transition during the VT gap years in which the population was small but stable prior to the gap, and larger and more variable after the gap. More concerningly, the CMSA predicts much higher average recruitment rates during the VT gap (for which no data are available for estimating recruitment) than at any single year before or after. The inflated average recruitment rates during the VT gap period are subsequently used for estimating

mean HSC recruitment rate for the HSC simulation models (thereby increasing estimated population resilience to harvest) – but unfortunately these high recruitment rates cannot be verified empirically. If average recruitment rates were computed from only those years in which recruitment could be verified empirically (i.e., excluding estimates from the VT gap years) the expected resilience of the HSC population to harvest would be substantially reduced.

(5) **The proposed ARM framework lacks 'null model' benchmarks and independent performance validation**. Null models are simplified representations of a system that lack many or all the proposed mechanisms that may help to explain the system dynamics; the typical null model in statistics assumes all observed variation is the result of a single random error process. By comparing complex models such as those used in the revised ARM with one or more null-model benchmark(s), researchers can determine whether the more complex models represent useful learned knowledge about a system (Koons et al. 2022). If a complex model fails to outperform a null model in terms of bias or precision (typically using independent validation data), the complex model is likely to be improperly specified or "overfitted" (whereby parameters are fitted to "noise" rather than true signal; Radosavljevic and Anderson 2014) and therefore not useful for prediction. The CMSA model fails to outperform even the simplest statistical null model (single intercept term with sampling error) for at least one data source (the VT swept-area estimate of female multiparous abundance). For the REKN component of the revised ARM, it would be informative to compare the performance of the REKN simulation model against a null model that omits any effect of female HSC abundance. It was recently demonstrated (Koons et al. 2022) that the ARM framework for guiding North American mallard harvest was unable to outperform a null model, and it would be instructive to pose a similar challenge to the REKN simulation model. If either model fails to outperform a null model, it should prompt managers to acknowledge that our current understanding of the effects of harvest on HSC populations remains insufficient for robust forecasting (Dietze 2017), and that a more precautionary approach may be warranted.

(6) **Lack of transparency**. The public still has no access to the data and code used for estimating

REKN population parameters, simulating REKN and HSC population dynamics, and running optimization routines (the CMSA code and data were made available). Without this data and code, it is difficult to fully assess the proposed ARM framework and to run scenario tests. If granted access to the code and data, there are a number of important null model tests (see above) and scenario tests that can be run, including (1) developing and testing the HSC and REKN models against a "null model" benchmark, (2) determining the 'optimal' female HSC harvest rates from the "canonical" versions of the HSC and REKN models in the absence of defined harvest limits, and (3) running the REKN simulation model under a scenario representing near-total collapse of the HSC population. The concerns identified above, which arise from analysis of the limited data and code made available to date, demonstrate, at a minimum, that such further testing is warranted. It seems prudent to delay implementation of the new ARM framework until the public and outside experts have had adequate time to scrutinize the statistical and simulation models that play such a central role in this proposed decision-making framework.

SUPPORTING EVIDENCE AND ANALYSES

The remainder of this report provides additional supporting details for the six major areas of concern identified above, including results and figures from re-analyses of the data presented in the ARM report.

1. The fitted relationship between HSC abundance and REKN vital rates (survival and fecundity) is of insufficient magnitude to forecast a decline in mean projected REKN population growth even under a total collapse of the HSC population

Including a model of REKN population dynamics as part of the previous and revised versions of the ARM framework implicitly acknowledges that reduction of the HSC population could, under some circumstances, have a negative impact on REKN populations. This assumption has a strong empirical basis, as multiple lines of evidence suggest that HSC eggs are an extremely important resource for migrating REKNs during their spring migration (e.g., Karpanty et al. 2006; Niles et al. 2009; USFWS 2014; USFWS 2021). Therefore, it is surprising that the fitted relationship between HSC abundance and REKN survival used in the revised ARM is very weak and appears to be

overwhelmed by random among-year variation (Fig. 47 from ARM Report; Fig. 9 from Supplemental Report; hereafter, I will use the notation 'ARM Fig. 47/9'). In fact, it appears from the ARM report that estimated REKN survival rates have generally decreased weakly over time despite an estimated increase in HSC abundance (ARM Fig. 44/7). Years with the lowest HSC abundance in the study period (at or near the lowest HSC abundances ever recorded in Delaware Bay) are coincident with the highest estimated REKN survival rates (ARM Fig. 47/9). Given this weak fitted relationship, simulated REKN abundance based on this model seems unlikely to be very sensitive to changes in HSC abundance. Indeed, a 'back of the envelope' calculation based on the REKN vital rates presented in the ARM report (and the slightly modified numbers presented in the Supplement) shows that the mean population growth rate (Lambda) of the REKN population is likely to remain at or above replacement levels (Lambda ≥ 1) even at HSC population size equal to zero (Fig. 1). This calculation was produced by using the mean survival from Supplemental Table 8, mean recruitment estimated from Supplemental Fig. 7b, and the standardized logistic regression coefficients from Supplemental Table 9 (effect size = 0.37 for survival and -0.14 for recruitment) to model REKN survival and recruitment as a function of HSC abundance. As a brief aside, the regression coefficients presented in the ARM report (e.g., effect of HSC on survival) are standardized and are on the logit (log-odds) scale, making them difficult to interpret. A quick example may help to aid interpretation of the effect size of this relationship: given a coefficient of 0.37 (the mean regression coefficient for the relationship between HSC abundance and REKN survival from the ARM Supplement, Table 8), a loss of 1 million female horseshoe crabs from Delaware Bay would result in REKN survival rate declining by only 0.004 (from 0.93 to 0.926). This is consistent with visual inspection of ARM Fig. 47/9.

Although I did not have access to the code and data used to fit the relationships between HSC abundance and REKN survival and recruitment, the relationships I used to generate Fig. 1 closely match the relationships presented in ARM Fig. 46/8 (Fig. 2). Interestingly, the value for mean recruitment provided in Supplemental Table 8 (ρ_{mean} = 0.063) yields a declining REKN population (Lambda = 0.99) even under average conditions from 2005 to 2017. Since this result is inconsistent with the reported Lambda of 1.04 during that same period from ARM Table 25 (and the generally increasing population trajectories indicated in ARM Fig. 58/15), I chose to use the

7

mean annual recruitment estimated from Supplemental Fig. 7b, which I calculated to be 0.109 (or geometric mean of 0.099). Using these mean recruitment values resulted in a Lambda of 1.035 (for arithmetic mean) or 1.027 (for geometric mean), more closely resembling but still below the reported baseline Lambda of 1.04 from the ARM report; setting baseline Lambda to 1.04 would only make a stronger case that REKN populations would not be expected to decline under an HSC population collapse (Fig. 1). This simulation exercise makes it very clear that the REKN model used in the revised ARM would not be able to predict or explain the decline in the REKN population observed during the 1990s, which has been attributed to unregulated harvest of HSCs in Delaware Bay (Niles et al. 2009; USFWS 2014). If this framework is unable to explain the decline of the REKN population in the first place, it does not appear to be an appropriate tool for helping to reverse the decline and promoting the recovery of this threatened subspecies.

Note that the population vital rates used to generate Fig. 1 represent point estimates. Because there was uncertainty associated with the estimate of Lambda (CI from 1.00 to 1.06; ARM Table 25), and with the effect size of HSC abundance on survival rate (CI from 0.12 to 0.63; ARM supplemental Table 9), some simulation runs (i.e., those with small Lambda and larger effect size sampled randomly from the joint posterior distribution) are likely to indicate REKN population decline at low HSC abundances. It is likely that these (probably rare) simulations drive the shape of the REKN "harvest function" yielded by the approximate dynamic programming algorithm. However, without access to the IPM and simulation code, I am not able to formally test the behavior of the REKN simulation model under scenarios of HSC population decline or collapse.

Scenario: HSC population collapse
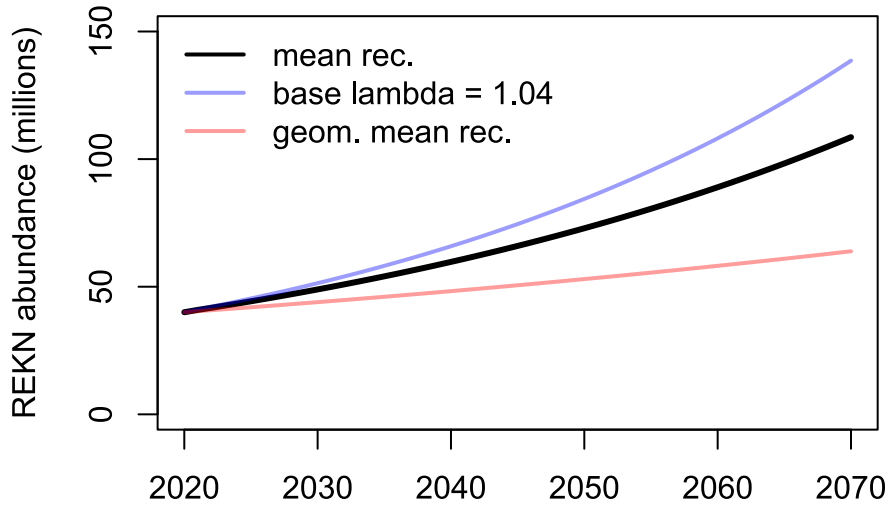
Figure 1. Results from a 'back of the envelope' calculation of REKN population growth under a scenario with depleted HSC population (female HSC abundance = 0 based on numbers presented in the ARM report. Mean recruitment rate was computed in three ways: arithmetic mean of values from ARM Supplemental Fig. 7b ("mean rec"), the geometric mean of these same values ("geom. mean rec."), and a value fitted to ensure a population growth rate (Lambda) of 1.04, as indicated in the ARM report. Although somewhat simplistic, this figure illustrates that the reduction in REKN survival due to the collapse of HSCs in Delaware Bay appears to be insufficient to induce a meaningful REKN population decline. This figure is based on a simple age-structured population model and does not incorporate a density-dependence mechanism (the revised ARM includes a density ceiling that prevents the REKN population from growing above ~150k).
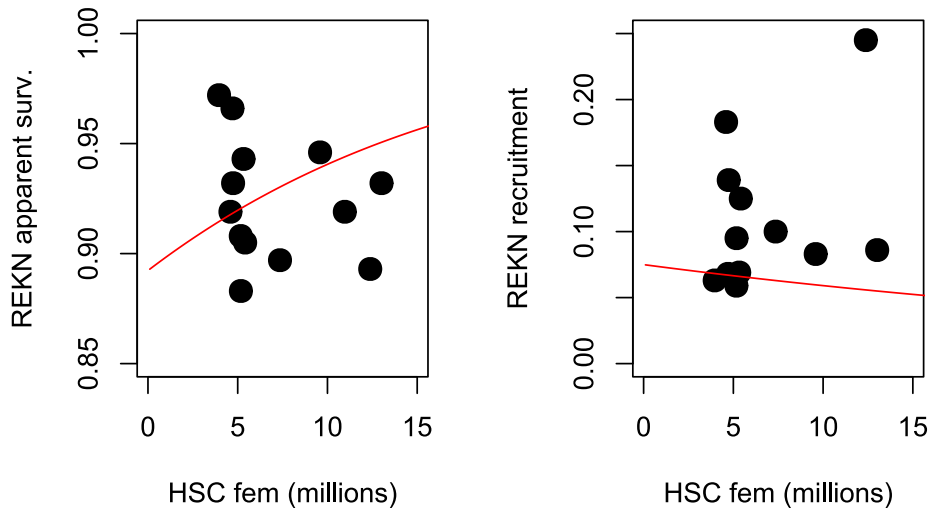
Figure 2. Relationships between female HSC abundance and REKN survival (left panel) and recruitment (right panel), recreated from information in the ARM supplemental report for the purpose of calculating the expected REKN population response to changes in the HSC population. Solid black dots represent annual vital rates estimated from ARM Supplement Fig. 9, and the red lines represent the fitted relationships presented in ARM Supplement Table 9.

Due to the weakness of the HSC/REKN relationship used in the revised ARM, and due to the complexity of the Integrated Population Model (IPM) framework used to represent the REKN population in the revised ARM, the relationship between HSC abundance and REKN population vital rates are likely to be unstable (sensitive to new data and alternative model specifications). Therefore, it is not implausible that the fitted relationship may disappear (become "non-significant") – or even flip sign to become a negative relationship – when the IPM is fitted to additional observations. This outcome would pose an existential problem for the ARM framework, decoupling the two-species framework and rendering the REKN model unusable in the context of management. There does not appear to be a contingency plan for this outcome. More generally, the REKN IPM appears to have gone through several distinct versions before researchers settled on a final set of decisions to incorporate into the final model (there are several important differences between an earlier version of the IPM presented in Tucker [2019] and the ARM report). Ideally, the results from alternative representations of the REKN system should be considered in aggregate to better represent structural uncertainty about this system (Williams 2011).

The linked two-species modeling framework in the revised ARM assumes the relationship between REKN and HSC is independent of REKN densities (i.e., it assumes a prey-dependent functional response). Under this assumption, larger REKN populations do not require larger abundances of HSC females (i.e., more HSC eggs deposited) to support adequate per-capita weight gain; in other words, the ARM model assumes that a REKN population of 40k would experience the same per-capita survival and fecundity as a population of 400k for a given abundance of female HSC. Implicitly, this assumes a lack of interference among REKN individuals, and no decline in the mean quality or accessibility of HSC egg resources at elevated REKN abundances (Karpanty et al. 2011). Some researchers have argued convincingly that a ratio-dependent functional response – in which per-capita prey consumption depends on the ratio between prey and predator abundances – is likely to be more realistic for simulation models with discrete time steps that span the entire reproductive periods of predator and prey (Abrams and Ginzburg 2000), such as the linked two-species model used in the revised ARM.

The previous ARM framework used data gathered from multiple sources of data outside

Delaware Bay to parameterize the simulation models. The revised ARM attempts to use Delaware Bay data sources wherever possible – which is a significant advance in many ways, as the revised ARM is "fine-tuned" for the system and can be updated relatively easily as new data are collected. However, this modeling decision also limits the analyses to a small geographic area over a short period of time, potentially ignoring relevant evidence from other regions and/or time periods. Furthermore, the time frame over which data are available for fitting the population models used in the revised ARM represents a limited scope of historical variation during which populations of REKN and HSC were relatively small in comparison with earlier estimates. Using these models to forecast system dynamics under conditions outside the range of values used to fit the model (e.g., lower HSC abundances, higher REKN abundances) therefore requires extrapolation, which can be highly uncertain (and often inaccurate). Since both the HSC and REKN simulation models tend to produce forecasts that differ from current conditions (e.g., larger numbers of both species), and because the optimization routine relies on these simulated results, the management recommendations emerging from the revised ARM rely on highly uncertain extrapolations about HSC and REKN population dynamics and about how these two species may interact (analogous to extrapolations of species and community distributions under climate change; Araujo and Rahbek 2009). On one hand, the ARM framework is designed to be able to refine management policies as new data become available and as sources of uncertainty are reduced (Nichols et al. 2007). On the other hand, it does not seem prudent to implement management "experiments" that could potentially imperil a threatened or endangered species (TES), even under the rubric of adaptive management.

In summary, the relationship between HSC abundance and REKN survival appears to be too weak to induce a decline in REKN abundance (Fig. 1). If all HSCs in Delaware Bay disappeared today, the model would continue to predict a generally stable or increasing population of REKN over the next 50 years. Therefore, the revised ARM model would be unable to predict the decline of REKNs that was observed in recent decades, and which has been attributed in part to the decline in the HSC population (Niles et al. 2009; USFWS 2014). This lack of consistency between the revised ARM model and recent historical observations raises significant doubts about the ability of this model to accurately reflect future risks to the REKN population or to guide HSC

harvest decisions in a way that promotes REKN survival and recovery. Furthermore, the decision to include a REKN population model as part of the ARM framework (in both the original and revised versions) presupposes that HSC harvest could result in risk to the REKN population; the apparent inability of the ARM model to predict a decline in REKN abundance under a total HSC population collapse violates this premise and undermines the apparent purpose of the model.

2. The HSC population simulation model fails to propagate uncertainty about mean recruitment rates

The HSC recruitment process is the most consequential empirically fitted component of the HSC simulation model. Other elements of the HSC simulation model are not fitted to data – for example, natural mortality rate, the biomedical mortality rate, and bait harvest rates are fixed by the modelers. In the revised ARM, the recruitment process is fitted to data indirectly via the CMSA model; annual male and female recruitment estimates were used to fit a bivariate log-normal distribution (defined by a mean and standard deviation for each sex, along with a covariance between sexes – all on a logarithmic scale), which was then used to represent annual recruitment in the simulation model. The only other parameter fitted in the CMSA model – initial abundance – is not directly used in the simulation model. Recruitment is critical for any assessment of population resilience to harvest, since (in the absence of immigration, which is not included in the revised ARM), it is the only process that enables the population to overcome sources of mortality. Therefore, it is not surprising that the HSC simulation model is highly sensitive to changes in mean (log) fecundity (ARM Fig. 33; note that when I omit any reference to the supplemental report, I am referring to the primary ARM report). Given the high sensitivity of the HSC simulation model to the (log) mean HSC recruitment for males and females, it is critical that uncertainty about these parameters is properly represented in simulation models. However, the revised ARM framework incorrectly treats incertitude about annual recruitment rates as representative of temporal process variance (natural year-to-year fluctuations) rather than as a mixture of parameter uncertainty and process variance (Link and Nichols 1994; Regan et al. 2002; McGowan et al. 2011). This is a subtle but consequential error, as sources of uncertainty will tend to "regress to the mean" (with good years cancelling bad years) instead of propagating over time.

To estimate the parameters for the log-normal recruitment process in the revised ARM, the following steps were taken: (1) log-normal distributions were separately fitted to each estimate of primiparous abundance (separately for each year and sex), based on estimates of parameter uncertainty (95% confidence intervals) derived from the CMSA results, (2) this collection of lognormal distributions (representing parameter uncertainty) was used to simulate annual male and female primiparous abundance for the years represented in the CMSA model (confusing parameter uncertainty with temporal process variation), and then (3) data from these simulations were used to fit a bivariate lognormal distribution (via maximum likelihood) for representing annual HSC recruitment in the ARM model. In general, parameter uncertainty should be represented in simulation models by drawing a single sample per replicate from a distribution of values representing parameter uncertainty (or by running replicates with "worst-case" and "best case" values for key parameters). However, the "canonical" version of the HSC projection model fails to address parameter uncertainty – most notably, uncertainty about the mean HSC recruitment rate, to which the HSC projection model is highly sensitive (ARM Fig. 33). Therefore, there is more uncertainty about the future of the HSC population in Delaware Bay than the revised ARM acknowledges. It is important to note that a sensitivity analysis was run in which expected recruitment was allowed to vary across simulation replicates within ca. 5% or 10% of the median recruitment value. This sensitivity test demonstrates an appropriate method for modeling parameter uncertainty; however, this test fails to represent the extent of uncertainty about the median HSC recruitment, which extends far beyond 10% of the mean estimated value (Fig. 3). Furthermore, this treatment of uncertainty was only run as a scenario test and was omitted from the 'canonical' version of the ARM that is proposed for use in managing the HSC harvest in Delaware Bay.

Interestingly, the REKN projection model in the revised ARM appears to represent parameter uncertainty appropriately. The key parameters of the REKN model were estimated using an Integrated Population Model (IPM), which were fitted in a Bayesian framework. In this framework, parameter uncertainty is represented by a joint posterior distribution that embodies the set of values that are consistent with the observed data. Furthermore, temporal process variation in the REKN population model is treated by explicitly modeling annual variability in key

vital rates (survival and recruitment) via annual random effects fitted with hyperparameters (Kery and Schaub 2011). This Bayesian hierarchical approach enables parameter uncertainty and process variation to be interpreted and modeled separately in a straightforward and intuitive manner. Specifically, parameter uncertainty is incorporated by running multiple replicates with different values drawn from the joint posterior distribution, and temporal process variation is included by sampling from the hyperparameters across years within each replicate (Goodman 2002).

To enable sensible propagation of parameter uncertainty in the HSC simulation model (analogous to the REKN model in the ARM), I constructed and fitted a hierarchical Bayesian version of the CMSA model. This model was fitted using the same data and model structure as the CMSA model included in the revised ARM. However, instead of estimating annual recruitment separately for each year and sex, the Bayesian CMSA model included an explicit representation of temporal process variance in recruitment (i.e., a "random effect" describing inter-annual variation in recruitment). This temporal process model was specified using a bivariate lognormal distribution exactly analogous to the HSC simulation model included in the ARM model, which included "hyperparameters" for male and female (log) mean recruitment, male and female (log) standard deviation, and a correlation term. By estimating temporal process variation directly, the Bayesian CMSA closely mirrors the HSC simulation model (analogous to the direct relationship between the IPM and the REKN simulation model), circumventing the multi-step process used in the ARM to generate the bivariate lognormal distribution from the CMSA results, and (most importantly) enabling the parameters of the bivariate lognormal distribution to be estimated directly from the data. To simulate HSC abundance over time, parameters for each replicate were drawn from the joint posterior distribution (representing parameter uncertainty), and temporal process variation within each replicate was simulated by sampling from the bivariate lognormal distribution. For the simulations, I incorporated the same restrictions in the stock-recruitment relationships indicated in the ARM report (driven by abundance and sex ratios for the years in which recruits were expected to have hatched).

Results from the Bayesian CMSA model indicate substantial uncertainty around mean HSC recruitment rates for both males and females (Fig. 3). Simulations (50 year time horizon) from

this model in the absence of any direct anthropogenic sources of mortality (no bait harvest, biomedical mortality or discard mortality) indicate that the future of the HSC population in Delaware Bay is uncertain; the population has a 17.4% chance of declining below 4 million females (combined multiparous and primiparous abundance) at least once in the next 50 years, equivalent to the lowest abundances estimated from 2003 – 2019 (period for which the CMSA model was fitted) (Fig. 4). This no-harvest scenario also had a 3.8% probability of falling below 3 million females over the 50-year simulation, well below any estimate from the VT swept area surveys. In contrast, the HSC projection model in the revised ARM indicates a large and sustainable HSC population under a scenario with no bait harvest but including other anthropogenic sources of mortality including biomedical harvest and discard mortality (ARM Fig. 30; note that this figure does not reflect changes in mean HSC recruitment following peer review—the Supplement does not update this figure but contains other figures indicating a sustainable HSC abundance even with a bait harvest; Supplemental Fig. 15). Simulations from the Bayesian CMSA also indicate a much higher probability of decline under a scenario in which males and females are harvested at their respective maximum allowable rates (but are not subject to biomedical and discard mortality); this scenario had a 33% probability of declining below 4 million females over the next 50 years, 11% probability of declining below 3 million females, and a 2% probability of declining below 2 million females (Fig. 4). This scenario also appeared to disrupt the age structure in many simulations, resulting in fewer multiparous adults than primiparous adults. In contrast, the HSC simulation model in the revised ARM suggests a stable or increasing HSC population even under maximum allowable harvest scenarios that also include biomedical and discard mortality (ARM Fig. 31; see above caveat). Finally, a scenario in which both female and male HSCs were harvested at a rate of 2 million per year (much higher than the current maximum rate) results in a high probability of decline or even extirpation over the 50-year simulation; there was a >99% probability of declining to below 3 million females, a 92% probability of declining below 1 million females, and a 12% chance of falling below 10k females (Fig. 4). In contrast, the HSC simulation model in the revised ARM predicted a relatively sustainable population of HSC even under this extreme scenario, with no risk of population collapse (ARM Fig. 32; note that the HSC simulation model in the supplemental report may not

sustain this level of harvest due to the reduced mean recruitment rate relative to the model used to generate ARM Fig. 32).
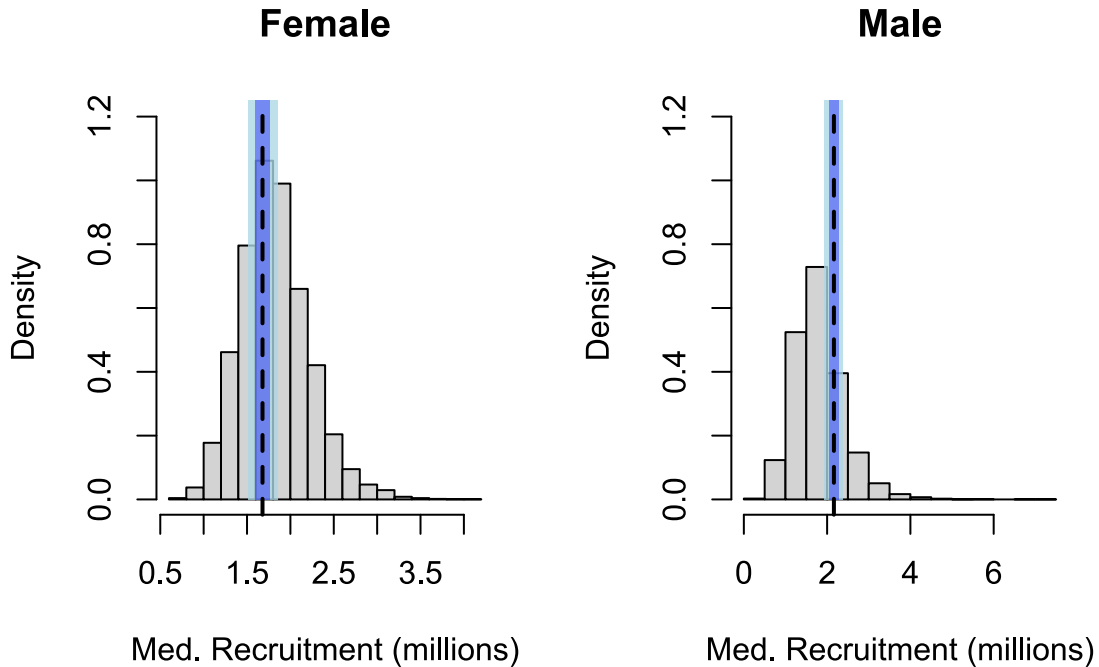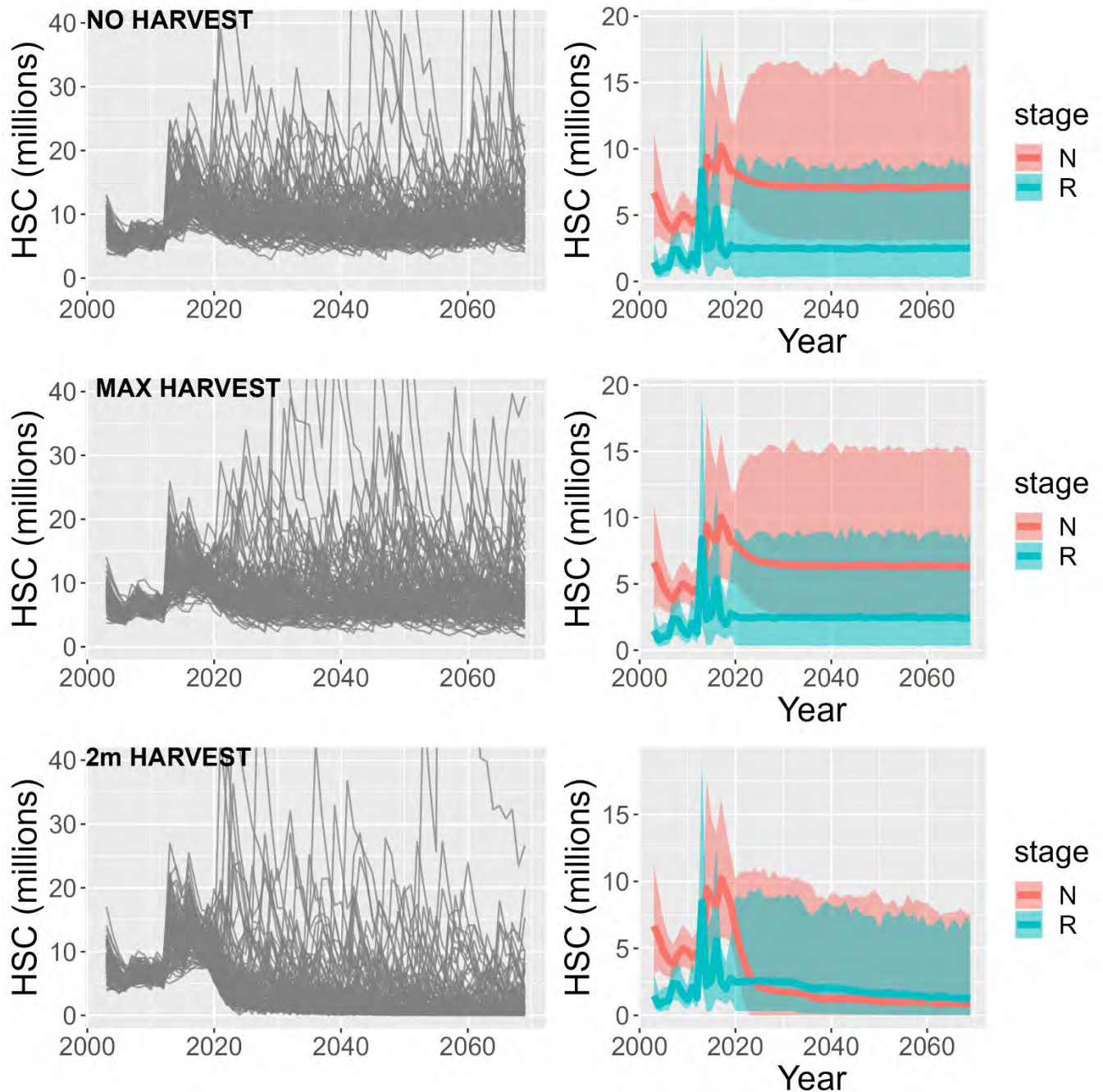

Figure 3. Posterior distributions representing parameter uncertainty for median female and male HSC recruitment rates, fitted using a Bayesian reanalysis of the CMSA model from the revised ARM (same data and model structure used to fit the CMSA model). Vertical dashed lines denote the median HSC recruitment values used in the base HSC projection model in the revised ARM. Light and darker blue shaded polygons represent the "added variation in expected recruitment" sensitivity tests from the ARM report (e.g., Fig. 69, 70). Note that the true range of parameter uncertainty falls well beyond the bounds of these sensitivity tests.

This critique is focused primarily on uncertainty about the annual HSC recruitment (primiparous abundance) parameters since they represent the ultimate source of projected resilience (or non-resilience) to harvest pressures and are therefore the most consequential fitted parameters in the CMSA simulation model. However, there are several other sources of uncertainty that should be accounted for in the HSC simulations. For example, natural mortality of HSC is set at exactly 0.3 (30%) across all sexes and age classes (primiparous and multiparous) in the revised ARM model, whereas there is substantial uncertainty about this parameter. The value of 0.3 was based on tag recovery data (assuming negligible harvest), but other lines of evidence seem to suggest natural mortality may be closer to 20% or even lower (as noted in the ARM

report). Lower estimates of mortality (higher survival and greater longevity) could imply lower resilience to harvest of adults (Midwood et al. 2015). Interestingly, natural mortality is an estimable parameter in the CMSA model; when modeled as a free parameter in the Bayesian CMSA, the model suggests that natural mortality is lower than 30%, but higher for females than males (note that Figs 3 and 4 are based on a model with natural mortality set at 30%, to match the ARM models). Other sources of uncertainty in the HSC population model include discard mortality (where 5% mortality was assumed for trawl and dredge surveys, while 12% mortality applied for gill nets) and biomedical mortality (assumed to be 15%). Although the ARM report documents a limited set of sensitivity analyses that were designed to test the degree to which key results changed under alternative parameter values (including mortality; ARM Table 18, 19), the relatively small set of sensitivity tests does not appear to comprehensively address these sources of uncertainty and seem inadequate for characterizing uncertainty about this system. Furthermore, uncertainty about these processes is not propagated through the HSC projection models.

In summary, if sources of error in the recruitment process are properly accounted for, the outlook for the HSC population in Delaware Bay is uncertain even in the absence of any harvest pressures. Based on a reanalysis of the existing data (using the same model specification used in the CMSA and HSC projection model), I found that harvest at the current maximum allowable rates has a high risk (11%) of causing the female HSC population to decline below the lowest levels ever recorded (3 million females). The HSC population models presented in the ARM report and supplement are not useful because they mis-characterize the risk of harvest pressures to the HSC population in Delaware Bay.

Figure 4. Female HSC population simulations run using fitted parameters (joint posterior distribution) from a Bayesian CMSA model, with uncertainty propagation performed in a manner analogous to the REKN projection model. The top row depicts simulations run under a no exploitation scenario (no bait harvest nor biomedical/discard mortality), the middle row depicts maximum allowable harvest rates (but also without biomedical and discard mortality), and the bottom row depicts an extreme harvest scenario (2 million females, 2 million males harvested annually). The left-hand panels depict trajectories of total abundance (primiparous and multiparous) for individual simulation replicates. Right-hand panels depict the 95% credible intervals for primiparous abundance (R) and multiparous abundance (N). None of these scenarios include biomedical or discard mortality.

3. The Catch Multiple Survey Analysis (CMSA) appears to exhibit poor fit to both training and independent data, raising concerns about its use in projecting future HSC abundance

The CMSA model explains little (and, in at least one case, none) of the variation in the data sources used to train this model (comprising three different trawl surveys conducted in and around Delaware Bay; here I present results for the female CMSA only) (Fig. 5). Notably, the CMSA performs worse than a statistical null model (all variation is assumed to be random "noise") for predicting the multiparous female abundance estimated from the VT trawl surveys, with R^2 of -0.42 for the full time series (negative R-squared value indicates the CMSA model performs worse than the null model). In contrast, the CMSA results appear to exhibit relatively good fit ($R^2 > 0.5$) to the recruitment data (primiparous abundance) from the VT trawl surveys (Fig. 5; ARM Fig. 21). However, this is not a fair test; with only one source of data for estimating annual primiparous abundance (the VT trawl surveys) – and with a separate recruitment parameter fitted for each year – the CMSA recruitment results are practically guaranteed to resemble the observed recruitment data.

For the remainder of the datasets used to train the CMSA (DE and NJ trawls), it is instructive to note that the majority of the observed variance 'explained' can be attributed to the apparent difference in mean HSC abundance before and after the period 2013-2016 (during which the Virginia Tech trawl surveys were not conducted and therefore no estimates of recruitment were available; hereafter, "VT gap", see below). Indeed, for the DE surveys the R-squared value drops to negative values for the periods before (R^2 = -0.07) and after (R^2 = -0.03) the VT gap period (versus R^2 = 0.14 for the full time series). Similarly, for the NJ trawl survey, the R-squared value drops to 0.11 for the period before the gap and falls below zero for the period after the VT gap (R^2 = -0.05; compared to R^2 = 0.57 for the full time series). More concerningly, the CMSA can "explain" the apparent increase in the HSC population after the VT gap period only by estimating extremely high recruitment during the VT gap period (during which no recruitment information was available; see below for more details). Because no data were available for fitting recruitment (primiparous abundance) during the VT gap, the CMSA model was free to "fill in" whatever recruitment estimates produced the best match to available data (DE and NJ surveys were the only available data sources during this period)—even if these recruitment estimates

were unrealistically high or low (with no data available for comparison, there was no penalty for producing unrealistic estimates). If the CMSA is only able to fit the training data via unrealistic estimates of recruitment (see below), this strongly suggests a poorly specified model and raises serious doubts about using the CMSA results to represent and forecast the HSC population in Delaware Bay.
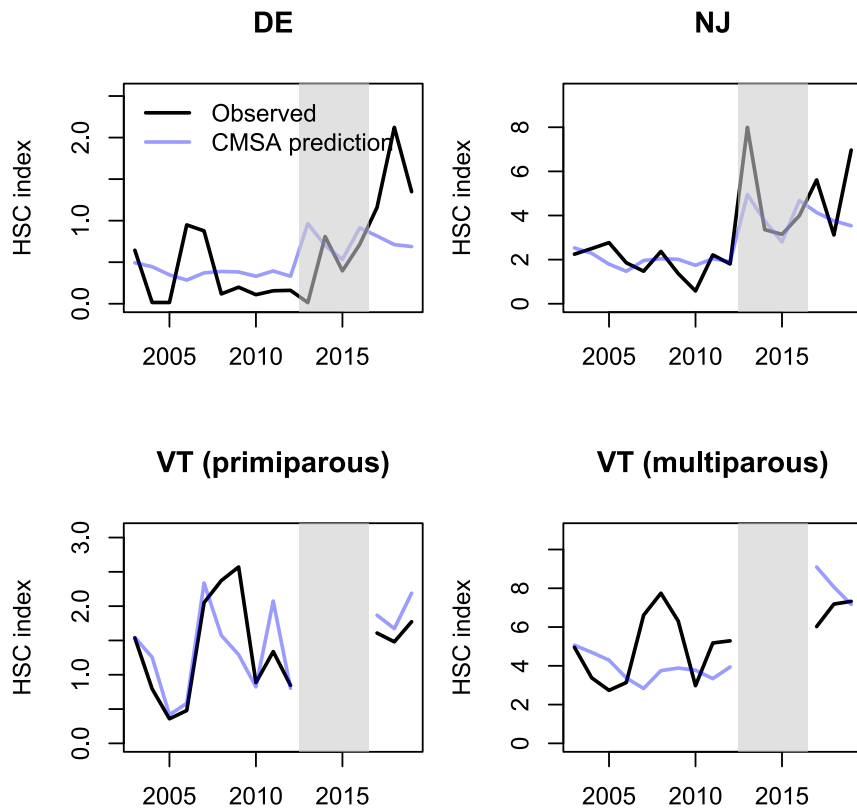
Figure 5. *Illustration of the fit of the CMSA model to data on female HSC abundance derived from three trawl surveys: DE, NJ, and VT (the same sources of data that were used to fit the CMSA model). This figure presents the same information as ARM Fig. 21/4. The CMSA model performs well in predicting primiparous abundance (bottom left) but exhibits poorer performance for predicting adult (multiparous) abundance (bottom right) or total abundance (top row). The CMSA predicts little to no variation in adult/total abundance besides the difference in apparent mean abundance before and after the "VT gap years" (gray regions).*

Given the lack of fit to training data, the HSC simulation model is unlikely to perform well when predicting to independent validation data (data not used to fit the model). Indeed, when the CMSA results are challenged against the Delaware Bay HSC Spawning Surveys (e.g., Zimmerman et al. 2020; https://www.delawarebayhscsurvey.org/), which provides an independent estimate of relative HSC abundance for this region, there is no detectable

relationship between these two independent estimates of HSC abundance (Fig. 6). This lack of fit to both training and validation data raises doubt about the utility of the CMSA results, which are central to all aspects of the proposed ARM, from fitting the HSC/REKN relationship to forecasting HSC abundance, to guiding annual decisions about HSC bait harvest.
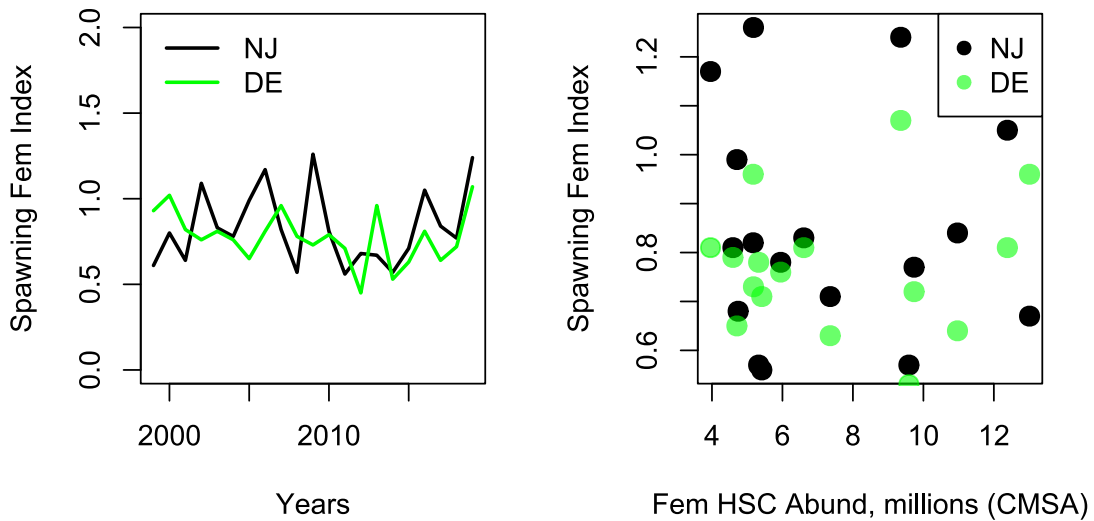
Figure 6. *Comparisons of standardized HSC spawning female counts from DE and NJ beaches (an index of relative female HSC abundance analogous to trawl surveys) with (left) each other and (right) with the CMSA estimates of female HSC abundance in Delaware Bay (in millions). The two spawning surveys exhibit very little correlation between the NJ and DE sides of Delaware Bay from 1999 to 2018 (left panel; correlation = 0.25). In addition, there is no detectable relationship between spawning counts (on either the NJ or DE sides) and CMSA estimates of female HSC abundance (right panel).*

In summary, the CMSA model does not perform well when predicting to the training data (the three sources of data used to fit the model). Although the model can explain some of the apparent difference in mean HSC abundance before and after the 'VT gap years', this 'ability' is driven by inflated recruitment rate estimates during the VT gap years that cannot be verified empirically (see below). Furthermore, the CMSA model explains virtually none of the observed variation in HSC spawning abundance from the same period, which represents an independent index of HSC population size. The poor performance of the CMSA model in predicting observed variations in HSC abundance in Delaware Bay calls into question the utility of this model – which is central to all aspects of the ARM model – as a robust system for characterizing and predicting

the HSC population in Delaware Bay.

4. The "gap years" in the VT trawl survey data raise concerns about HSC recruitment estimates from the Catch Multiple Survey Analysis (CMSA)

As noted previously, the CMSA is fundamental to the proposed ARM framework. For the HSC population simulation models, the primary role of the CMSA is to parameterize HSC recruitment rates (which are the most consequential empirically derived inputs for the HSC simulation model). Unfortunately, of the three trawl surveys used to fit the CMSA models, the only survey that provides information for estimating recruitment – the Virginia Tech (VT) trawl surveys – was not conducted during a critical four-year period from 2013 to 2016 (referred to in this report as the "VT gap", during which no direct information was available for estimating annual HSC recruitment; note that the missing survey years were actually 2012-2015, but the VT results were lagged forward within the CMSA to ensure comparability with the DE and VT trawls). The lack of information on primiparous abundance during the VT gap years leads to several nonsensical results in the CMSA model. For example, in one year (2013; the first VT gap year) the estimated number of new female recruits is near 10 million – approximately 8 times larger than the average estimated recruitment rate from the 10-year period from 2003 to 2012 and 4 times larger than the maximum estimate during this 10-year time frame (ARM Supplemental Table 3). The following year (2014), the point estimate for primiparous abundance goes down to 2, i.e., 2 primiparous female individuals across Delaware Bay. Furthermore, the standard error estimates for primiparous abundance during the VT gap years are very large – in fact, the upper bound on the confidence intervals approaches infinity for one year (2014).

The CMSA results suggest that the HSC population underwent a substantial state transition during the VT gap years in which the population was small but stable prior to the gap, and larger and more variable after the gap. In the fitted CMSA model, this state transition appears to be driven by extremely high recruitment rates during the VT gap years. Concerningly, the CMSA model (including the Bayesian version of the CMSA model described above) predicts much higher mean annual recruitment rates during the VT gap (for which no data are available for estimating recruitment) than at any single year before or after (Fig. 7). Specifically, mean

annual recruitment during the VT gap years was estimated at 4.2 million (using the arithmetic mean, per the ARM report), versus 1.2 million before the gap and 1.9 million after the gap (using the geometric mean to represent the median of a lognormally distributed sample, per the ARM report). The inflated mean recruitment rates during the VT gap period are subsequently used for estimating the average HSC recruitment rate for the HSC simulation models (thereby increasing estimated population resilience to harvest) – but unfortunately these high recruitment rates cannot be verified empirically.

In summary, the CMSA model estimates abnormally high annual recruitment rates during the VT gap years (Fig. 7). These very high estimates are unverifiable, as no data on HSC recruitment was collected during these years. In the original ARM report, the average annual recruitment used in the HSC simulation model relied heavily on the inflated estimates of recruitment during the VT gap years, discounting the pre-gap years entirely. After peer-review, the ARM was altered to consider all years instead of discarding lower estimates from the pre-gap years. Nonetheless, the revised ARM model continues to treat the mean recruitment rate during the VT gap as reliable, allowing these inflated estimates to contribute to the estimate of average annual HSC recruitment used for the HSC simulation models (which are highly sensitive to the estimate of average recruitment; ARM Fig. 33). If the extremely high recruitment estimates during the VT gap years were to be excluded from this estimation process out of precaution, the average annual HSC recruitment rate would drop substantially (Fig. 7), further reducing the expected resilience of this population to harvest pressures. Ultimately, the inflated estimates of recruitment during the VT gap years are likely to be an artifact of the CMSA model specification (and the lack of data on recruitment for those years) and are unlikely to be reflective of true HSC recruitment rates. However, there remains no way to verify HSC recruitment rates during this period. Given this uncertainty, a conservative (precautionary) approach would be to exclude the VT gap years when computing recruitment for the HSC population simulations (Fig. 7).
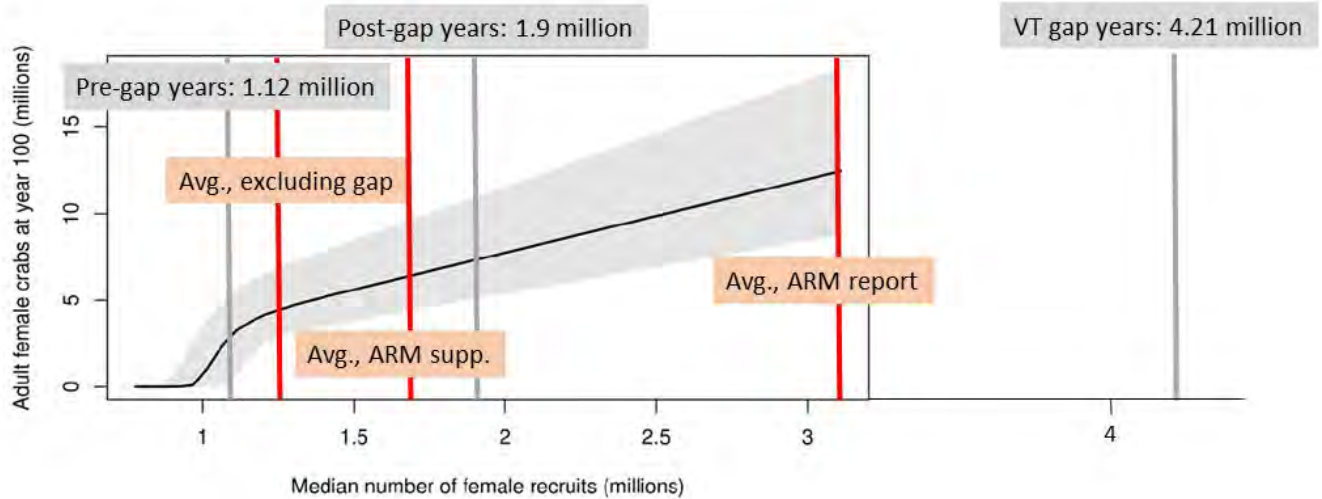
Figure 7*. Annotated version of ARM Fig. 33, which (in its original form) illustrates the sensitivity of HSC simulation results to changes in average HSC recruitment rates. Annotations reflect the average female recruitment before, after and during the VT gap years (in gray), the average recruitment value used in the original 2021 ARM report (red, far right), the value used in the supplemental report produced after peer-review (red, middle) and the analogous estimate computed by excluding the VT gap years (red, left). Average recruitment estimated for the VT gap years (arithmetic mean of 4.21 million based on the latest CMSA results) falls well outside the range of estimates during years for which recruitment was an observable process (and well outside the range of the x-axis of the original figure). The ARM report ignored recruitment estimates from the pre-gap years, giving very high weight to the inflated estimates during the VT gap years. Based on the peer-review, which suggested that the pre-gap years should not be excluded from the estimation of average recruitment rates, the current proposed value (described in the ARM supplement) is much lower than the value used in the ARM report (1.67 million vs. 3.1 million). However, the new proposed value continues to include unverifiable estimates from the VT gap years. If the VT estimates were excluded out of precaution, the average annual HSC recruitment would drop to 1.26 million, perilously close to the sustainability threshold identified in this figure (i.e., ARM Fig. 33).*

5. The proposed ARM framework lacks 'null model' benchmarks and independent performance validation

Null models are simplified representations of a system that lack many or all the explanatory mechanisms hypothesized to operate in the system. In statistics (e.g., linear regression analysis) the typical null model assumes all system variation is a result of unexplained variance in the form of random noise (often a single random error process). In other contexts, null models may include additional processes/mechanisms but omit a key focal mechanism, enabling researchers to test whether that focal mechanism contributes usefully to predictive performance. In the context of adaptive harvest management, a null model would at least omit

consideration of the impacts of harvest processes on system dynamics, which ultimately informs management decisions (Koons et al. 2022). By comparing complex models such as those used in the revised ARM with one or more null-model benchmark(s), researchers can determine whether the more complex models represent useful learned knowledge about a system (Koons et al. 2022). If a complex model fails to outperform a null model in terms of bias or precision (typically using independent validation data), the complex model is likely to be improperly specified or "overfitted" (whereby parameters are fitted to "noise" rather than true signal; Radosavljevic and Anderson 2014) and therefore not useful for prediction.

In the context of the HSC fishery in Delaware Bay, it would be informative to compare the performance of the HSC simulation model against a null model that omits all information about HSC harvest from the model fitting process; this would enable assessment of our current understanding of how estimated rates of harvest affect the HSC population. Given the poor fit of the HSC simulation model to training and validation data (see above), the HSC simulation is unlikely to outperform simpler null models. In fact, the CMSA model fails to outperform the simplest standard null model (single intercept term with sampling error) for at least one data source (the VT swept-area estimate of female multiparous abundance) despite its complexity (~20 parameters for the CMSA vs 1 parameter for describing expected abundance each year). If the HSC simulation model fails to outperform a model in which population dynamics are driven by noise instead of harvest, it should prompt managers to acknowledge that our current understanding of the effects of harvest on HSC populations remains insufficient for robust forecasting (Dietze 2017).

For the REKN component of the revised ARM, it would be informative to compare the performance of the REKN simulation model against a null model that omits any effect of female HSC abundance. It was recently demonstrated (Koons et al. 2022) that the ARM framework for guiding North American mallard harvest was unable to outperform a null model, and it would be instructive to pose a similar challenge to the REKN simulation model. Given that all the deterministic processes (fixed effects) included in the IPM model were very weak (i.e., the HSC effect on survival and fecundity; see above) or "non-significant", it is already apparent that random noise overwhelms most signal in the training data regarding how the HSC population

affects REKN population dynamics. Therefore, it is likely that information about the HSC/REKN relationship would explain little if any of the variation in independent validation data. Furthermore, the lack of a relationship between the HSC model (CMSA) and the number of spawning females observed on coastal beaches (see above) makes it even more unlikely that the current REKN population model would outperform a null model that excludes any effect of HSC abundance (since the HSC/REKN relationship is based on the consumption by REKNs of HSC eggs deposited by spawning females).

In summary, null model benchmarks should be incorporated into the ARM framework to ensure that effective learning is occurring and that managers acknowledge uncertainty about how their decisions affect the populations they are charged with managing (Koons et al. 2022). If one or both simulation models that form the core of the revised ARM framework fail to outperform null models, it would strongly suggest that the ARM framework's current level of understanding about how management decisions are likely to affect the HSC and REKN populations is insufficient for robust forecasting of population-level risk to either species from HSC harvest. Although the ARM process is designed to treat management actions as opportunities for learning – updating harvest recommendations as new data become available (Nichols et al. 2007) – the fact that one of these species is federally threatened (USFWS 2014) justifies a more precautionary approach for risk management.

6. Lack of transparency

The public still has no access to the data and code used for (1) estimating REKN population parameters via a Bayesian integrated population model (IPM), (2) simulating REKN and HSC population dynamics, and (3) running the optimization routines via approximate dynamic programming (ADP). The CMSA code and data were made available, which enabled me to re-analyze the HSC survey data and run informative scenario tests (see above). Without the data and code for other components of the ARM model, it is not possible to re-analyze the data, test key assumptions, or simulate population dynamics under different hypothetical scenarios. Given the substantial concerns generated by the data and code that has been made publicly available to date (discussed above), such further re-analysis, testing, and simulation is warranted. If granted access to the code and data, there are several important questions that could be

addressed more thoroughly, including but not limited to:

1) How would HSC abundance projections change – and how would harvest functions change – under the lower mean recruitment estimate produced by excluding anomalous estimates from the VT gap years?

2) What would happen to the REKN population projections if female HSC abundance were set to zero?

3) Does the REKN projection model outperform a null model that excludes any effect of HSC abundance?

4) In the REKN IPM, does the effect of HSC abundance disappear (or flip sign to become a negative relationship) under alternative plausible model specifications?

5) What proportion of variation in apparent survival in the REKN IPM model is explained by the HSC effect vs. random among-year variation?

6) Does an index of HSC spawning or HSC egg densities explain more variation in REKN survival and fecundity than the CMSA-derived estimate of HSC abundance?

CONCLUSION

In this report I have outlined six major concerns about the revised ARM. First, the modeled relationship between REKN vital rates and HSC abundance does not appear to be strong enough to induce an expected decline in the REKN population even under a catastrophic collapse of the HSC population. The apparent inability of the model to predict a major population response of REKNs to the depletion of the Delaware Bay HSC stock invalidates the premise of including a REKN population model within the ARM framework, which implicitly assumes that (1) HSC eggs are a critical resource for REKN populations and (2) HSC harvest could inhibit or slow the recovery of the REKN population, at least under some circumstances. The apparent inability of the ARM model to show a strong population-level effect of HSC harvest on REKN populations is inconsistent with the observed decline of the REKN population in recent decades, which many researchers have attributed to increased HSC harvest rates in the 1990s. Therefore, the REKN model included as part of the revised ARM does not appear to be a useful tool for assessing and managing risks to the REKN population from HSC harvest – or for promoting recovery of the REKN population.

In addition, I have identified several concerns about the HSC data analysis and simulation models. First, the HSC model in the revised ARM does not appropriately address key sources of uncertainty – particularly with respect to HSC fecundity (the source of potential harvest resilience). When these sources of uncertainty are addressed, the outlook for the HSC population is more uncertain than indicated in the ARM report. My analyses indicate that harvest at the maximum allowable levels could put the population in jeopardy (~11% risk) of decline below 3 million females – well below the minimum level previously recorded – within the next 50 years. In addition, the Catch Multiple Survey Analysis (CMSA), which is central to all aspects of the ARM, appears to exhibit poor fit to both training and independent data. I was unable to detect any correlation between the CMSA estimate of female HSC abundance and the estimated number of spawning females on coastal beaches in Delaware Bay. Finally, the estimate of HSC recruitment (which determines harvest resilience in the projection models) used in the revised ARM incorporates questionable (and highly inflated) estimates from a four-year period during which direct information on HSC recruitment was not available. Taken together, the above concerns strongly suggest the ARM model is not a valid tool for managing risk to the HSC population in Delaware Bay.

My final concerns are more general. First, I suggest that both the REKN and HSC models should be subjected to more rigorous evaluation, including tests for whether these models are able to outperform "null model" benchmarks that assume no useful learned knowledge about population dynamics and population response to harvest and harvest management. Ecological null models provide a useful benchmark for gauging the degree to which knowledge is accrued through the adaptive management process, and a mechanism for keeping modelers and managers "honest" by acknowledging an incomplete or inadequate understanding of the systems they are charged with managing. My analysis demonstrates that the CMSA model fails to outperform the simplest statistical null model for at least one data source. Finally, I was not provided access with much of the data and code used to generate the models used in the revised ARM (except for the CMSA code and data). Given the concerns that are apparent based on analysis of the limited code and data made available to date, it seems prudent to, at a minimum, delay implementation of this framework until the public and outside experts have had adequate

time to scrutinize the statistical and simulation models that play such a central role in this proposed decision-making framework.

Despite the lack of transparency, I was able to run several informative re-analyses and scenario tests with the information provided in the ARM report and supplement, and with the CMSA code and data. Based on my analysis, there is sufficient evidence to conclude that the ARM framework is not useful for assessing the resilience of the HSC population to harvest pressures, nor for managing risk to the REKN population due to HSC harvest.

Referenced literature:

Abrams, P. A., & Ginzburg, L. R. (2000). The nature of predation: prey dependent, ratio dependent or neither?. Trends in Ecology & Evolution, 15(8), 337-341.

Araújo, M. B., & Rahbek, C. (2006). How does climate change affect biodiversity?. Science, 313(5792), 1396-1397.

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

Dietze, M. C. (2017). Ecological forecasting. Princeton University Press.

Goodman, D. (2002). Predictive Bayesian population viability analysis: a logic for listing criteria, delisting criteria, and recovery plans. Population viability analysis, 447, 454.

Karpanty, S. M., Fraser, J. D., Berkson, J., Niles, L. J., Dey, A., & Smith, E. P. (2006). Horseshoe crab eggs determine red knot distribution in Delaware Bay. The Journal of wildlife management, 70(6), 1704-1710.

Karpanty, S. M., Cohen, J., Fraser, J. D., & Berkson, J. (2011). Sufficiency of horseshoe crab eggs for red knots during spring migration stopover in Delaware Bay USA. The Journal of wildlife management, 75(5), 984-994.

Kéry, M., & Schaub, M. (2011). Bayesian population analysis using WinBUGS: a hierarchical perspective. Academic Press.

Koons, D. N., Riecke, T. V., Boomer, G. S., Sedinger, B. S., Sedinger, J. S., Williams, P. J., & Arnold, T. W. (2022). A niche for null models in adaptive resource management. Ecology and evolution, 12(1), e8541.

Link, W. A., & Nichols, J. D. (1994). On the importance of sampling variance to investigations of temporal variation in animal population size. Oikos, 539-544.

Lovich, J. E., Ennen, J. R., Yackulic, C. B., Meyer-Wilkins, K., Agha, M., Loughran, C., ... & Madrak, S. (2015). Not putting all their eggs in one basket: bet-hedging despite extraordinary annual reproductive output of desert tortoises. Biological Journal of the Linnean Society, 115(2), 399-410.

McGowan, C. P., Runge, M. C., & Larson, M. A. (2011). Incorporating parametric uncertainty into population viability analysis models. Biological Conservation, 144(5), 1400-1408.

Midwood, J. D., Cairns, N. A., Stoot, L. J., Cooke, S. J., & Blouin-Demers, G. (2015). Bycatch mortality can cause extirpation in four freshwater turtle species. Aquatic Conservation: Marine and Freshwater Ecosystems, 25(1), 71-80.

Nichols, J. D., Runge, M. C., Johnson, F. A., & Williams, B. K. (2007). Adaptive harvest management of North American waterfowl populations: a brief history and future prospects. Journal of Ornithology, 148(2), 343-349.

Niles, L. J., Bart, J., Sitters, H. P., Dey, A. D., Clark, K. E., Atkinson, P. W., ... & Veitch, C. R. (2009). Effects of horseshoe crab harvest in Delaware Bay on red knots: are harvest restrictions working?. BioScience, 59(2), 153-164.

Radosavljevic, A., & Anderson, R. P. (2014). Making better Maxent models of species distributions: complexity, overfitting and evaluation. Journal of biogeography, 41(4), 629-643.

Regan, H. M., Colyvan, M., & Burgman, M. A. (2002). A taxonomy and treatment of uncertainty for ecology and conservation biology. Ecological applications, 12(2), 618-628.

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

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2016). False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant.

Tucker, A. (2019). Stopover ecology and population dynamics of migratory shorebirds. Dissertation. Auburn University.

Tucker, A. M., McGowan, C. P., Lyons, J. E., DeRose-Wilson, A., & Clark, N. A. (2021). Species-specific demographic and behavioral responses to food availability during migratory stopover. Population Ecology, 64(1), 19-34.

USFWS 2014. Endangered and Threatened Wildlife and Plants; Threatened Species Status for the Rufa Red Knot. Federal Register / Vol. 79, No. 238

USFWS 2021. Endangered and Threatened Wildlife and Plants; Designation of Critical Habitat for Rufa Red Knot (*Calidris canutus rufa*). Federal Register / Vol. 86, No. 133

Williams, B. K. (2011). Resolving structural uncertainty in natural resources management using POMDP approaches. Ecological Modelling, 222(5), 1092-1102.

EXPERT REPORT

Romuald N. Lipcius, Ph.D.

29 September 2022

Contents

1 Scope of Work

I was asked by representatives of EARTHJUSTICE to evaluate the Atlantic States Marine Fisheries Commission's Report and Supplemental Report to the 2021 Revision to the Adaptive Resource Management (ARM) Framework dealing with horseshoe crab (*Limulus polyphemus*) fishery management and implications for red knot (*Calidris canutus*) conservation. The red knot (RK hereafter) has been listed as "threatened" under the Endangered Species Act, and relies on horseshoe crab eggs buried along beaches of Delaware Bay to feed as it migrates along North and South America. The conclusions in the ARM report relate to an amendment proposed through the Atlantic States Marine Fisheries Commission (ASMFC) that would likely allow female horseshoe crab (HSC hereafter) harvest in Delaware Bay for the first time since 2012 and thereby potentially reduce food provisions (HSC eggs) needed by migrating RK. My primary goal is to evaluate the evidence in favor of the amendment objectively and determine if the amendment is justified.

In forming my opinions, I reviewed and considered various data sources regarding the HSC fishery and RK conservation along the Mid-Atlantic coast, with emphasis on Delaware Bay. My opinions are also based on my extensive experience conducting research and providing technical advice on fishery management and conservation of various marine species (see Section 8). My compensation is not contingent upon the conclusions or outcome of my review.

2 Summary Opinion

Based on my analysis and my expertise in conservation, fisheries and fishery management, I conclude to a reasonable degree of scientific certainty that:

The proposed amendment that would allow harvest of female horseshoe crabs is not justified by the available scientific evidence, due to various *risk-prone* decisions and assumptions that underlie the Adaptive Resource Management framework and model. The proposed amendment thereby poses a significant risk both to the Horseshoe Crab population and Red Knot recovery.

3 Abbreviations and Definitions

ARM: Adaptive Resource Management framework
HSC: Horseshoe Crab (*Limulus polyphemus*)
RK: Red Knot (*Calidris canutus*)
VTS: Virginia Tech HSC survey
DES: Delaware HSC survey
NJS: New Jersey HSC survey
Risk-prone: Conservation or management actions based on overly optimistic assumptions about the status of a population. The assumptions may be about data sources, observations or data, and often involve ignoring information to the contrary of optimistic conclusions about population status. For endangered or threatened species, a risk-averse, rather than risk-prone, strategy based on the precautionary principle is critical for population recovery, population conservation, and sustainable resource management.

4 Opinions

The following specific opinions describe various lines of evidence indicating that the HSC population is not in a healthy state and has not fully recovered despite a prohibition on female harvest since 2012. The different lines of evidence are effectively "red flags" leading to the conclusion that the current and proposed management strategies are risk-prone, such that harvest restrictions should not be relaxed at present. To the contrary, further management actions or improvements to the current management plan are necessary to stimulate HSC recovery. Furthermore, due to the lack of *substantial* improvement of the HSC spawning stock (i.e. mature females), the existing HSC management strategy has not significantly enhanced food availability for the threatened RK and therefore its recovery. A shift to risk-averse management based on the precautionary principle is essential for HSC and RK recovery.

4.1 Low Newly Mature Female, Recruit and Spawning HSC Abundance

An expectation from the female harvest prohibition is a rebound in young mature females and recruitment of immature males and females into the HSC population. In 2019 and 2020, abundance of newly mature females was at an all-time low; recruitment of immature females and males was extremely low and unchanged since before the prohibition; and female abundance in the spawning survey dropped sharply in 2019. These are warning signs that the HSC population has not fully recovered and may even be declining. Thus, female harvest should not be raised.

4.2 Smaller Body Size of Mature Female HSC

An expectation of the female harvest prohibition is that female body size would increase, given constant recruitment, which is a typical response in fisheries worldwide when harvest pressure on older, larger females is reduced. On the contrary, mean size of mature female HSC was smallest in the last 3 years (2018 to 2020) and of newly mature females in the last 2 years of the time series from 2002 to 2020, despite the prohibition on female harvest since 2012. These data are inconsistent with the previous expectation and the premise that the female segment of the HSC population has rebounded.

4.3 Loss of Large Mature Female HSC and Lower Egg Production

Population egg production is a function of spawning stock (= mature females) biomass (i.e. weight). Hence, changes in size distribution of mature females will affect total egg production, particularly the loss of large HSC females which contribute disproportionately to total egg production. Consequently, using only HSC abundance to estimate reproductive output and egg production is ignoring main biological drivers of population egg production–size structure and biomass–of the HSC spawning stock. Size distribution of mature females has shifted to smaller females. Abundance of females larger than 300 mm prosomal width (i.e. females with the highest egg production) has dropped recently, particularly from 2018 to 2020. Recent low recruitment means that smaller mature females are not compensating for the loss of larger mature females. Consequently, total reproductive (egg) output has likely not improved, which hampers recovery of the HSC and RK populations.

4.4 HSC Sex Ratio

When HSC harvest has been restricted to males, the ratio of males to females should have decreased. In contrast, male:female sex ratios have actually increased from 1999 to 2019. This represents another warning sign that the current management strategy has not been effective, that population dynamics are not well understood, and that harvest of females should not be increased.

4.5 High Mature Female HSC Mortality

The combination of discard mortality and bait harvest mortality for females has increased substantially in recent years and is comparable to levels before the prohibition. Assuming that the prohibition has worked is therefore risk-prone. The collective bait harvest and discard mortality is not being controlled effectively and inhibits HSC recovery.

4.6 Reliance on HSC Density as the Indicator of HSC Population Status

Female density (catch per unit area) is a primary variable used in HSC surveys and the ARM framework model. Reliance solely on HSC density or abundance ignores other variables that commonly produce warning signs about the status of a stock, such as female size, female size-frequency distribution, spawning stock biomass and female:male sex ratio. These variables are often more sensitive indicators of problems in a population, meaning that they can detect problems more effectively than abundance estimates. Hence, the current management strategy is risk-prone by ignoring these more sensitive indicators.

4.7 Low HSC Egg Density

Recent data indicate that HSC egg densities in HSC spawning habitats and RK feeding grounds remain an order of magnitude below densities when RK and HSC were relatively abundant. The ARM process has decided to ignore patterns in HSC egg density because of methodological "uncertainty" in the data. Under conditions where a population is not in danger, this may be acceptable, but absolutely not when it represents a potential warning sign about a population in danger, such as the RK. Thus, lack of use of HSC egg density data, as a proxy for RK food availability, amounts to a failure to incorporate all available scientific information into the analysis to guide management decisions in a risk-averse manner.

4.8 Lack of Correlation of HSC Surveys

Data from the DES and NJS of HSC in Delaware Bay are assumed to be correlated with the VTS and used to fill in survey gaps in the VTS. Survey data when all three surveys were conducted are not correlated, and data from the DES and NJS were relatively higher than that from VTS. These results lead to an overestimation of HSC abundance during VTS gap years, which is indicative of a risk-prone assumption.

4.9 Degraded HSC Spawning Habitat and RK Feeding Grounds

Spawning habitat (e.g. beaches) for HSC and feeding grounds for RK have been lost throughout the stopover range of RK in the Mid-Atlantic. Loss of habitat is an additional stress that demands risk-averse management of mortality sources (e.g. fishing) which management can control. There may be variables that are beyond ASMFC's control, but that means they should be more precautionary

with variables they can control, and it's certainly not a valid basis for ignoring warning signs like reduced HSC egg density and abundance.

5   Evidence for Opinions

The VTS is based on robust experimental design principles, and is the only spatially widespread survey that includes the coastal zone along Delaware and New Jersey, as well as Delaware Bay. In addition, the VTS collects much more comprehensive demographic data, which enables more types of analysis. Thus, the VTS serves as a robust and independent measure of HSC population status. The remainder of the analysis therefore focuses on data from the VTS and other published information on horseshoe crabs and the red knot. All analyses were conducted using the statistical software package R, version 4.1.2 (2021).

5.1   Low Newly Mature Female, Recruit and Spawning HSC Abundance

An expectation from the female harvest prohibition is a rebound in young mature females and recruitment of immature males and females into the HSC population. In 2019 and 2020, abundance of newly mature females was at an all-time low; recruitment of immature females and males was extremely low and unchanged since before the prohibition; and female abundance in the spawning survey dropped sharply in 2019. These are warning signs that the HSC population has not fully recovered and that female harvest should not be raised.

Data from the VTS on abundance of newly mature female HSC in 2019 and 2020 were at the lowest levels in the time series since 2002, indicating low influx of young mature females into the spawning stock (Figure 1). Similarly, abundance of immature female and male HSC, representing future recruitment to the adult segment and spawning stock of the population, were at extremely low levels and unchanged from those before 2013 (Figure 1). Moreover, female abundance in the Delaware Bay Horseshoe Crab Spawning Survey dropped sharply in 2019 (Figure 2), despite the prohibition of female harvest since 2012.

5.2   Smaller Body Size of Mature Female HSC

An expectation of the female harvest prohibition is that female body size would increase, given constant recruitment, which is a typical response in fisheries worldwide when harvest pressure on older, larger females is reduced (Beverton and Holt, 1956; Gedamke and Hoenig, 2006). On the contrary, mean size of mature female HSC was smallest in the last 3 years (2018 to 2020) and of newly mature females in the last 2 years of the time series from 2002 to 2020, despite the prohibition on female harvest since 2012. These data are inconsistent with the previous expectation and the premise that the female segment of the HSC population has rebounded.

VTS data were examined in two ways (mean and mode of size-frequency histograms) to evaluate this expectation. First, the time series of mean size in the VTS (Figure 3) indicated that mean sizes of mature female HSC and of newly mature females from 2016 to 2020 were the smallest in the time series from 2002 to 2020, despite the prohibition of female harvest since 2012.

Given that the mean of a sample can be influenced by outliers, the size data were also examined using a non-parametric statistic, the mode. The median could not be calculated because the raw data were unavailable for this analysis. The mode for each year was visually estimated from the size-frequency histograms of mature females (Appendix Figures 10 and 11). As with the mean, modal sizes of mature females from 2018 to 2020 were the lowest in the time series (Figure 4). In contrast, modal sizes of mature males were relatively unchanged (Figure 4).
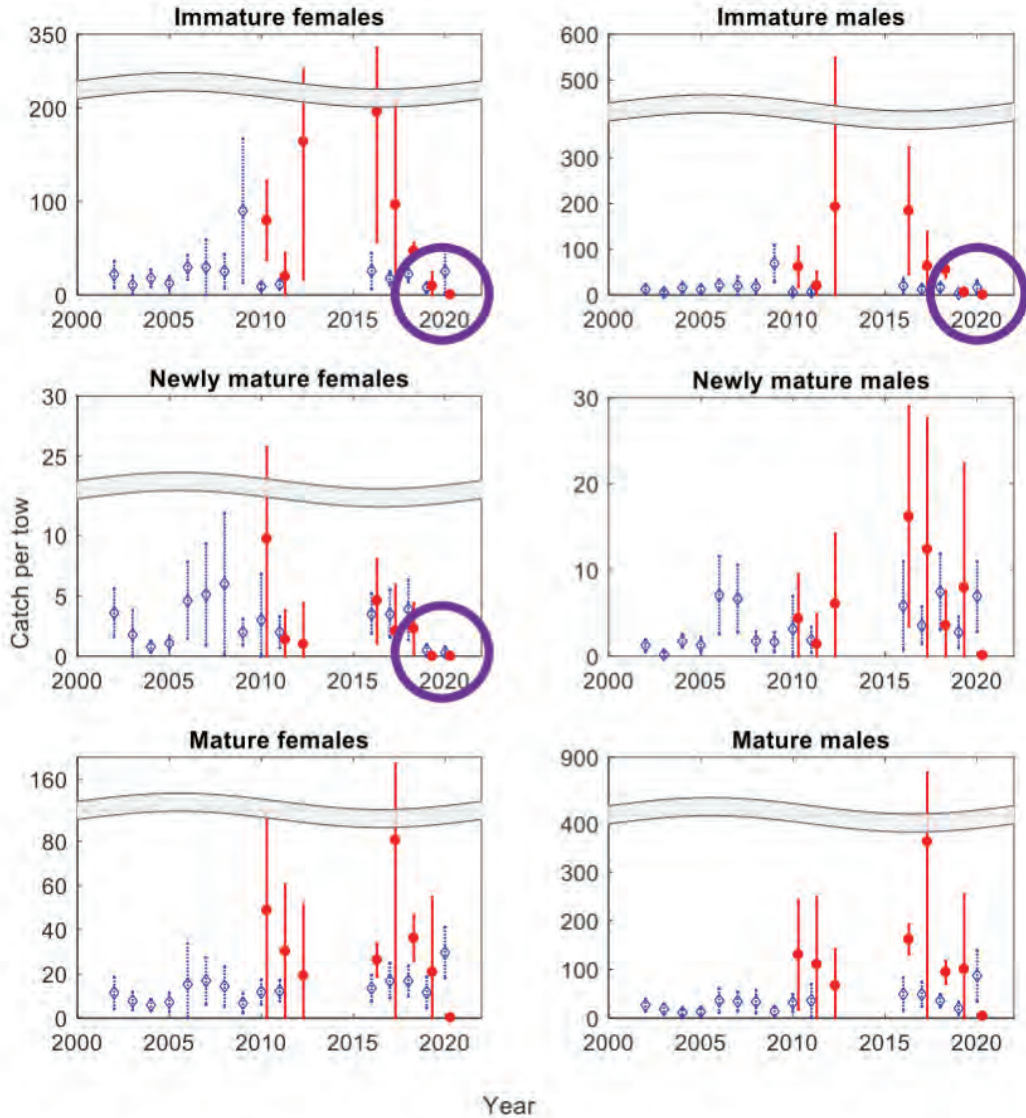
6

Figure 1: Densities of HSC males and females from Figure 3 of the VTS report (Hallerman and Jiao, 2021). Purple circles have been added to highlight the warning signs that the HSC population has not fully recovered.

Mean body size of spawning females could decrease over time if there was high recruitment of smaller, newly mature females shifting down the average size. However, the opposite (weak recruitment) appears to be the case, as described in section 5.1.

5.3   Loss of Large Mature Female HSC and Lower Egg Production

Population egg production is a function of spawning stock (= mature females) biomass (i.e. weight). Hence, changes in size distribution of mature females will affect total egg production, particularly large HSC females which contribute disproportionately to total egg production. Consequently, using only HSC abundance to estimate reproductive output and egg production is ignoring the main biological drivers of population egg production–size structure and biomass–of the HSC spawning stock. Size distribution of mature females has shifted to smaller females. Abundance of females larger than 300 mm prosomal width (i.e. females with the highest egg production) has dropped recently,
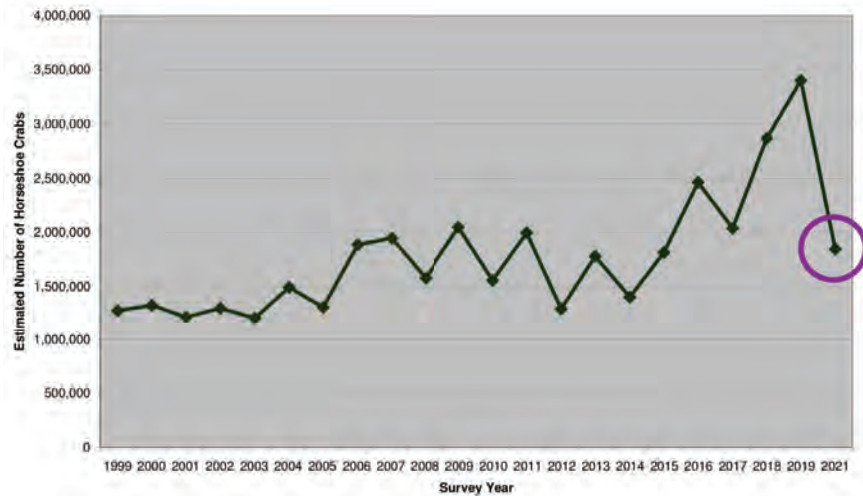
Figure 2: Spawning horseshoe crab survey data, highlighting low abundance of spawning horseshoe crabs in 2021 Swann and Hall (2019).
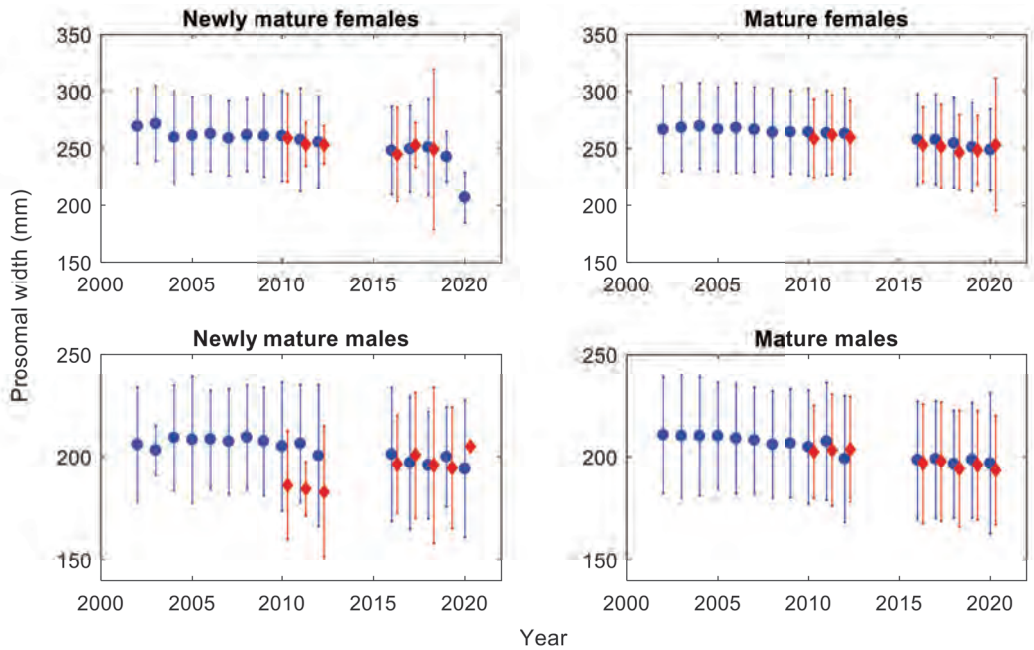

Figure 3: Mean sizes of newly mature and mature female and male horseshoe crabs over 2010 to 2020, with gap years from 2012 to 2015, from the VT survey in the coastal Delaware Bay area (Hallerman and Jiao, 2021).

particularly from 2018 to 2020. Recent low recruitment means that smaller mature females are not compensating for the loss of larger mature females. Consequently, total reproductive (egg) output has likely not improved, which hampers recovery of the HSC and RK populations.

For an individual HSC female, her egg production is directly proportional to individual weight, which is an exponential (not linear) function of prosomal width (Figure 5), as in other species of horseshoe crabs (Chatterji, 1995) and marine species in general (Barneche et al., 2018).

Changes in size distribution of mature females, particularly large HSC females which contribute disproportionately to total egg production due to the exponential increase in weight with size (Figure 6), will reduce population egg production. This was validated for an HSC population by
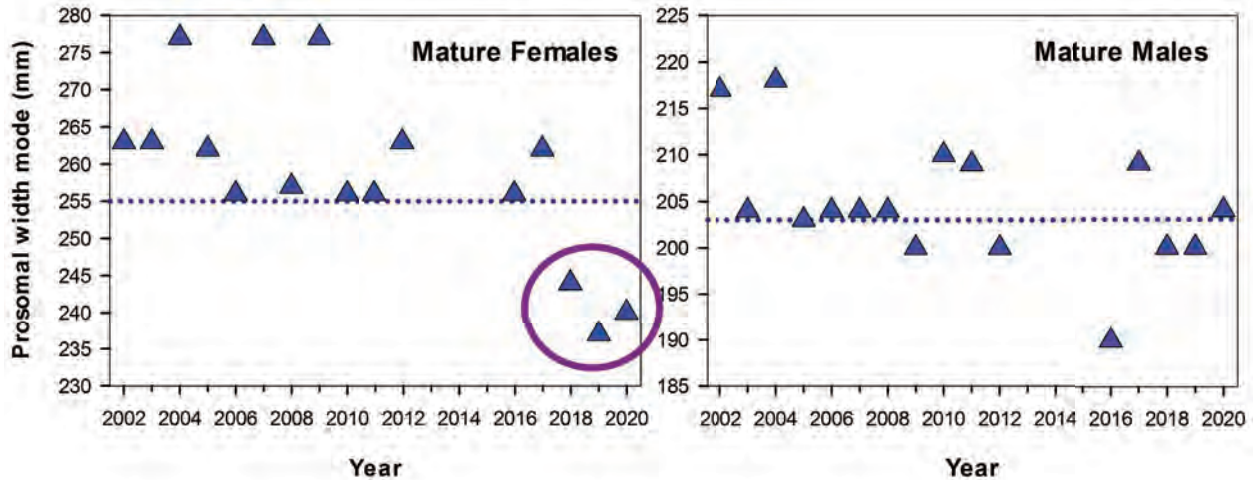
8

Figure 4: Size modes of mature female and male horseshoe crabs over 2002 to 2020 (gap years from 2013 to 2015) from the VTS in the coastal Delaware Bay area. Mode sizes were estimated from Figures 10 and 11.
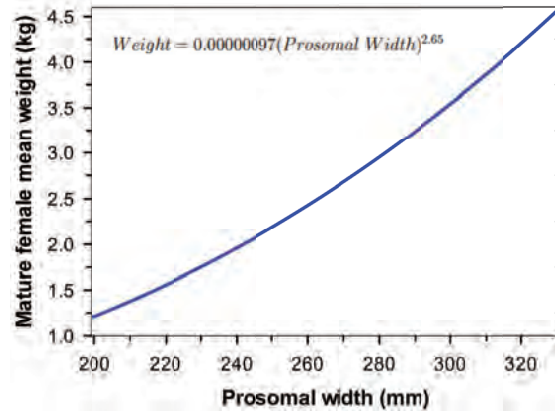

Figure 5: Exponential relationship between mature female HSC weight (kg) and prosomal width (mm) derived from Table 3 in Graham et al. (2009).

Leschen et al. (2006), who concluded that "larger females held a larger number of eggs (63,500) than smaller females (14,500) [and] laid a higher percentage of the eggs they contained. Thus they not only contain more eggs, but are more effective at laying them as well."

Using only HSC abundance to estimate reproductive output and egg production is ignoring the main biological drivers of population egg production–size structure and biomass (weight)–of the HSC spawning stock. Abundance is a reliable proxy of HSC egg production only if size structure of the spawning stock is unchanged over time, which is not the situation with the HSC spawning stock. Size distribution of mature females has shifted to smaller females (Figures 3 and 4), and recruitment does not account for the recent shift in size distribution because abundance of newly mature and immature females in the past few years has been well below average (Figure 1).

Abundance of females larger than 300 mm prosomal width (i.e. females with the highest egg production) has dropped recently, particularly from 2018 to 2020 (Appendix Figures 10 and 11), which has substantially reduced egg production. Note in Figures 10 and 11 that females larger than 300 mm prosomal width were apparent in 6 of 8 years from 2002 to 2009 (Figure 10), but only in 1 of 8 years from 2010 to 2020 (Figure 11). Moreover, the recent low recruitment means that
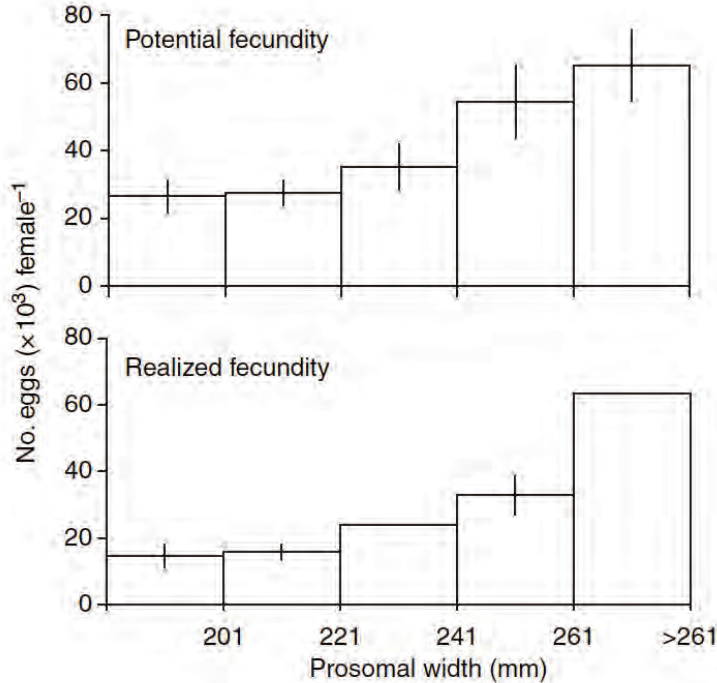
9

Figure 6: Positive relationship between HSC female fecundity and prosomal width (Leschen et al., 2006).

smaller mature females are not compensating for the loss of larger mature females. Consequently, total reproductive (egg) output has likely not improved, which hampers recovery of the HSC and RK populations.

5.4 HSC Sex Ratio

When HSC harvest has been restricted to males during the prohibition, the ratio of males to females should have decreased. In contrast, male:female sex ratios have actually increased from 1999 to 2019. This represents another warning sign that the current management strategy has not been effective, and that harvest of females should not be increased.

To assess HSC sex ratio over time, particularly since the prohibition on female harvest, I examined sex ratio data from the 2019 Delaware Bay Horseshoe Crab Spawning Survey, Table 5 (Figure 7). The time series shows an initial drop in the ratio of males to females during 2013, shortly after the prohibition on female harvest began. However, the ratio of males to females has increased since 2014 and even reached the highest ratios in the time series during 2018 and 2019.

5.5 High Mature Female HSC Mortality

The combination of discard mortality and bait harvest mortality for females has increased substantially in recent years and is comparable to levels before the prohibition. Assuming that the prohibition has worked is therefore risk-prone. The collective bait harvest and discard mortality is not being controlled effectively and inhibits HSC recovery.

Total mortality of females due to the bait fishery and its discards has increased substantially in recent years and is comparable to levels before the prohibition (Figure 8). Note that there is still a small amount of direct mortality due to the bait fishery (Figure 8), possibly due to inaccurate identification of female HSC by fishers. Thus, the prohibition on female harvest has not been
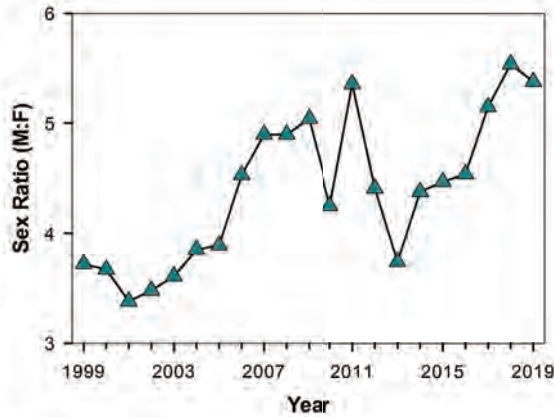
10

Figure 7: Sex ratio from the Delaware Bay Horseshoe Crab Spawning Survey Swann and Hall (2019).

effective in reducing female HSC mortality, and any further increase in female harvest is risk-prone and a danger to the HSC population and RK recovery.
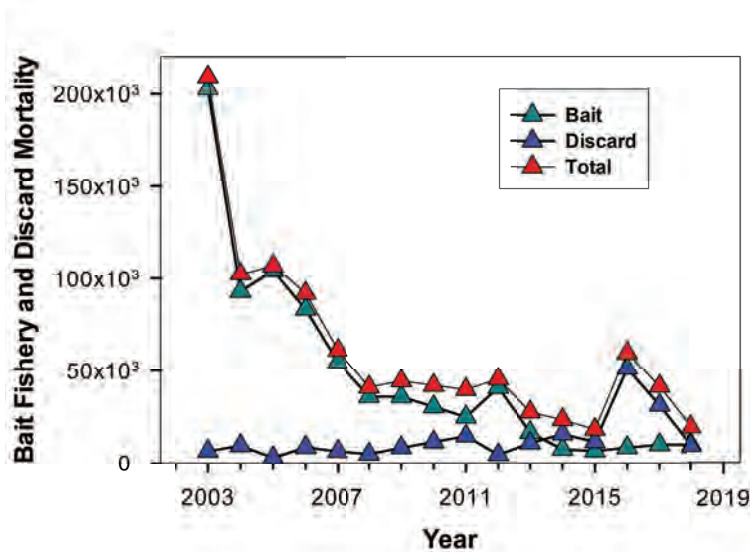

Figure 8: HSC mortality due to the bait fishery and discards (Adaptive Resource Management Subcommittee, 2022).

5.6 Reliance on HSC Density as the Indicator of HSC Population Status

Female density (catch per unit area) is a primary variable used in HSC surveys and the ARM framework model. Reliance solely on HSC density or abundance ignores other variables that commonly produce warning signs about the status of a stock, such as female size, female size-frequency distribution, spawning stock biomass and female:male sex ratio (Free et al., 2020; Punt et al., 2020). These variables are often more sensitive indicators of problems in a population, meaning that they can detect problems more effectively than abundance estimates alone. Hence, the current management strategy is risk-prone by ignoring these more sensitive indicators.

11

5.7   Low HSC Egg Density

Recent data indicate that HSC egg densities in HSC spawning habitats and RK feeding grounds remain an order of magnitude below densities when RK and HSC were relatively abundant. The ARM process has decided to ignore patterns in HSC egg density because of methodological "uncertainty" in the data. Under conditions where a population is not in danger, this may be acceptable, but absolutely not when it represents a potential warning sign about a population in danger, such as the RK. Thus, lack of use of HSC egg density data, as a proxy for RK food availability, amounts to a failure to incorporate all available scientific information into the analysis to guide management decisions in a risk-averse manner.

To assess changes in HSC egg density over time, I compared data for egg density before the peak of HSC harvest during 1985, 1986, 1988 and 1990 with data after the peak of HSC harvest from 1999 to 2021 (Smith et al., 2022). While the time series from 1999 to 2021 shows egg density increasing from an average of about 3,000 eggs per m^2 in 2000 to 9,000 eggs per m^2 in 2021 (Figure 6), egg density remains over an order of magnitude lower than that before the peak of HSC harvest during 1985 to 1990 (Figure 6).
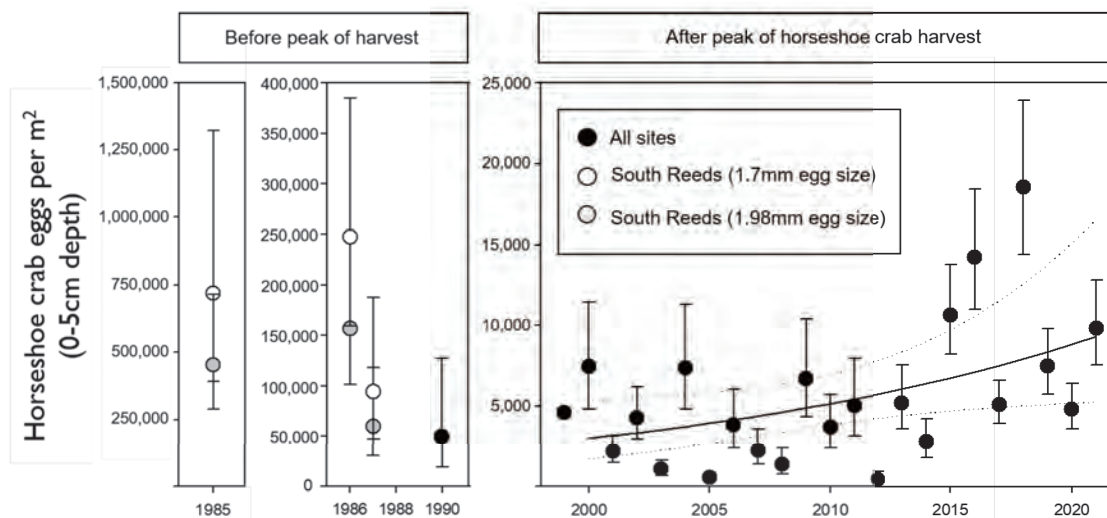

Figure 9: HSC egg density from spawning beaches, emphasizing the order of magnitude lower egg densities in recent years relative to historical levels in the spawning beaches. Note the different range of values in the left and right graphs. Figure from Smith et al. (2022).

5.8   Lack of Correlation of HSC Surveys

Data from the DES and NJS of HSC in Delaware Bay are assumed to be correlated with the VTS and used to fill in survey gaps in the VTS. Survey data when all three surveys were conducted are not correlated, and data from the DES and NJS were relatively higher than that from VTS. These results lead to an overestimation of HSC abundance during VTS gap years, which is indicative of a risk-prone assumption.

To evaluate the assumption of coherence between the three surveys, and justification for use of the DES and NJS in the four years when VTS data were unavailable, correlation between the three surveys was investigated. Data used in the analysis are those in Tables 1 and 2 from Adaptive Resource Management Subcommittee (2022) for indices VTS Multiparous Females, DES Adult and

NJS Ocean Trawl from 2003 to 2012, when indices were available for all three surveys prior to the 2012 prohibition.

Data for female and male HSC abundance from the three surveys were not correlated (Table 1), such that the use of data from two surveys (NJS and DES) to estimate data from the VTS survey during gap years when the VTS did not collect data is invalid. Furthermore, the NJS and DES produced data that were relatively higher than data from the VTS (positive intercepts in Table 1), indicating that the replacement data for the VTS using DES and NJS overestimate HSC abundance from the VTS.

Table 1: Correlation analysis for mature female HSC from VTS, NJS and DES.

Parameter	Estimate	Standard Error	t value	P
Females				
DES as a function of VTS: $r^2 = 0.01$				
Intercept	0.23	0.37	0.61	0.56
Slope	0.02	0.07	0.28	0.79
NJS as a function of VTS: $r^2 = 0.001$				
Intercept	1.96	0.67	2.91	0.02
Slope	-0.01	0.13	-0.07	0.95
Males				
DES as a function of VTS: $r^2 = 0.12$				
Intercept	0.03	0.23	0.12	0.91
Slope	0.02	0.02	1.03	0.34
NJS as a function of VTS: $r^2 = 0.03$				
Intercept	2.25	0.71	3.15	0.02
Slope	-0.03	0.06	-0.52	0.62

5.9 Degraded HSC Spawning Habitat and RK Feeding Grounds

Spawning habitat (e.g. beaches) for HSC and feeding grounds for RK have been lost throughout the stopover range of RK in the Mid-Atlantic. Loss of habitat is an additional stress that demands risk-averse management of mortality sources (e.g. fishing) which management can control. There may be variables that are beyond ASMFC's control, but that means they should be more precautionary with variables they can control, and it's certainly not a valid basis for ignoring warning signs like reduced HSC egg density.

A major threat to horseshoe crab population involves habitat degradation and loss, and is expected to worsen in the future due to sea level rise (Botton et al., 2022). Spawning habitat loss has been significant due to various factors such as shoreline management (e.g. bulkheading), coastal disturbances and sea-level rise (Smith et al., 2017, 2020). In some cases, whole beaches have been lost (Smith et al., 2017). Given that habitat loss is not under control by ASMFC, precautionary management demands consideration of such stressors to the population by control of fishery harvest to compensate for external stressors.
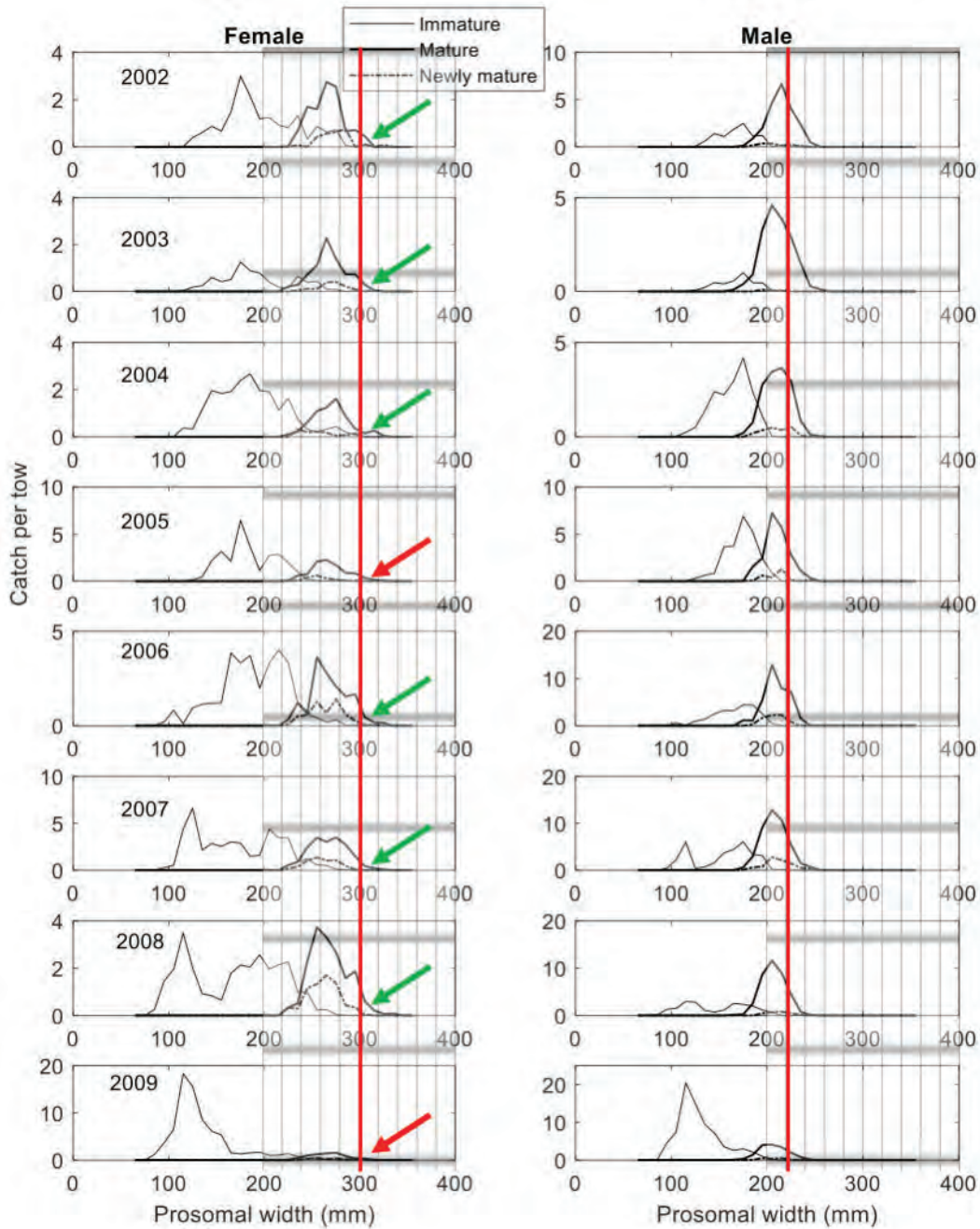
5.10   Appendix Figures

Figure 10: Size frequencies of mature female and male horseshoe crabs over 2002 to 2009 from the VT survey in the coastal Delaware Bay area (Hallerman and Jiao, 2021). Vertical red lines and grid cells were added for reference. Green arrows indicate years when mature females larger than 300 mm prosomal width were apparent, and red arrows when not.
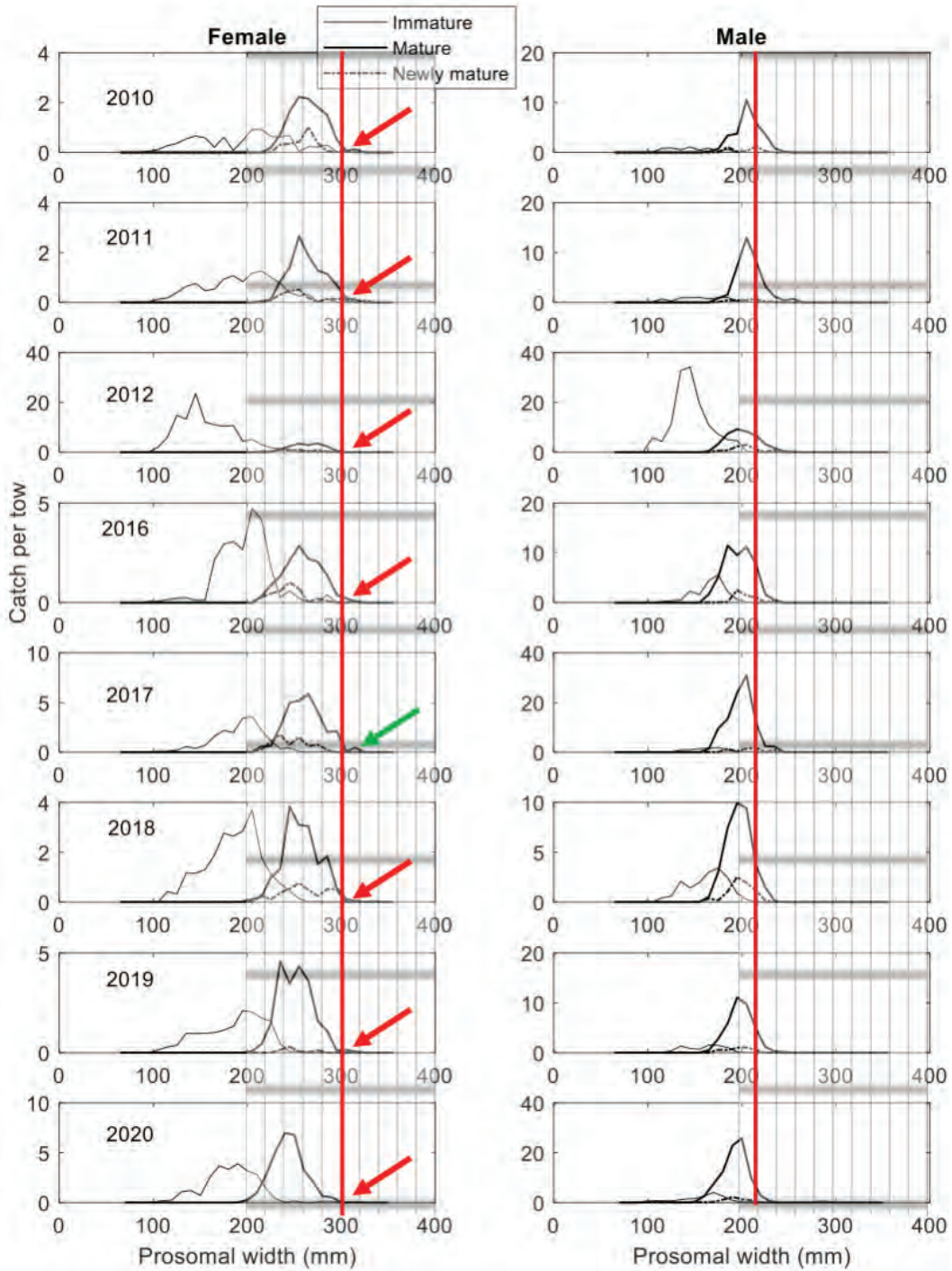
Figure 11: Size frequencies of mature female and male horseshoe crabs over 2010 to 2020, with gap years from 2013 to 2015, from the VT survey in the coastal Delaware Bay area (Hallerman and Jiao, 2021). Vertical red lines and grid cells were added for reference. Green arrows indicate years when mature females larger than 300 mm prosomal width were apparent, and red arrows when not.

15

6   Acknowledgements

I am extremely grateful to Dr. John Hoenig for his ideas and comments which greatly improved this report.

7   Literature Cited

Adaptive Resource Management Subcommittee (2022). 2021 revision to the adaptive resource management framework and peer review report. Technical report, Atlantic States Marine Fisheries Commission.

Barneche, D. R., Robertson, D. R., White, C. R., and Marshall, D. J. (2018). Fish reproductive-energy output increases disproportionately with body size. Science, 360(6389):642–645.

Beverton, R. and Holt, S. (1956). A review of methods for estimating mortality rate of exploited populations, with special reference to source of bias in catch sampling. Rapp. P.-V. Reun. Cons. Int. Explor. Mer., 140 (1): 67, 140:67–83.

Botton, M. L., Loveland, R. E., Munroe, D., Bushek, D., and Cooper, J. F. (2022). Identifying the major threats to American horseshoe crab populations, with emphasis on Delaware Bay. International Horseshoe Crab Conservation and Research Efforts: 2007-2020: Conservation of Horseshoe Crabs Species Globally, page 315.

Chatterji, A. (1995). Fecundity of the Indian horseshoe crab, Tachypleus gigas (Muller) from Balramgari (Orissa). Pakistan Journal of Marine Sciences, 4(2):127–131.

Free, C. M., Jensen, O. P., Anderson, S. C., Gutierrez, N. L., Kleisner, K. M., Longo, C., Minto, C., Osio, G. C., and Walsh, J. C. (2020). Blood from a stone: performance of catch-only methods in estimating stock biomass status. Fisheries Research, 223:105452.

Gedamke, T. and Hoenig, J. M. (2006). Estimating mortality from mean length data in nonequilibrium situations, with application to the assessment of goosefish. Transactions of the American Fisheries Society, 135(2):476–487.

Graham, L. J., Botton, M. L., Hata, D., Loveland, R. E., and Murphy, B. R. (2009). Prosomal-width-to-weight relationships in American horseshoe crabs (Limulus polyphemus): examining conversion factors used to estimate landings. Fishery Bulletin, 107(2):235–244.

Hallerman, E. and Jiao, Y. (2021). Results of the 2020 horseshoe crab trawl survey: Supplemental report to the 2021 revision to the adaptive resource management framework. Technical report, Virginia Polytechnic Institute and State University.

Leschen, A. S., Grady, S. P., and Valiela, I. (2006). Fecundity and spawning of the Atlantic horseshoe crab, Limulus polyphemus, in Pleasant Bay, Cape Cod, Massachusetts, USA. Marine Ecology, 27(1):54–65.

Punt, A. E., Tuck, G. N., Day, J., Canales, C. M., Cope, J. M., de Moor, C. L., De Oliveira, J. A., Dickey-Collas, M., Elvarsson, B. ., Haltuch, M. A., et al. (2020). When are model-based stock assessments rejected for use in management and what happens then? Fisheries Research, 224:105465.

Smith, D., Brockmann, H., Beekey, M., King, T., Millard, M., and Zaldivar-Rae, J. (2017). Conservation status of the American horseshoe crab, (*Limulus polyphemus*): a regional assessment. Rev Fish Biol Fisheries, 27:135–175.

Smith, J., Niles, L., Hafner, S., Modjeski, A., and Dillingham, T. (2020). Beach restoration improves habitat quality for American horseshoe crabs and shorebirds in the Delaware Bay, USA. Marine Ecology Progress Series, 645:91–107.

Smith, J. A. M., Dey, A., Williams, K., Diehl, T., Feigin, S., and Niles, L. J. (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, in press.

Swann, B. L. and Hall, W. (2019). The 2019 Delaware Bay horseshoe crab spawning survey. Technical report, `https://www.delawarebayhscsurvey.org/survey`.

8   Qualifications & Credentials

The qualifications, experience and scientific recognition that allow me to provide an informed, expert opinion on this matter are described below. My academic and professional credentials include: Professor (2000-present), Associate Professor (1993-2000), and Assistant Professor (1986-1993) of Marine Science, Virginia Institute of Marine Science, William & Mary, Department of Fisheries Science; Senior Postdoctoral Fellow, Smithsonian Institution (1997-1999); Postdoctoral Fellow, U.S. National Research Council (1985-1986); Adjunct Professor, Anne Arundel Community College (1984-1985); Postdoctoral Fellow, Smithsonian Institution (1984-1985); and Assistant Professor, Florida A & M University (1981-1984; while Ph.D. student at FSU). I received my Ph.D. from Florida State University in 1984 (major: Biological Science; minor: Statistics).

My scientific expertise and research specialties include Marine Conservation Ecology, Fisheries Management, Mathematical Biology, Ecological Statistics, and Ecology and Management of Crustaceans and Molluscs. Over the span of my career, I have 121 publications in peer-reviewed scientific journals, numerous technical reports, and 80 research grants totaling over $20 million from agencies including the National Science Foundation, National Oceanic and Atmospheric Administration, U.S. Army Corps of Engineers, National Undersea Research Program, Department of Defense, and various others.

I have 45 years of experience with eastern oyster, blue crab, Caribbean spiny lobster, queen conch, Nassau grouper and various marine bivalves; 36 years experience as the Commonwealth of Virginia's expert on blue crab ecology and fishery management; provision of formal opinions to Virginia Marine Resources Commission, Chesapeake Bay Commission, Chesapeake Bay Stock Assessment Committee, and Chesapeake Bay Program Fisheries Goal Implementation Team; 18 years experience as scientific advisor on oyster restoration to U.S. Army Corps of Engineers, NOAA Chesapeake Bay Office, and Chesapeake Bay Program Fisheries Goal Implementation Team; 33 years as Chief Scientist of the Blue Crab Winter Dredge Survey; Co-Principal Investigator of the Blue Crab Stock Assessment in Chesapeake Bay; and member of technical teams for Gulf of Mexico and Chesapeake Bay oyster and blue crab stock assessment, conservation and restoration.

Scientific honors, recognition and awards include: (i) Coastal America Partnership Award from the Executive Office of the President of the U.S., (ii) Kavli Fellowship from U.S. National Academy of Sciences, (iii) Aldo Leopold Leadership Fellow Award, (iv) Outstanding Faculty Award for Advisory Service, Virginia Institute of Marine Science, and (v) Outstanding Faculty Award for Research, Virginia Institute of Marine Science.