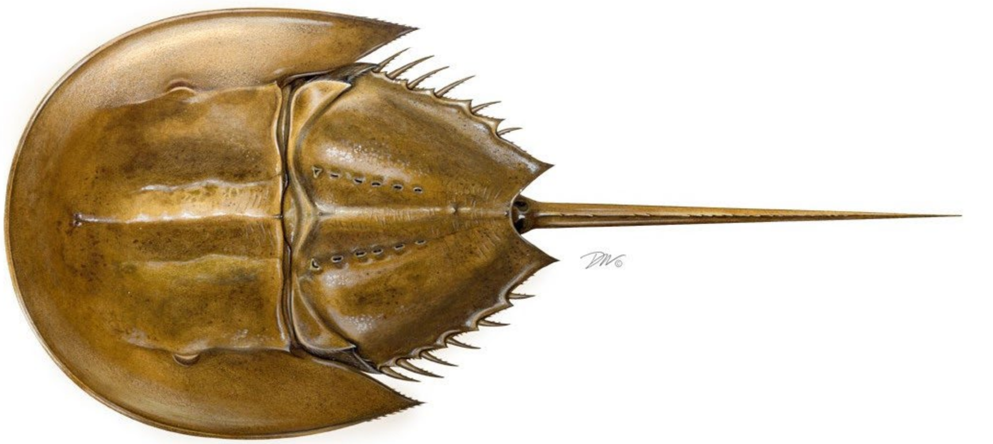
Atlantic States Marine Fisheries Commission

Technical Response to External Review of the
2022 ARM Framework Revision



Prepared by the
Adaptive Resource Management Subcommittee
John Sweka (Chair), US Fish and Wildlife Service
James Lyons (Vice Chair), USGS Patuxent Wildlife Research Center
Kristen Anstead, Atlantic States Marine Fisheries Commission
Linda Barry, New Jersey Division of Fish and Wildlife
Jason Boucher, National Marine Fisheries Service
Margaret Conroy, Delaware Division of Fish and Wildlife
Steve Doctor, Maryland Department of Natural Resources
Conor McGowan, USGS Alabama Cooperative Fish and Wildlife Research Unit
Clint Moore, USGS Georgia Cooperative Fish and Wildlife Research Unit
Bryan Nuse, University of Georgia
Caitlin Starks, Atlantic States Marine Fisheries Commission
Wendy Walsh, US Fish and Wildlife Service

In Collaboration with
Anna Tucker, USGS Iowa Cooperative Fish and Wildlife Research Unit

EXECUTIVE SUMMARY

The Adaptive Resource Management (ARM) Framework was developed in 2009 and implemented through Addendum VII in 2012 to set horseshoe crab harvest in the Delaware Bay at a level that does not limit the red knot stopover populations. In the decade since its implementation, more data on red knots and horseshoe crabs have been collected in the region, programming software advanced, and better population models were developed for the two species. Therefore, the ARM Framework was revised in 2022 by the ARM Subcommittee, a group that includes shorebird and horseshoe crab biologists and modelers. The ARM Revision was evaluated and endorsed by an independent panel of scientific experts through the Atlantic States Marine Fisheries Commission's (ASMFC) external peer review process.

While the ARM Revision represents significant advances in modeling and data use, the conversation around the revised ARM Framework quickly focused on the allowance of female horseshoe crab harvest when horseshoe crab population estimates are sufficiently high as to not limit red knot populations. The original ARM Framework had a technical flaw where it recommended 0 female horseshoe crab harvest when the adult female population was estimated to be less than 11.2 million, as it did from 2013-2022, or maximum female harvest (210,000 female horseshoe crabs) when the population was estimated to be greater than 11.2 million females, as it did in 2023. Rarely were the intermediate harvest levels selected by the model, as was shown through a simulation study. To correct this, the ARM Revision allowed a *gradual* increase of female harvest from 0-210,000 females as population estimates of female horseshoe crabs increased. The nuance of this change was lost in the discourse as stakeholders greatly opposed female harvest at any level, despite the original ARM Framework also recommending female harvest in recent years. In response to the concern over possible female harvest, Earthjustice, a non-profit public interest organization, hired experts to do their own technical review of the ARM Revision in 2022 and again in 2023 before the annual meeting of the Horseshoe Crab Management Board (Board) to set harvest specifications for the Delaware Bay region. During the October 2023 meeting, the Board tasked the ARM Subcommittee with responding to the 2023 technical review from Earthjustice.

The ARM Subcommittee seeks to always use the best scientific information available and welcomes scientific review and critique. As such, the Subcommittee has considered the comments provided by Earthjustice thoroughly. The following report outlines the ARM Subcommittee's responses to the six major criticisms listed by Dr. Kevin Shoemaker, a population ecologist at the University of Nevada, Reno, hired by Earthjustice as an external peer reviewer. Briefly, the ARM Subcommittee maintains that the red knot and horseshoe crab population models used in the ARM Framework currently represent the best use of the available data. Red knot survival rates and horseshoe crab population trends from the ARM Revision are consistent with other published values or data sources in the Delaware Bay region. This includes horseshoe crab egg density data, which were not provided to the ARM Subcommittee, but were subsequently published in the literature and show a similar trend to the horseshoe crab relative abundance indices. Additionally, the ARM Subcommittee responds to the comments in Dr. Shoemaker's report regarding the overparameterization or goodness of fit for the integrated population model for red knots and assert that this criticism misrepresents the work in the ARM Revision.

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

TABLE OF CONTENTS

INTRODUCTION

The Atlantic States Marine Fisheries Commission (ASMFC) has been managing the harvest of horseshoe crabs in the Delaware Bay region using an Adaptive Resource Management (ARM) Framework since 2012. The ARM Framework uses linked population dynamics models between horseshoe crabs and red knots to determine a harvest level of male and female horseshoe crabs such that the fishery has the opportunity to benefit from the harvest of horseshoe crabs and the population growth of red knots is not limited by that harvest. The original ARM Framework recommended an annual harvest of 500,000 male and 0 female horseshoe crabs from 2013 – 2022. These harvest recommendations have likely contributed to in an increase of both male and female horseshoe crab abundance over the last decade.

The original ARM Framework was theoretical in nature because the underlying population dynamics models for both species were based heavily on literature values of life history parameters and not specific to Delaware Bay. Since its inception, more Delaware Bay-specific data have been collected and the ARM Framework was revised in 2021 and adopted for management use in 2022 (ASMFC 2022). ASMFC 2022, henceforth referred to as "the ARM Revision," documents the many advantages of the revised ARM Framework including the use of more Delaware Bay-specific data, modern modeling software, and more advanced models for horseshoe crabs and red knots. Although shorebird advocates supported the original ARM Framework for managing horseshoe crab harvest, they have expressed strong disagreement with the use of the ARM Revision for making horseshoe crab harvest recommendations primarily because of the female horseshoe crab harvest strategy. This disagreement spurred intense review and scrutiny of the ARM Revision by Earthjustice and the outside experts they hired to critique the data and modeling. The first public comment by Earthjustice was submitted to ASMFC in September 2022 and contained critiques by Dr. Kevin Shoemaker (University of Nevada, Reno) and Dr. Romuald Lipcius (Virginia Institute of Marine Science). The second public comment was submitted in September 2023 and contained additional critique by Dr. Shoemaker. During the October 2023 meeting, the Board tasked the ARM Subcommittee, which includes red knot and horseshoe crab biologists and modelers, with responding to the 2023 critique by Dr. Shoemaker.

The ARM Revision (ASMFC 2022) did modify the female horseshoe crab harvest strategy from that of the original ARM Framework (ASMFC 2009a, 2012). The original ARM Framework had some technical flaws in the algorithm that optimized horseshoe crab harvest which resulted in an "all or nothing" harvest strategy for female horseshoe crabs. A simulation study showed that the original ARM Framework would recommend either 0 female harvest, or the maximum female harvest (210,000) if the female horseshoe crab population reached a threshold of 11.2 million individuals. Intermediate harvest levels would rarely, if ever, be recommended. The "all or nothing" harvest flaw in the original ARM Framework was observed when 0 female harvest was recommended from 2013 – 2022 and then in 2023, it recommended maximum harvest (210,000 female horseshoe crabs) because female horseshoe crab population estimates exceeded the threshold of 11.2 million. Conversely, the ARM Revision allows female harvest to gradually increase with increasing female horseshoe crab abundance. Despite detailed explanation for this difference between the two ARM Framework versions, shorebird advocates have strongly objected to the possibility of any female harvest, regardless of the population level of female horseshoe crabs and despite the fact that the original ARM Framework also allowed for female harvest.

With the publication of the ARM Revision and the discourse around the change in female harvest recommendations, Earthjustice solicited an external peer review of the technical work. The following represents the ARM Subcommittee's response to six major criticisms outlined by Dr. Shoemaker in his 2023 peer review. Each criticism is followed by a few bulleted summary points of the response and then a more detailed technical response to the criticism. While the ARM Subcommittee was not tasked with responding to the 2022 critiques, some responses to the major criticisms not included in the 2023 report have been provided in an appendix as supplemental information.

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

- High survival and long lifespans are common for red knots and other shorebirds of similar size and life histories.
- Survival rates used in the ARM are calculated from the tagging data for red knots in the Delaware Bay region and are comparable with other published survival values.
- The tagging data were critically analyzed by the ARM Subcommittee to represent the best available data and caveats to the survival estimates were provided in the ARM Revision. The analysis of the tagging data and its use in the modeling was commended by the peer review panel.

<u>Technical Response:</u> Dr. Shoemaker asserts that red knot annual survival probability is more likely closer to 0.8 than the 0.9 used in the revised ARM Framework, corresponding to an expected lifespan of about 5 years. There is not strong evidence for this lower annual survival probability for *rufa* red knot. In fact, previous studies of *rufa* red knot in Delaware Bay (McGowan et al. 2011) and Florida (Schwarzer et al. 2012) also estimated annual survival probability at approximately 0.9. In a separate published analysis, only using data collected by the state of Delaware, Tucker et al. (2022) estimated red knot annual survival probability at 0.89, and at 0.91 for ruddy turnstones, a species with similar body size and a similar annual life cycle. Additionally, observations of birds more than 5 years old are common in the mark-recapture data set (approximately 20% of birds), with a maximum of 17 years between physical recaptures. These observations are a conservative minimum estimate of lifespan. Further, it is worth noting that almost all vertebrate species with delayed maturation life cycles, like red knots, that do not recruit to the breeding population until their third year, exhibit high adult survival rates. This is especially true when annual reproductive output is low, as it is with red knots, which lay only four eggs in a single nest per year.

Outside of the Delaware Bay system, high survival and long lifespans are also reported for red knots and other shorebirds of similar size and annual cycle. For example, Piersma et al. (2016)

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

In his report, Dr. Shoemaker claims that the survival estimates in the ARM are biased by individual misidentification, or flag misreads. Before analyzing the data, the ARM Subcommittee conducted a thorough QA/QC, including filtering records to only lime and dark green flags that were first deployed by New Jersey or Delaware, removing records of 5 duplicate flags (n = 36), flags apparently resighted before they were deployed (n = 711), and flags that were never deployed (n = 1). Removal of these records represents only 0.35% of the total resightings. Members of the ARM Subcommittee have worked extensively on the issue of flag misreads, including conducting a thorough simulation study investigating the situations in which misreads might bias survival estimates and the implications of that bias (Tucker et al. 2019). The key points from that work are: 1) misreads disproportionately affect survival estimates from the first years of the study, causing apparent negative trends in survival over time, and 2) there is an important tradeoff to consider between potential bias due to misreads and loss of precision if data filtering is applied. In that paper, the authors suggest a data filtering step of removing all observations of flags that were only seen once in a year as a way to potentially mitigate misidentification errors. However, there are nuances to consider when determining whether this is necessary, because this data filtering will inevitably remove some number of valid observations, and the authors identify thresholds that depend on study length and error rate. For a 10-year study, removing single observations becomes beneficial if the error rate is >5%; below that rate the bias is minimal relative to the detrimental effects of removing valid observations. In the Delaware Bay mark-recapture dataset, the misread error rate is between 0.38% (712 impossible observations/187,587 total) and 4.5% (8,448 single observations). Additionally, the characteristic apparent negative trend in survival over time that would indicate bias due to misreads is not observed. To examine this further, the distribution of the number of resightings in a year for every flag (Figure 1) was plotted, with and without removing single observations. The shape of the resulting histogram indicates that removing these records results in fewer flags being seen once in a year than would be expected, i.e., that the data filtering removes a large number of valid records (> 3,000). The integrated population model uses the mark-recapture data to estimate survival as well as parameters related to stopover site use within each year. There were concerns that removing single observations would bias estimation of within-year parameters, and because the error was below the thresholds identified by Tucker et al. (2019) and the characteristic negative trend in survival was not observed, single observations were kept in the data set for the analysis.

The ARM Revision (ASMFC 2022) contains a thorough discussion of this topic on pages 63-64, in which several hypotheses for the disagreement in annual survival probability estimates from the older studies was described. Dr. Shoemaker points to lower estimates of survival from studies from the early 2000s, when red knot annual survival probability was estimated to be close to 0.8. It is likely that older estimates were negatively biased to some extent due to short study periods, low detection probably, and unmodeled temporary emigration from the system. It is also possible that during that time, when horseshoe crab populations were lower, red knot survival probability was truly lower. Alternatively, because permanent emigration from the system cannot be distinguished from mortality in older mark-recapture studies, a higher rate of permanent emigration (i.e., birds abandoning Delaware Bay for other spring stopover sites) would appear as lower survival probability. It is possible that there is a threshold of horseshoe crab abundance below which red knot survival probability might be expected to drop dramatically. If such a threshold exists, it was not observed over the time series included in the model (2005-2018).

It has also been proposed that southern-wintering birds (with longer migrations) have lower annual survival probabilities than northern-wintering birds. Declines in the number of red knots overwintering in Argentina (Niles et al. 2009) suggest a decline in the southern-wintering subpopulation and therefore it is possible that in more recent years a greater proportion of the Delaware Bay stopover population are northern-wintering birds. As discussed in the report, this is a key area for future research.

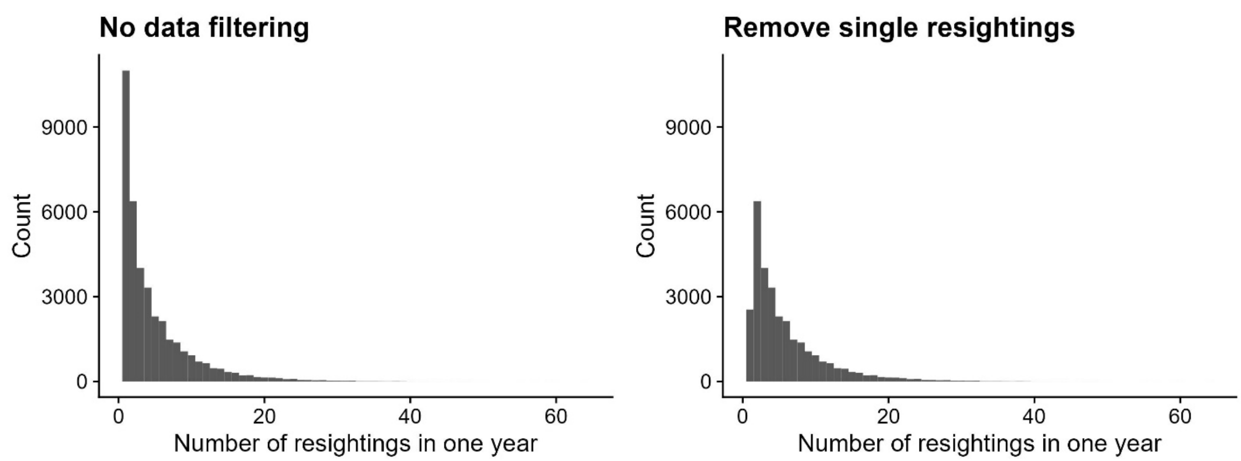


Figure 1. Histogram of the number of resightings per year for all lime and dark green flags deployed by New Jersey or Delaware from 2005 – 2018. The left panel is plotted without any data filtering. In the right panel, all flags that were seen only once (and not physically captured) were removed.

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

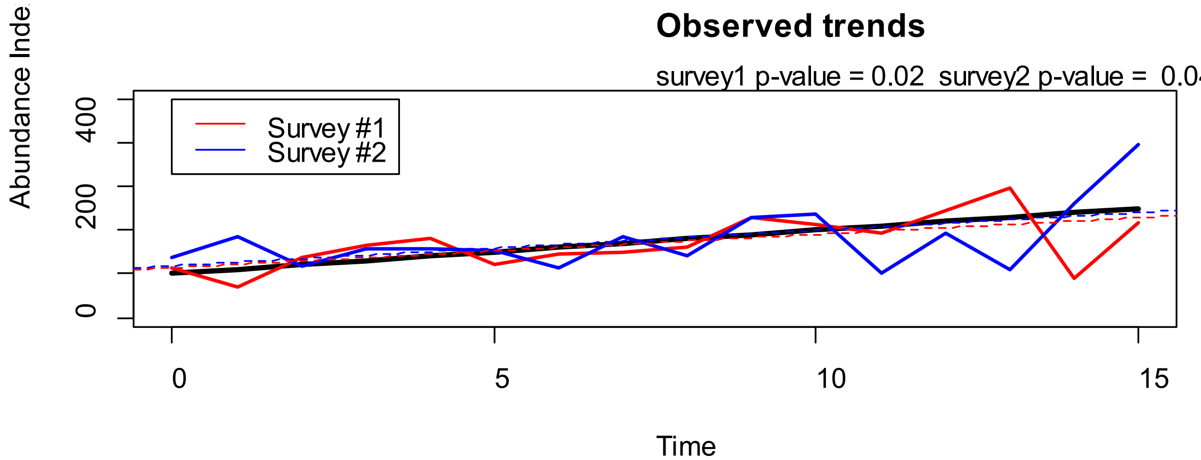
- The inclusion of trawl surveys as indices of horseshoe crab abundance may be imperfect but it is the best available science and its use has been approved by several independent peer reviews.
- Most of the criticisms and caveats relevant to trawl surveys would also apply to egg density and red knot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

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

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

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

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



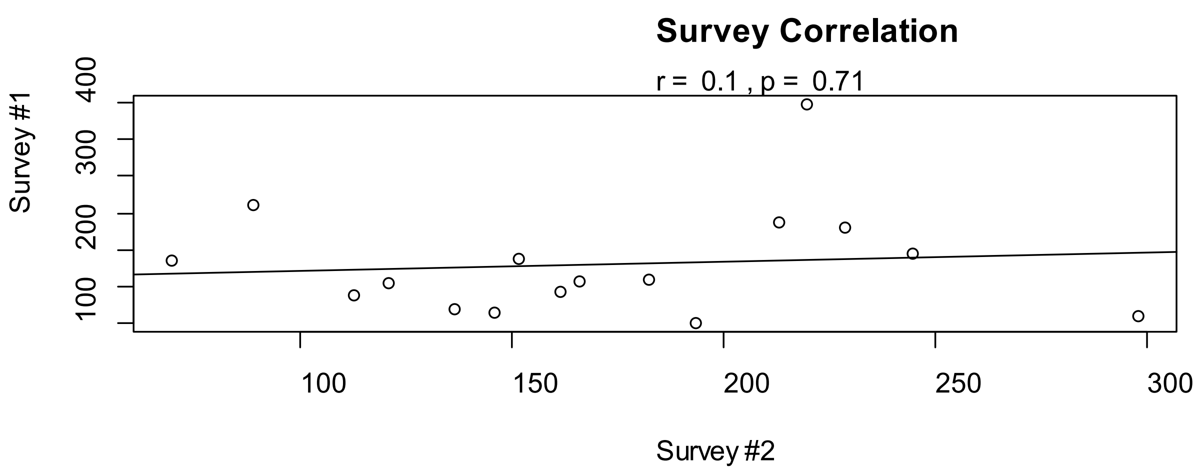


Figure 2. An example of simulated data to show that two surveys can both show an increasing trend but to be statistically uncorrelated with each other. The top graph shows the true change in hypothetical abundance (black line) and two randomly generated independent surveys of abundance each with a CV = 0.3. Dashed red and blue lines indicate linear regression lines for the two surveys. Both of these randomly generated surveys show statistically significantly increasing trends (p-values < 0.05), yet the correlation between the two is low (r = 0.10) and non-significant (p = 0.71).

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

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

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

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

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

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

Additionally, Dr. Shoemaker analyzed the egg density data from Smith et al. (2022) and accounted for differences in survey methodology through time. The results of his reanalysis showed no trend in egg density although Smith et al. (2022) showed a general increasing trend in recent years similar to that of horseshoe crab abundance from the CMSA (Figure 4). Dr. Shoemaker also conducted an analysis that shows the effect of egg density on red knot survival. However, this survival analysis is not documented in great detail and only includes data from the New Jersey side of the Delaware Bay. Thus, it is questionable whether this analysis is representative of the red knot population as a whole. If these analyses by Dr. Shoemaker are correct, it still begs the question of how to incorporate this into the ARM Framework. In Dr. Shoemaker's report, red knot survival is positively correlated with egg density but egg density has not changed over time; however, female horseshoe crab abundance has increased. Therefore, while egg density and female horseshoe crab abundance must ultimately be linked,

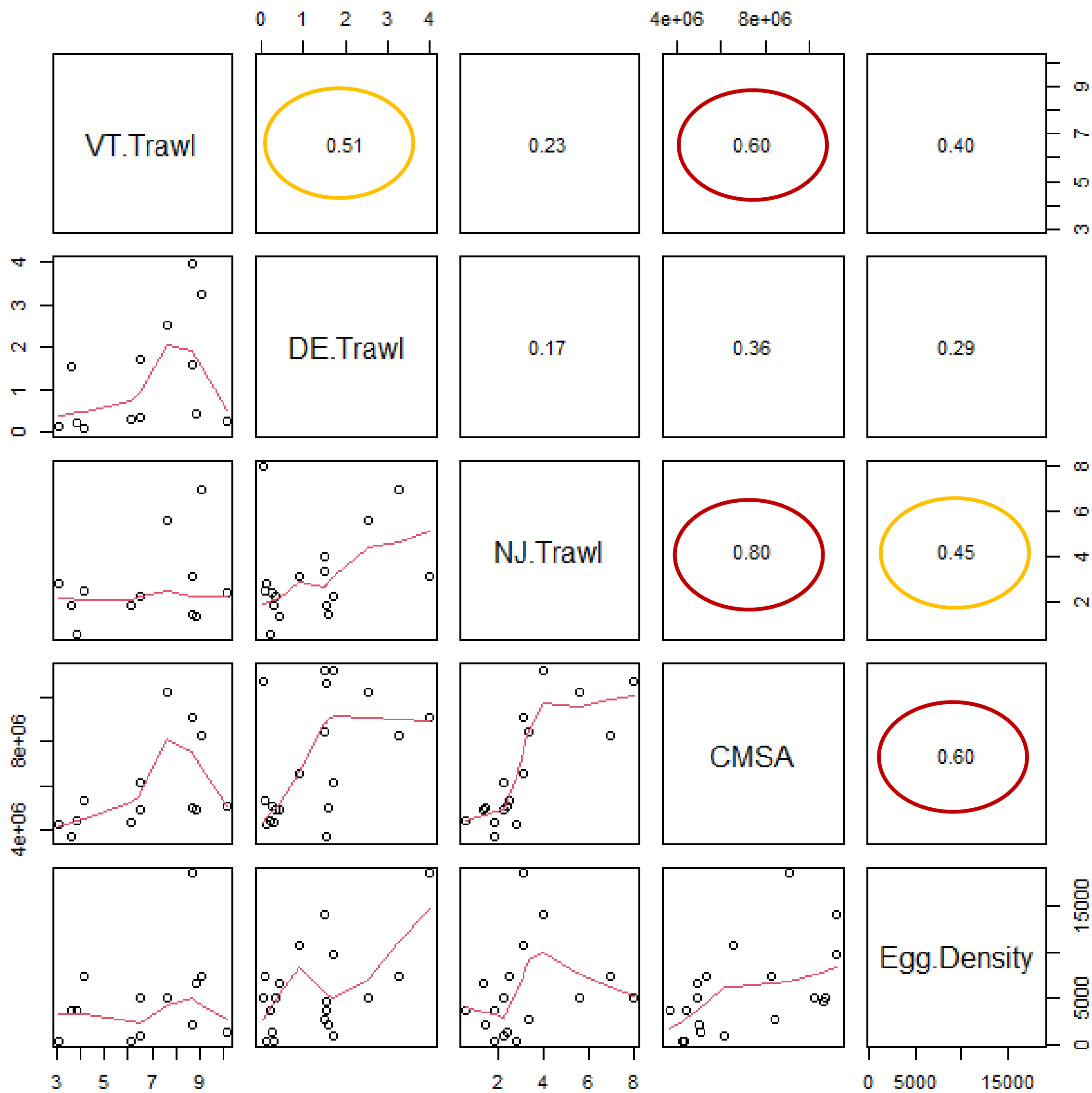


Figure 3. Scatterplot matrices (lower diagonals) and spearman correlation tests (upper diagonals) for female horseshoe crab abundance indices derived from the CMSA model (used as an estimate of horseshoe crab abundance in the ARM Framework), three trawl-based surveys conducted in the Delaware Bay area from 2003 to 2021 (female indices), and New Jersey surface egg densities from Smith et al. (2022). All time series are positively correlated and those correlation coefficients circled in red are significantly correlated at the P<0.05 level. Correlation coefficients circled in yellow are significant at the P<0.10 level.

this relationship is not evident in the data. The lack of an empirical relationship ultimately complicates any effort to quantify a model linking horseshoe crab abundance to red knot survival through egg density. Dr. Shoemaker falls short of proposing a way to do this. Regardless, for the time series of the CMSA model, egg density is positively correlated with the other time series of horseshoe crab abundance used. Because egg density data are not readily available to the ARM Subcommittee (either for the model development in 2021 or possibly on an annual basis that would be required for their inclusion), the data only cover New Jersey beaches, and their use and sampling design have been questioned over the years, the trawl surveys remain the best available data for horseshoe crab abundance in the ARM Framework.

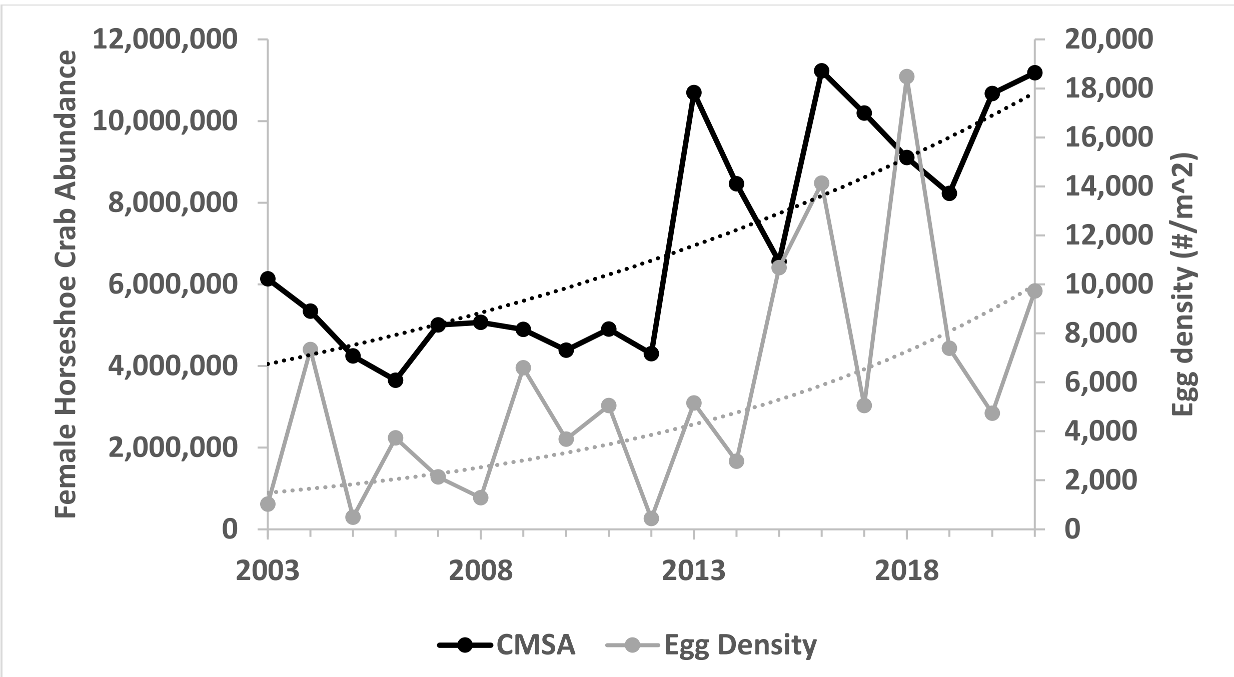


Figure 4. Increasing trends of female horseshoe crab abundance from the catch survey model and egg density time series, 2003-2021. Egg density data were digitized from Figure 2 in Smith et al. (2022).

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

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

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

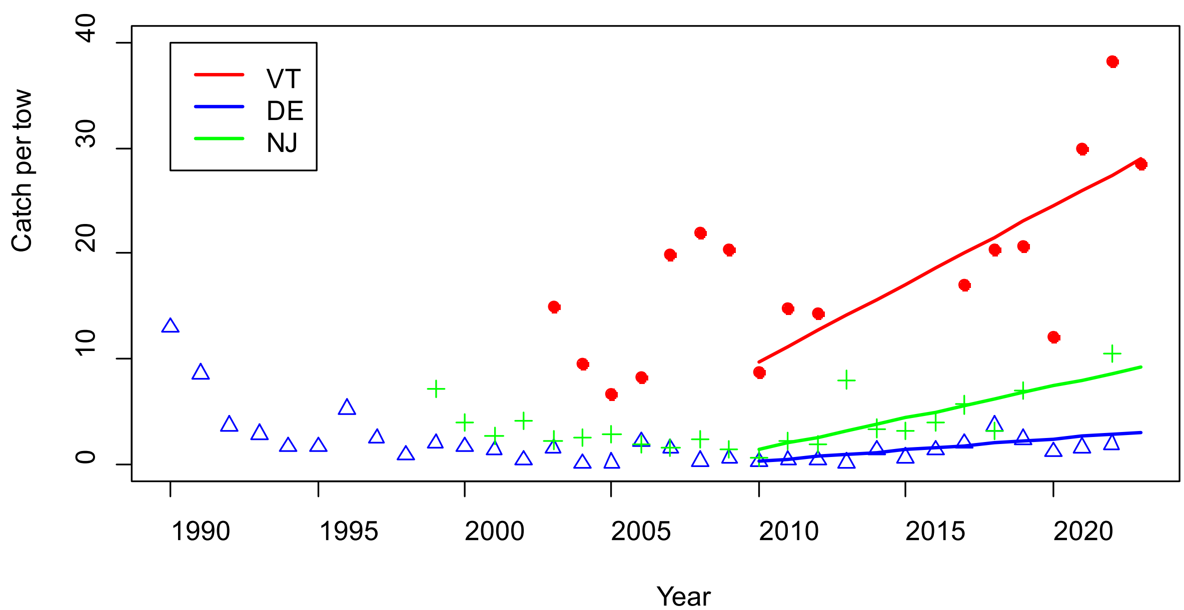


Figure 5. Time series of catch-per-tow of female horseshoe crabs from the three trawl surveys used to index abundance. Simple linear regression models for each survey since 2010 all show statistically significant ($p < 0.05$) increasing trends in abundance.

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

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

Shorebird stakeholders' views of the trawl surveys have evolved through time. In a 2009 publication questioning if harvest restrictions have worked to increase horseshoe crab abundance in the Delaware Bay region, Niles et al. (2009) included a graph of the Delaware Trawl Survey showing a declining trend in catch-per-effort as evidence that horseshoe crab abundance has declined. In Earthjustice's September 2022 comments to the Board, they argue that the "…original decision to rely exclusively on the Virginia Tech Trawl Survey reflected explicit stakeholder input," and that other trawl surveys are "not purpose-designed" and "disfavored" by stakeholders. Finally, in 2023, according to Dr. Shoemaker, "trawl-based indices of horseshoe crab abundance are a noisy and unreliable indicator of annual fluctuations in the horseshoe crab population…". If the view that the trawl surveys only capture random noise is accepted, and thus the increasing trend in the surveys since 2010 cannot be trusted, one should also question if the horseshoe crab population actually ever declined.

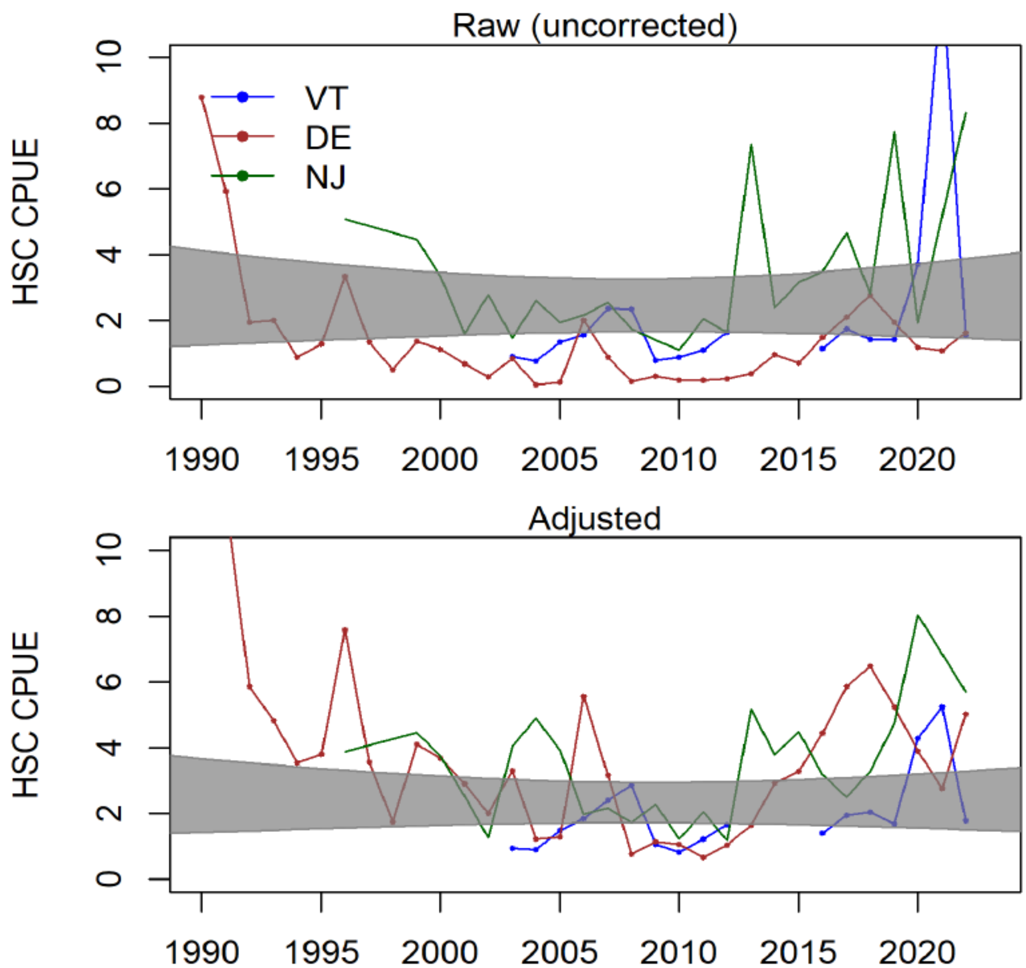


Figure 6. This graphic is taken from the 2023 Dr. Shoemaker report (Figure 12). The intent was to show there is no significant trend in female horseshoe crab abundance through time for the combined trawl surveys using a linear regression model over the entire time series (1990 – 2022). The reason for the lack of a significant trend, either increasing or decreasing (gray shaded area) is because the time series exhibits a "U" shaped pattern – decreasing until around 2010, and increasing afterwards.

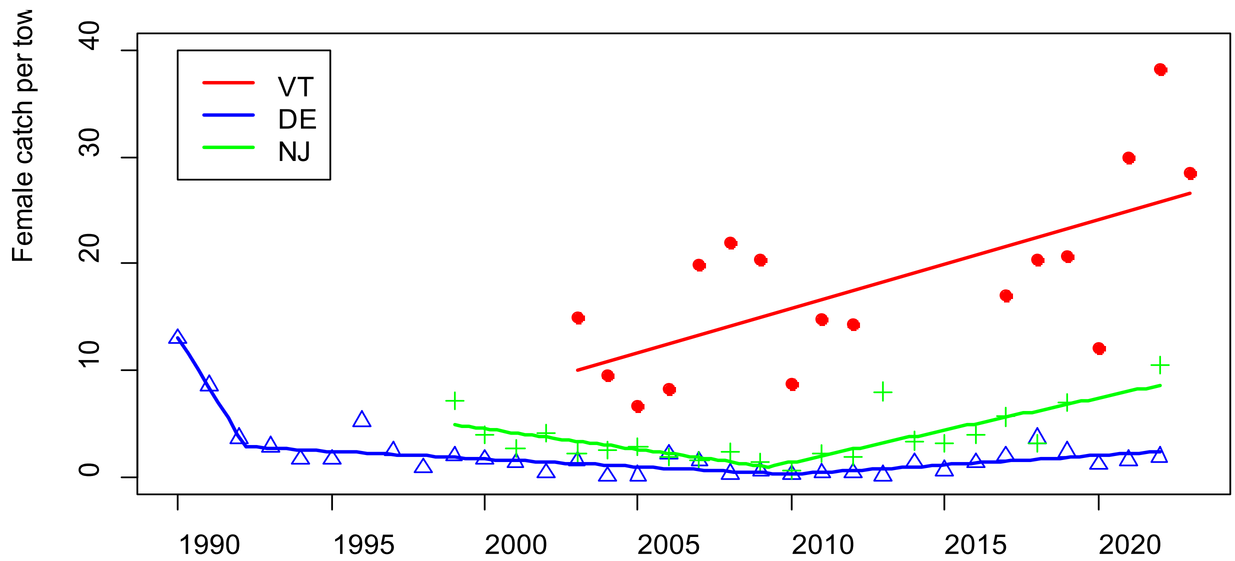


Figure 7. Trends in female horseshoe crab catch-per-tow from the three trawl surveys in the Delaware Bay region. Lines for the Delaware and New Jersey Trawls represent segmented regression models fit to those data. The results of the segmented regression analysis show that the slope of the trend for both the Delaware and New Jersey Trawls changed from being negative to being positive around 2010. The trend line for the Virginia Tech Trawl Survey is simply a linear model over the entire time series because the Virginia Tech Trawl Survey did not extend back in time as far as the other surveys.

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

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

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

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

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

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

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

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

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

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

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

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

CONCLUSIONS

Continuous scientific review and critique is welcome as that is how science advances. There will always be room for improvement in any modeling effort in the management of natural resources. This is part of the double-loop learning in an adaptive management effort whereby model design and management are periodically reevaluated (Fabricius and Cundill 2014; Williams and Brown 2018). In this specific case, however, advocacy is infused into the scientific debate. The 2022 ARM Revision represented some great advancements in the understanding of the population dynamics of horseshoe crabs and red knots, and their interactions during the double-loop of the adaptive management process. It is curious that these advancements have stirred so much controversy because the technical criticisms of the ARM Revision could have equally applied to the original ARM Framework. In fact, the original framework merited specific criticism because it relied on life history parameters informed by literature values taken from outside the Delaware Bay or based on expert opinion. The ARM Subcommittee questions if the true problem is not with the process or technical modeling, but rather with the final result and harvest recommendation.

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

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

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

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

SUPPLEMENTAL RESPONSES TO THE 2022 EXTERNAL PEER REVIEWS

The Management Board specifically tasked the ARM Subcommittee with responding to the 2023 critique by Dr. Shoemaker, and the responses above fulfill that task. However, the ARM Subcommittee felt it appropriate to address a few additional items presented in the 2022 public comment by Earthjustice that included critiques by Dr. Shoemaker and Dr. Romuald Lipcius, as well as the supplemental section to the 2023 critique by Dr. Shoemaker. These items are not in any particular order.

Criticism 7: The estimate of mean horseshoe crab recruitment and propagation of error within the horseshoe crab population dynamics model is inappropriate.

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

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

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

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

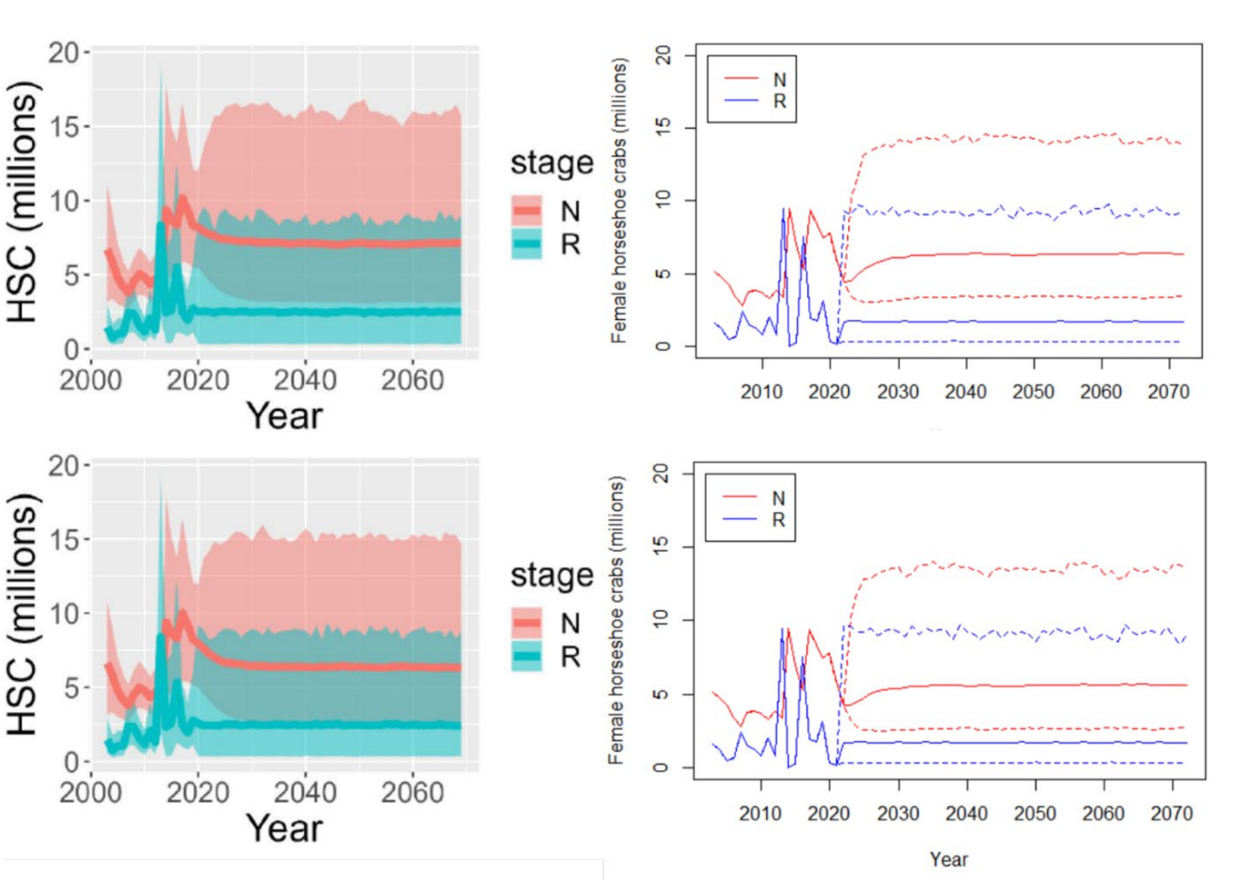


Figure 8. Comparison of the projection model of female horseshoe crabs by Dr. Shoemaker in Earthjustice's September 2022 comments (left graphs) with that from the revised ARM Framework (right graphs). The early part of the projected time series (2021 and prior) represent population estimates from the CMSA – both the version in the revised ARM Framework and Dr. Shoemaker's Bayesian version in Earthjustice's September 2022 comments.

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

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

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

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

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

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

Criticism 9: Demographic data indicate a declining horseshoe crab population.

- Declining individual size of horseshoe crabs began after harvest was greatly curtailed in the Delaware Bay region and is not indicative of overfishing.
- Assuming natural mortality has not changed, abundance of horseshoe crabs could not have increased if egg deposition and hatch had also not increased.
- Recent low estimates of female primiparous crabs do not necessarily represent recruitment failure. Male primiparous crabs did not decrease over the same time period.

<u>Technical Response:</u> In 2022, Dr. Lipcius argues demographic data are inconsistent with an increase in the horseshoe crab population such as the apparent decline in mean size of individual horseshoe crabs. It is true that mean size has decreased and the ARM Subcommittee agrees that in a typical finfish fishery a declining trend in mean size-at-age is indicative of overfishing (i.e., faster growing fish recruit to the fishing gear at younger ages and a fishery then selects against fast growing fish). However, horseshoe crabs are not finfish, they have a terminal molt, and stop growing after maturity is reached. One cannot apply the general rule-of-thumb that average size deceases with excessive exploitation to a species like horseshoe crabs, which stop growing once mature and are targeted by a commercial fishery at the mature stage. Fishing pressure for males, and especially females, greatly declined since the 1990s, yet it appears from the Virginia Tech Trawl Survey data on prosomal widths that the decrease in size occurred after 2008 (Wong et al. 2023), and after the fishery was curtailed in the mid-2000s. Alternative hypotheses for the reduction in size is density-dependent growth as the population rebuilt, or an ecosystem wide loss of productivity over the last 20 years resulting in fewer resources available for horseshoe crab growth. The ARM Subcommittee agrees that additional research is needed to explain the declining size of horseshoe crabs, but it is doubtful that it is tied to fishing mortality given how limited the harvest has been relative to the size of the population.

Dr. Lipcius makes an argument that with the decrease in mean size of mature female horseshoe crabs, individual fecundity would also decrease and total reproductive output has not increased. This hypothesis seems unlikely because horseshoe crab abundance would not have increased if natural mortality has not changed and there had there not been an increase in total egg deposition and hatch. Smith et al. (2022) shows a general increase in egg density in recent years, which also refutes this hypothesis.

Dr. Lipcius also argues that the recent low numbers of newly mature (primiparous) females in the Virginia Tech Trawl Survey indicate recent harvest is problematic. Intuitively, one would expect an increase in recruitment following the prohibition of female harvest. However, many factors influence year class strength of horseshoe crabs and there is a 9 to 10-year delay between when new crabs are spawned and when primiparous crabs are assessed. There could be density-dependent effects (nest disturbance by subsequently spawning females) at play, and inter-annual variation in survival over the 9 to 10-year period between the egg and primiparous stage could mask any differences in year class strength. Some very high years of newly mature males also occurred prior to the prohibition of female harvest (Wong et al. 2023). The observed variation in newly mature animals suggests year class strength is influenced by much more than female spawner abundance alone. Also, harvest pressure targets mature individuals and Virginia Tech Trawl Survey data shows a significant increase in mature individuals through time, especially in the last three years. There could not have been an increase in multiparous individuals without a preceding increase in primiparous individuals. Finally, we do not observe the same decline in primiparous males as observed in primiparous females. If harvest pressure caused a decline in female recruitment, as suggested by Dr. Lipcius, why would it not also cause the same decline in male recruitment? The recent years of low primiparous female abundance observations is something the ARM Subcommittee and Delaware Bay Ecosystem Technical Committee are discussing.

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

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

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

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

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

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

Table 1. The proportion of individual birds resighted at Mispillion Harbor only, at other sites only, or at both Mispillion and other sites in each year.

Year	Resighted in Mispillion Harbor only	Resighted at non-Mispillion sites only	Resighted at both Mispillion and other sites
2005	0.26	0.45	0.30
2006	0.28	0.40	0.32
2007	0.48	0.17	0.35
2008	0.48	0.30	0.23
2009	0.46	0.28	0.26
2010	0.12	0.69	0.20
2011	0.46	0.30	0.25
2012	0.30	0.46	0.24
2013	0.29	0.53	0.18
2014	0.36	0.43	0.20
2015	0.54	0.24	0.22
2016	0.25	0.62	0.14
2017	0.53	0.27	0.21
2018	0.48	0.29	0.23

REFERENCES

Anstead, K.A., J.A. Sweka, L. Barry, E.M. Hallerman, D.R. Smith, N. Ameral, M. Schmidtke, and R.A. Wong. 2023. Application of a catch multiple survey analysis for Atlantic horseshoe crab *Limulus polyphemus* in the Delaware Bay. Marine and Coastal Fisheries 15(5): e10250.

Atlantic States Marine Fisheries Commission (ASMFC). 2009a. A Framework for Adaptive Management of Horseshoe Crab Harvest in the Delaware Bay Constrained by Red Knot Conservation. Washington, D.C.

_____. 2009b. Horseshoe Crab Benchmark Stock Assessment and Peer Review Report. Washington, D.C.

_____. 2012. Addendum VII to the Fishery Management Plan for Horseshoe Crab. Arlington, VA.

_____. 2019. Horseshoe Crab Benchmark Stock Assessment and Peer Review Report. Arlington, VA.

_____. 2022. Revision to the Adaptive Resource Management Framework and Peer Review Report. Arlington, VA.

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

Fabricius, C., and G. Cundill. 2014. Learning in adaptive management: insights from published practice. Ecology and Society 19: 29 http://dx.doi.org/10.5751/ES-06263-190129

Fuller, A.K., D.J. Decker, M.V. Schiavone, and A.B. Forstchen. 2020. Ratcheting up Rigor in Wildlife Management Decision Making. Wildlife Society Bulletin 44:29-41. DOI: 10.1002/wsb.1064

Gillings, S., P.W. Atkinson, A.J. Baker, K.A. Bennett, N.A. Clark, K.B. Cole, P.M. González, K.S. Kalasz, C.D.T. Minton, L.J. Niles, R.C. Porter, I.D.L. Serrano, H.P. Sitters, and J.L. Woods. 2009. Staging behavior in Red Knot (Calidris canutus) in Delaware Bay: Implications for monitoring mass and population size. The Auk 126:54–63.

Lindberg, M., and E. Rexstad. 2002. Capture-Recapture Sampling Designs. Pages 251–262 in A.H. El-Shaarawi and W.W. Piegorsch, editors. Encyclopedia of Environmetrics. John Wiley & Sons, Chichester, UK.

McGowan, C. P., J.E. Hines, J.D. Nichols, J.E. Lyons, D.R. Smith, K.S. Kalasz, L.J. Niles, A.D. Dey, N.A. Clark, P.W. Atkinson, C.D.T. Minton, and W. Kendall. 2011. Demographic consequences of migratory stopover: linking red knot survival to horseshoe crab spawning abundance. Ecosphere 2:1–22.

_____, J.E. Lyons, and D.R. Smith. 2020. Decision Implementation and the Double-Loop Process in Adaptive Management of Horseshoe Crab Harvest in Delaware Bay. In Structured Decision Making: Case Studies in Natural Resource Management. Johns Hopkins University Press, Baltimore, MD.

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

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

Piersma, T., T. Lok, Y. Chen, C.J. Hassell, H.-Y. Yang, A. Boyle, M. Slaymaker, Y.-C. Chan, D.S. Melville, Z.-W. Zhang, and Z. Ma. 2016. Simultaneous declines in summer survival of three shorebird species signals a flyway at risk. Journal of Applied Ecology 53:479–490.

Schaub, M., and F. Abadi. 2011. Integrated population models: a novel analysis framework for deeper insights into population dynamics. Journal of Ornithology 152:1–11.

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

Smith, D.R., M.J. Millard, and S. Eyler. 2006. Abundance of adult horseshoe crabs (Limulus polyphemus) in Delaware Bay estimated from a bay-wide mark-recapture study. Fisheries Bulletin 104:456-464.

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

Tucker, A.M., C.P. McGowan, J.E. Lyons, A. DeRose-Wilson, and N.A. Clark. 2022. Species-specific demographic and behavioral responses to food availability during migratory stopover. Population Ecology 64:1–16.

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

Williams, B.K. 2011. Adaptive management of natural resources--framework and issues. Journal of Environmental Management 92(5):1346--1353.

_____, and E.D. Brown. 2018. Double-loop learning in adaptive management: the need, the challenge, and the opportunity. Environmental Management 62:995-1006.

_____, J.D. Nichols and M.J. Conroy. 2002. Analysis and Management of Animal Populations. Academic Press.

Wong, C., Y. Jiao, and E. Hallerman. 2023. Results of the 2022 Horseshoe Crab Trawl Survey: Report to the Atlantic States Marine Fisheries Commission Horseshoe Crab and Delaware Bay Ecology Technical Committees. Virginia Tech, Blacksburg, VA. pp. 28.