

Ecosystem Indicator Report Final Analysis

Katie Lankowicz

27 March, 2024

Contents

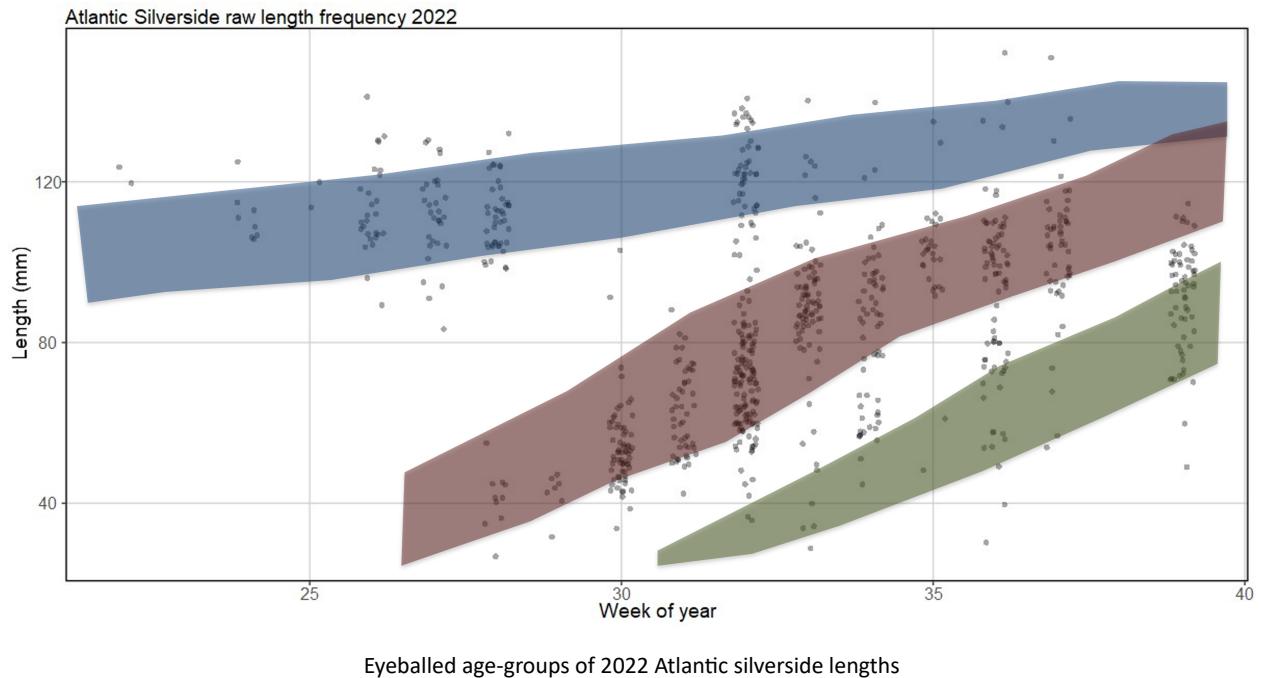
1 Growth	1
1.1 The problem	1
1.2 Kernel density estimation and modes	3
1.3 Mixed distribution modeling	4
1.4 Eyeballing	5
1.5 GLMs to model age-group length bounds	6
1.6 Modeling growth	7
1.7 Results	14

1 Growth

Last time we met, figuring out how to quantify growth rates was escaping me. I have since developed a mostly statistically-robust method of accomplishing this task. There's a bit of less-robust handwaving, but not too much. Let's go through it.

1.1 The problem

We sample a population of mixed age-based groups. I won't say cohort, because I think that will incite arguments about the nature of a "cohort" and obscure my point here. Those groups, by nature of being different ages, have distinct size distributions and growth rates. We can sometimes identify these groups just by eyeballing length distributions over time, like the theoretical groups I plotted below based on Atlantic silversides 2022 length distribution.



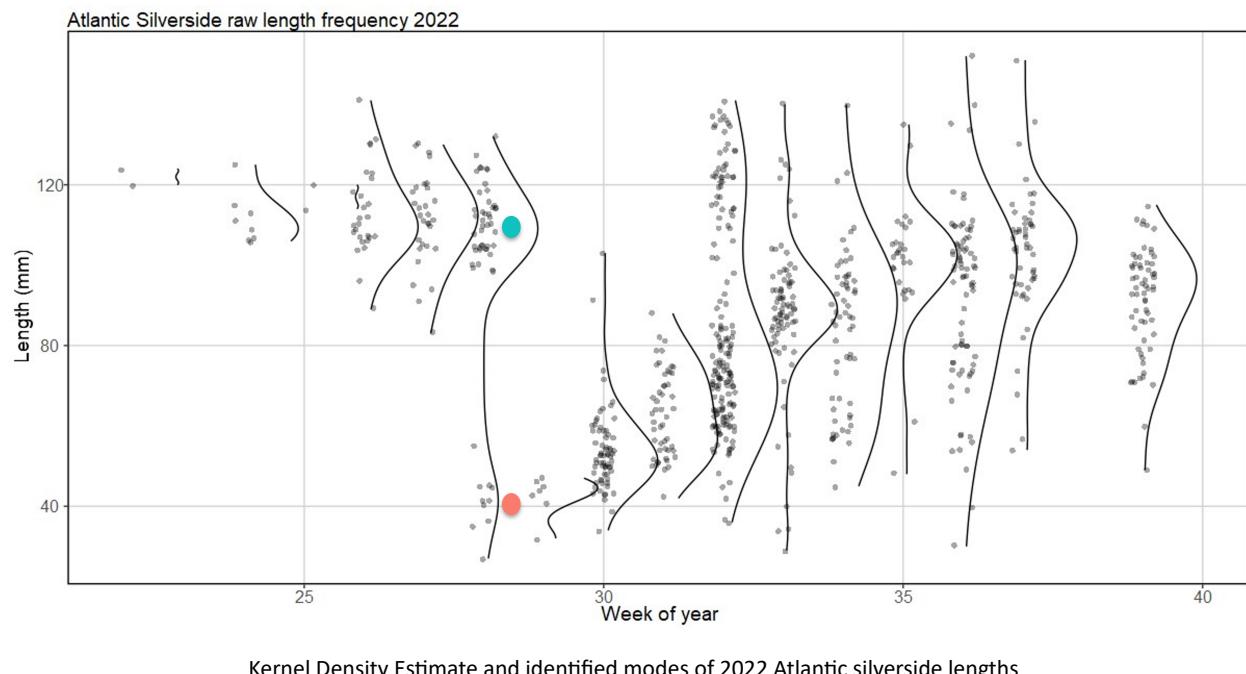
We need to separate the groups as best as possible for all species-years combinations so we can more clearly see realistic growth rates. Note that the eyeballed version presented here is from probably the most clear example of separate age groups out of all our species-year combinations, and the “eyeball” method will be neither consistent nor robust for more noisy data.

1.2 Kernel density estimation and modes

The first step of separating age-groups is to make some assumptions. Let's assume that the lengths of conspecifics in the same age group come from a normal distribution, and the range of each age-group's length distribution has to be reasonably compact (less than 40% of the maximum asymptotic body length for the species, as pulled from FishBase).

We then plot a kernel density estimate of lengths by species for each week of sampling. Considering our assumptions, we should see peaks of length-density around the mean length for each age group. Deterministic methods can be used to determine both the number of modes present in the length-density distribution and the length-values at those modes. These should theoretically be mean length for an age-group.

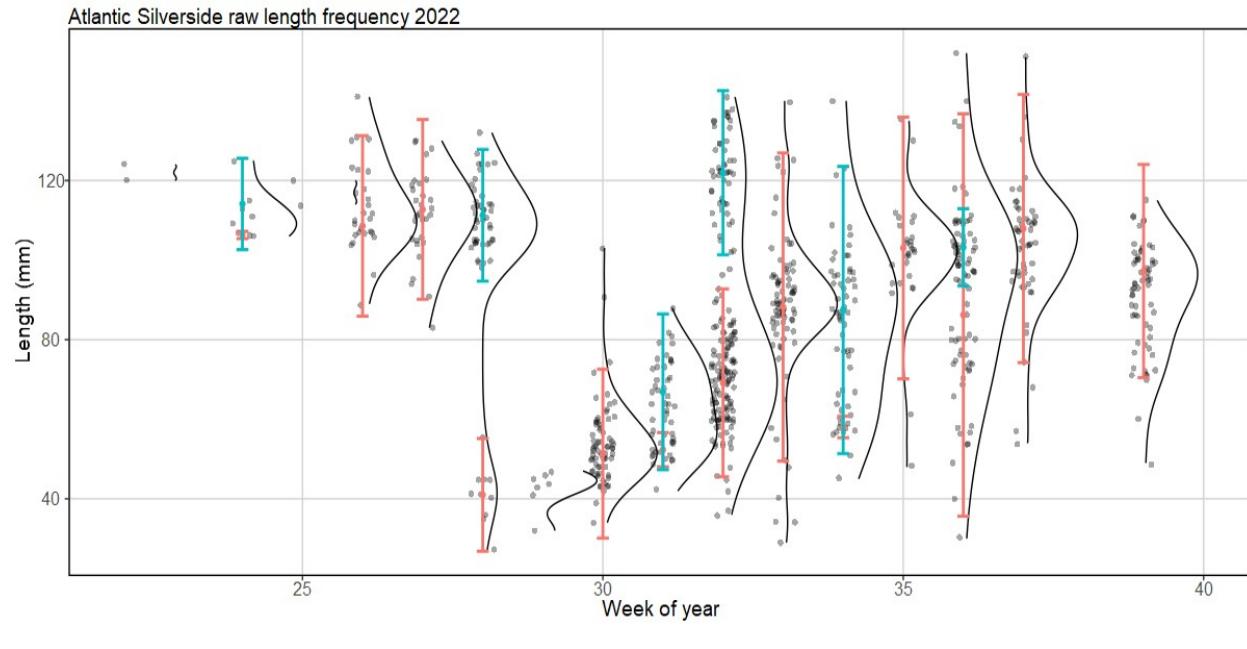
At this point, we must remove observations from weeks in which fewer than 7 fish of a species were lengthed— it isn't enough data to continue. This limits this analysis to the most frequently-encountered teleosts: alewife, atlantic herring, silverside, mummichog, sandlance, and winter flounder. The approach does not work for other commonly-encountered species (tomcod, stickleback spp, sculpin spp) because there aren't enough observations of lengths. Finally, I am hesitant to use this approach for green crabs. I don't know if my length and growth assumptions hold for crustaceans. I can try, if there is interest.



1.3 Mixed distribution modeling

From here, we use mixed distribution modeling to calculate the upper and lower bounds that create a 95% confidence interval around our identified mean lengths, effectively identifying the length range of each age-group. Again, our assumption is that distribution of lengths is normal, so we use an expectation-maximization (EM) algorithm for mixtures of univariate normals as our approach. The normal-mix EM model is given the lengths of all conspecifics per week and the number of modes identified in the previous step. It then returns mu (means of each component), sigma (standard deviations), and lambda (mixing proportions). For each component, we calculate an approximation to a 95% CI by finding 2 standard deviations around the mean.

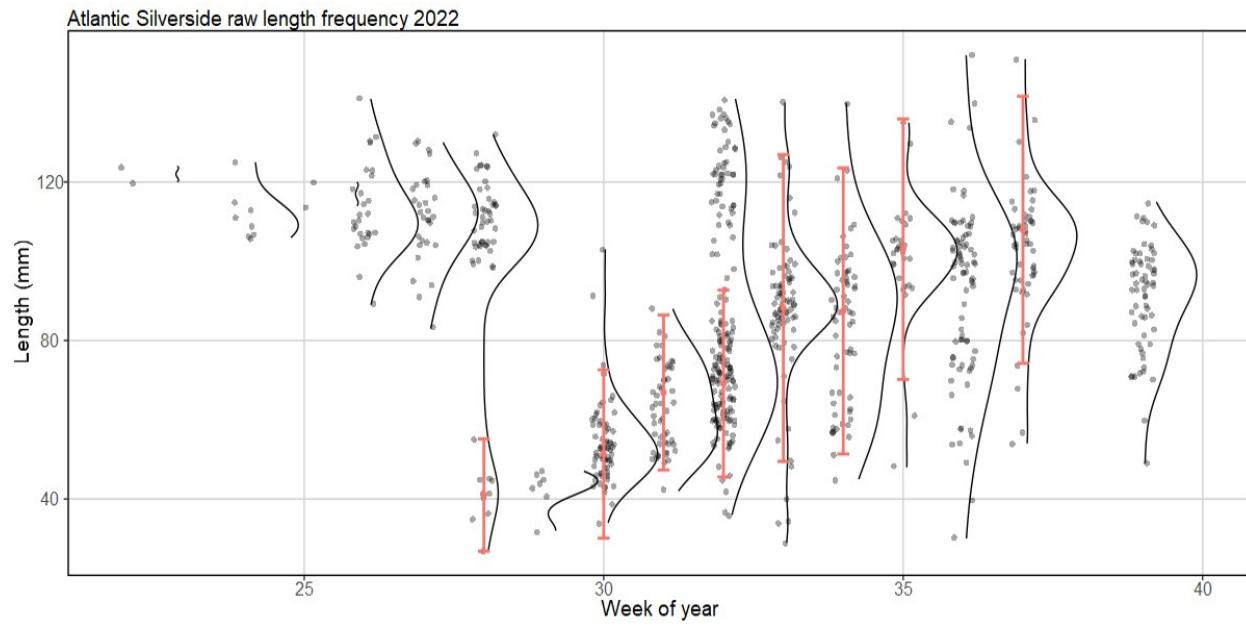
This gets us pretty close to where we want to be. We can remove distributions where the upper or lower bounds are biologically invalid (less than 0 cm or above the reported maximum asymptotic body length) or biologically improbable (range is greater than 40% of maximum asymptotic body length). We can also melt together distributions which overlap each other by more than 80%, as this isn't enough separation to clearly identify age-groups anyway.



1.4 Eyeballing

The next step is to give age-groups a distinct identifier and confirm that their ID remains the same between weeks. I haven't come up with a good method to automate this. I just go off my plots and returned mixed distribution model values and track age-groups by eye. Here's a theoretical example of what I do: in week 1, we identify one distribution of mean length 100 cm. In week 2, we identify two distributions, one of mean length 102 cm and one of mean length 20 cm. In week 3, we identify one distribution of mean length 50 cm. We would assume that the groups with mean length 100 cm in week one and mean length 102 cm in week 2 are the same age-group growing over time and make sure they have the same age-group identifier. The groups with mean length 20cm in week 2 and mean length 50 cm in week 3 are not linked to each other, nor to age-group 1. They need unique identifiers.

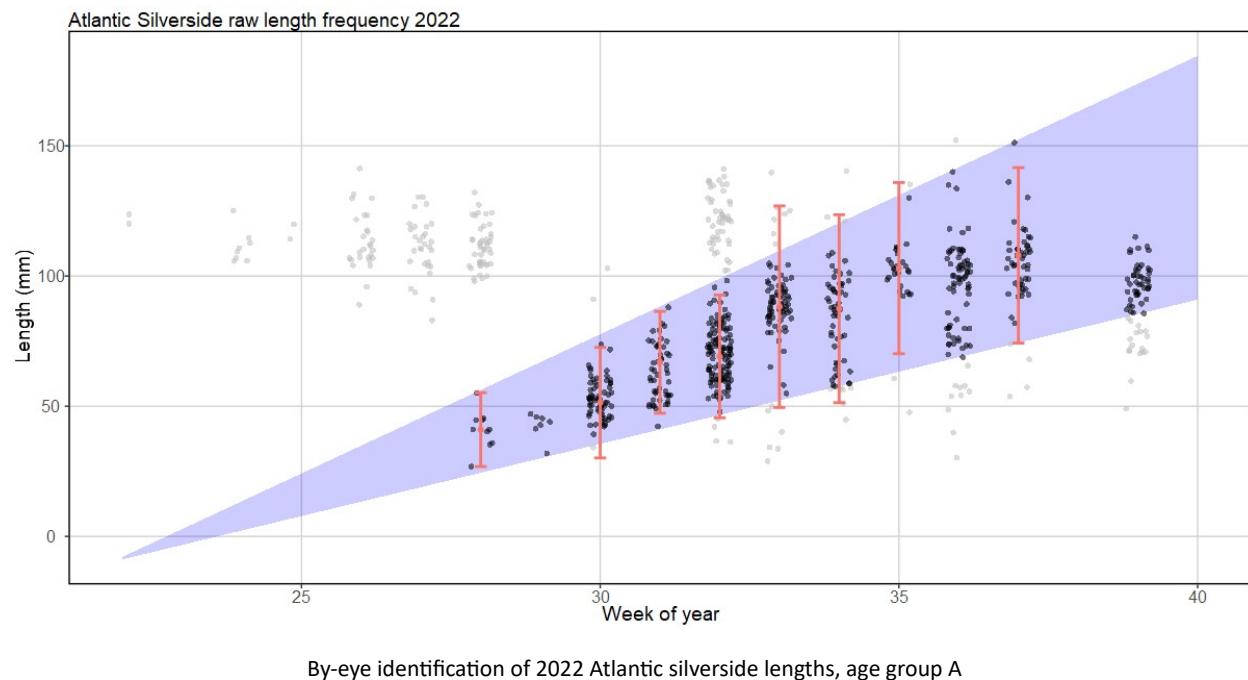
I'm sure there's some kind of way to code this out to do it automatically, but it wasn't worth my time to figure it out with all the edge cases. Importantly, I also did not want to *a priori* assume any maximum or minimum growth rates, which would have made this process much easier.



1.5 GLMs to model age-group length bounds

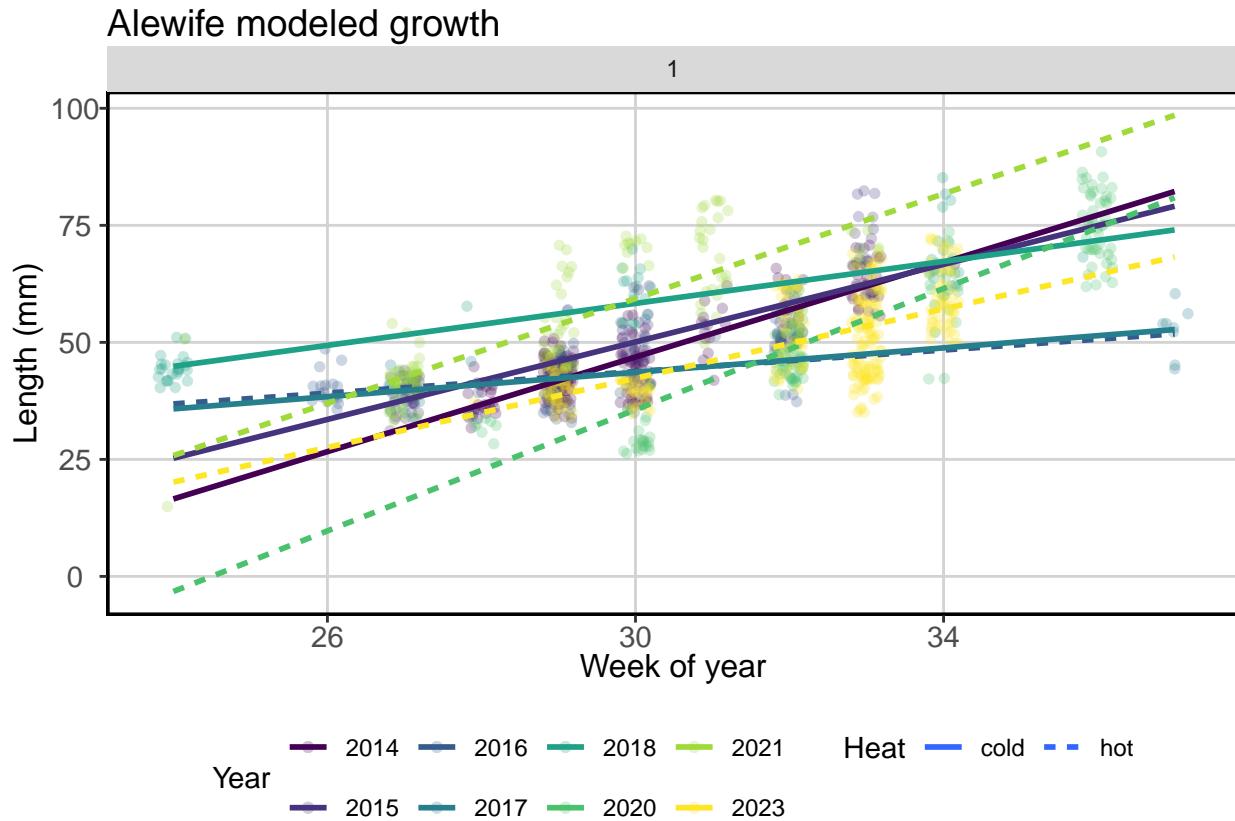
This leaves us with distributions of length for various age-groups for some weeks of a sampling year. Remember that we did remove some biologically improbable distributions, so sometimes we are left with no identified length distributions in a week. This is fine. The lower bounds, means, and upper bounds for each age group *generally* are monotonically increasing for the remaining weeks, but there is some variation. We also need to fill the gaps we made. The next step is therefore to model the upper and lower bounds of length in each week. Here, another assumption is made— that growth in this extremely short period (typically June 8 - August 28) is approximately linear. We therefore use generalized linear models to characterize length distribution bounds.

Once the GLM process has been completed, we can assign age-groups of fish accordingly.

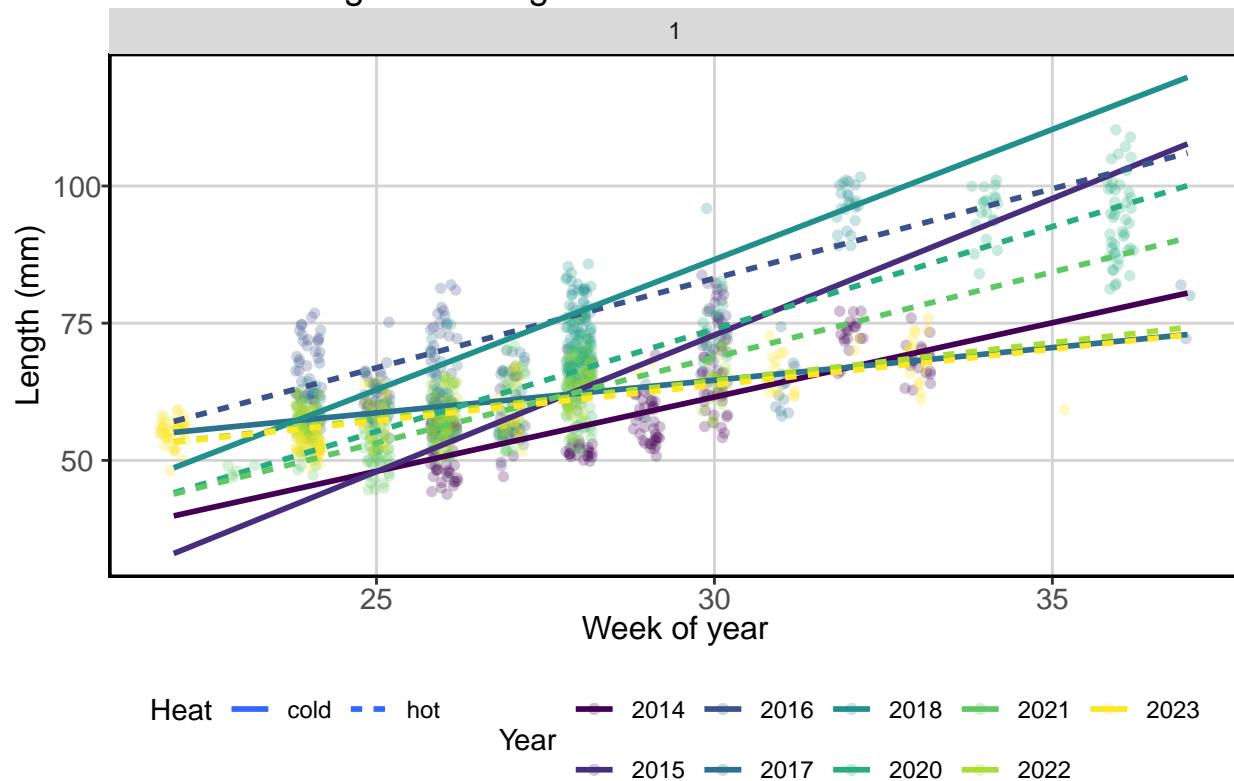


1.6 Modeling growth

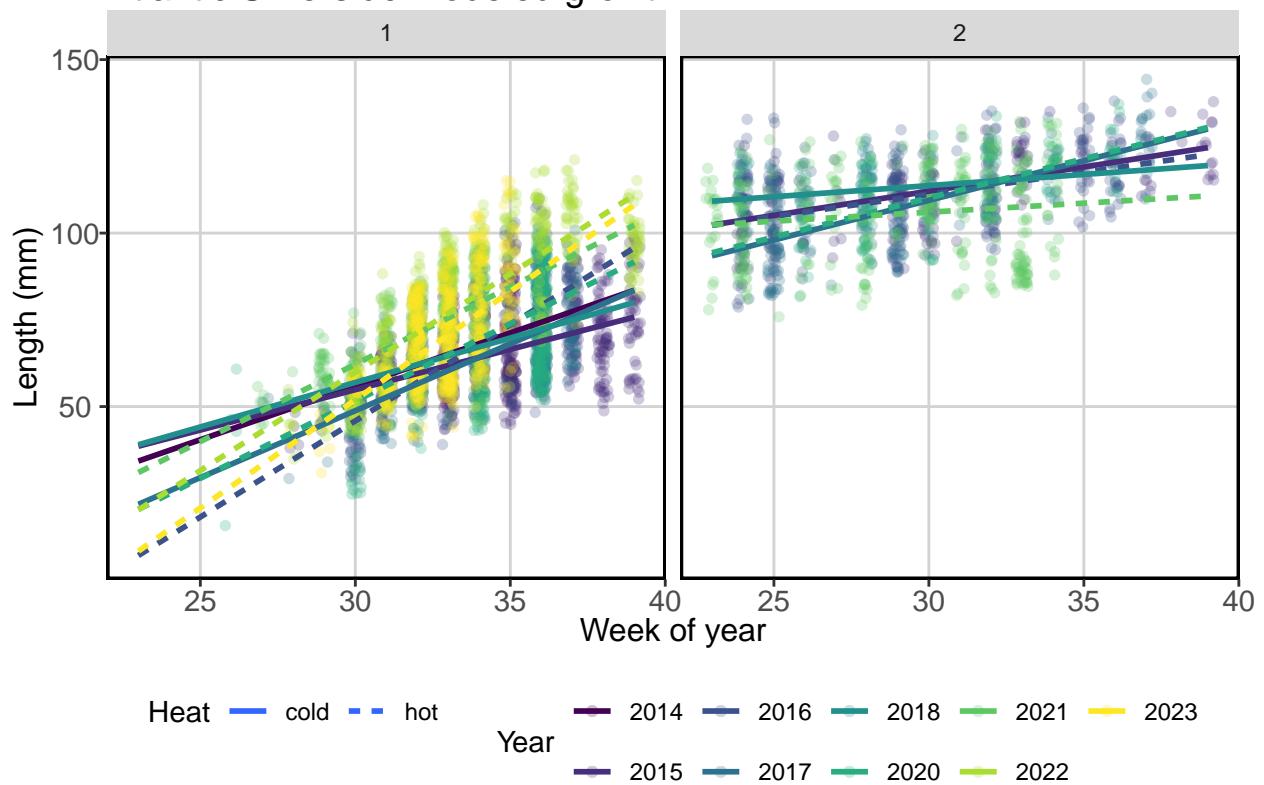
This last bit is what finally gives us the results we wanted. Now that we have assigned fish into age groups by length in each week, we can model the growth of those age groups. Again, we assume that growth is approximately linear over our short time period and use a generalized linear model. Growth should be equal to the slope of the GLM.



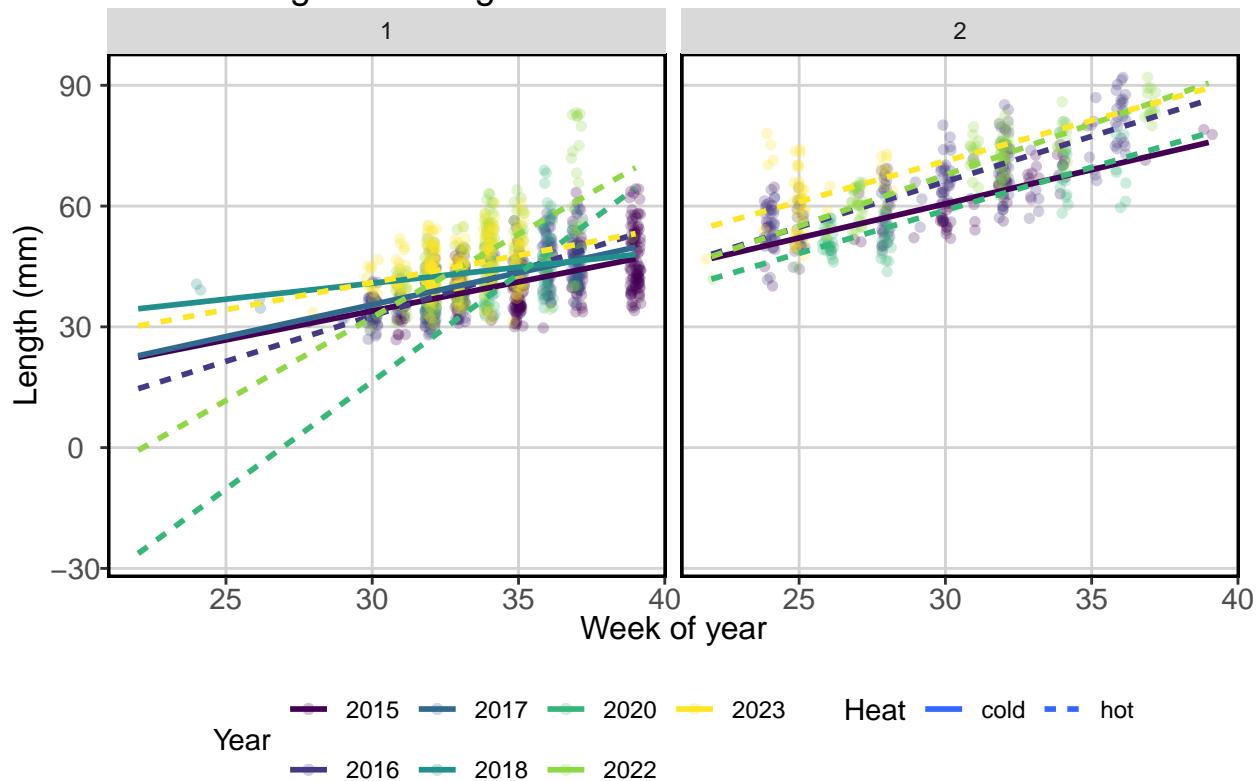
Atlantic Herring modeled growth



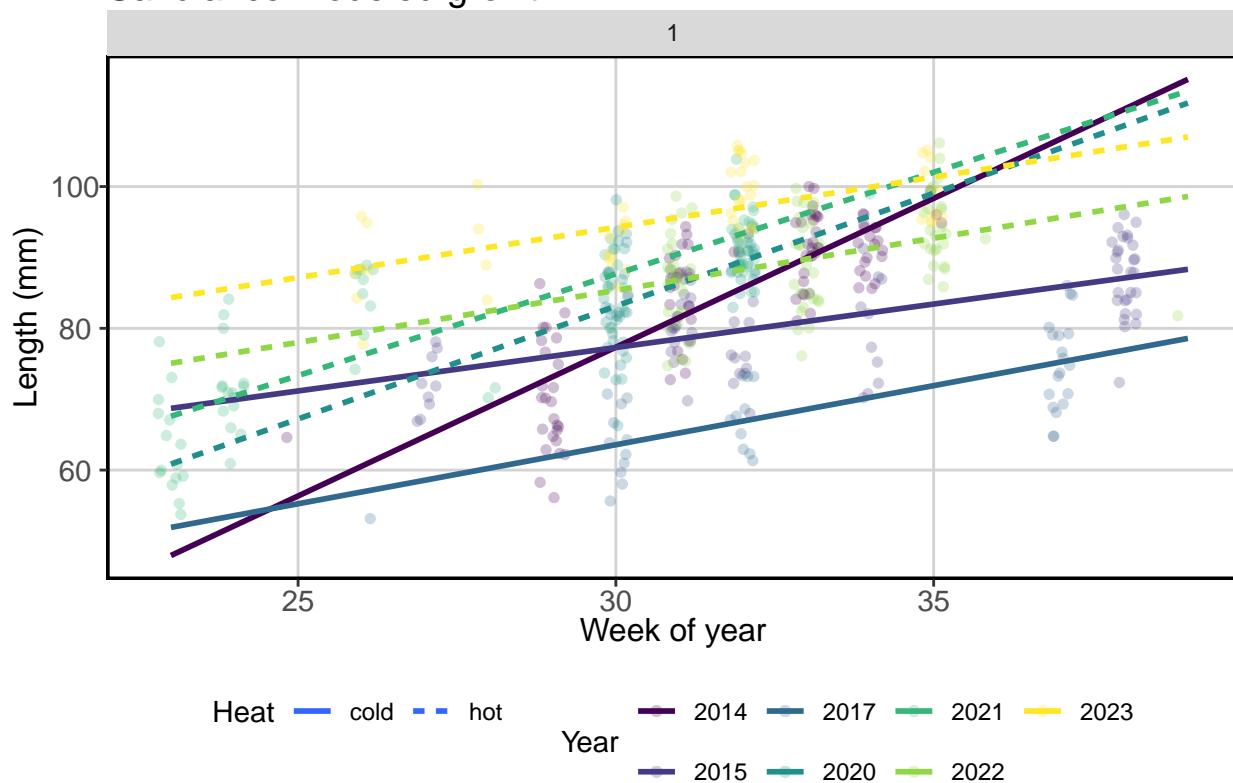
Atlantic Silverside modeled growth



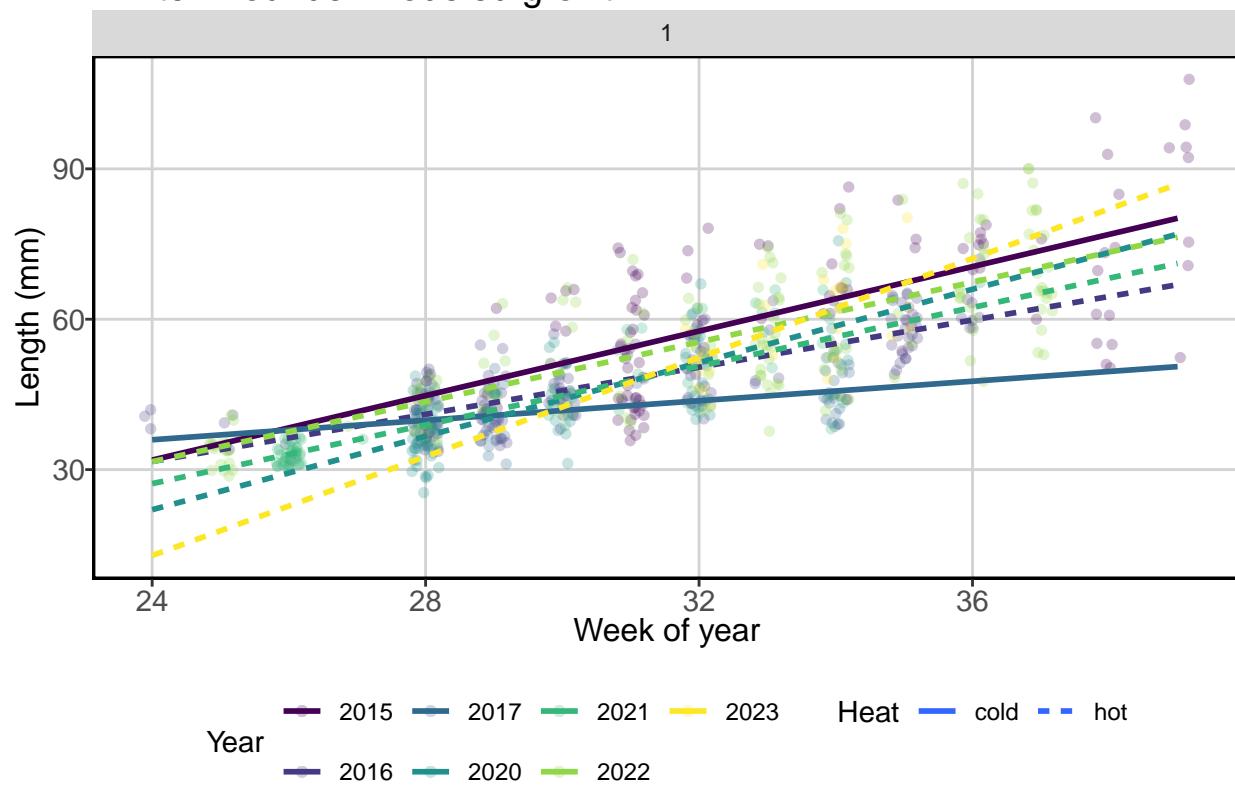
Mummichog modeled growth

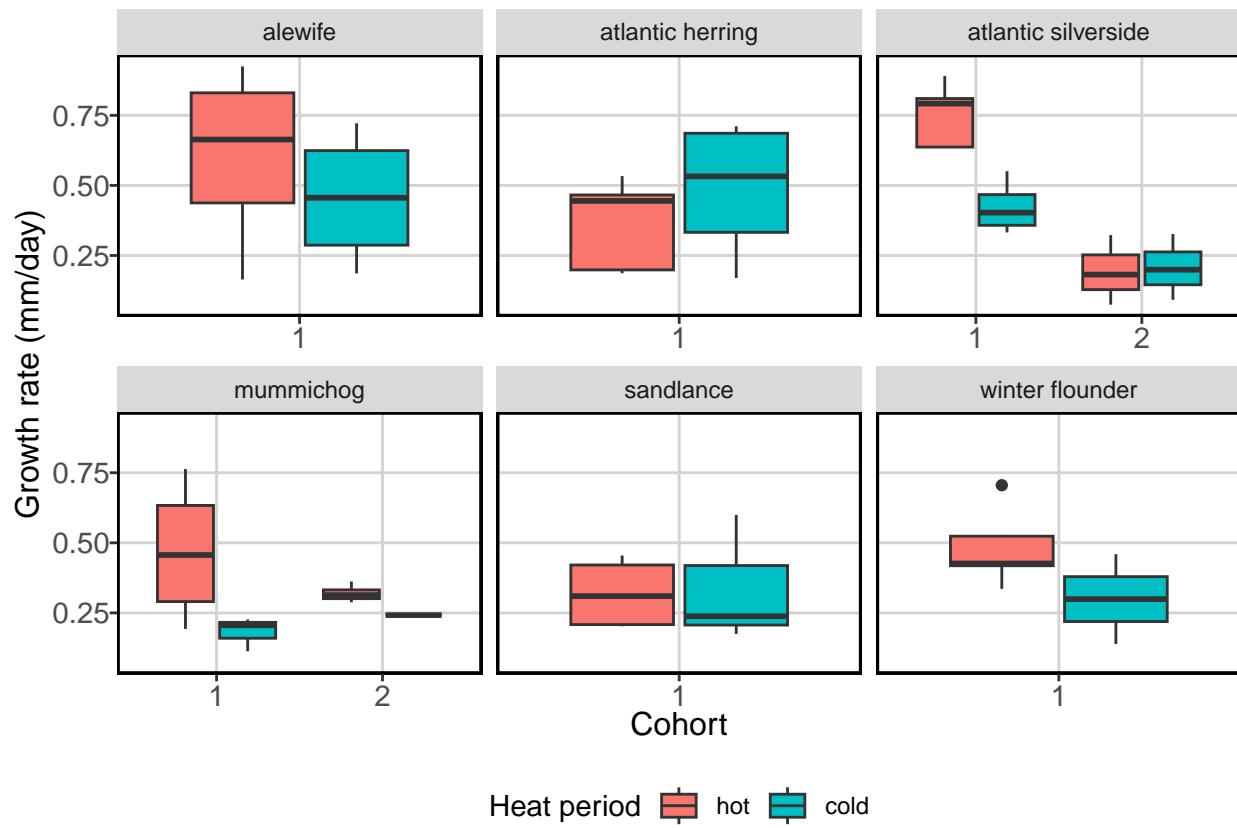


Sandlance modeled growth



Winter Flounder modeled growth





1.7 Results

Of the six modeled species, 2 are likely north of their populations' centers of gravity: atlantic silverside and mummichog. We would expect to see increased growth rates in hotter years. Because mummichog and silversides are short-lived (4 and 2 years expected lifespan, respectively) and lifelong residents of the inshore area, we were able to model growth rates for 2 age groups for both species. We found a trend of increased growth rates in hotter periods for age groups 1 and 2 for mummichogs, and age group 1 for silversides. Silverside age group 2 had a slightly increased growth rate in colder years. The only significant difference in growth rates for any species occurred in the silverside age-group 1: significantly higher growth rates in hotter years than colder years.

Three of the remaining species (alewife, Atlantic herring, and winter flounder) are likely south of their populations' centers of gravity. Increased temperatures should lead to slower growth rates. However, no significant differences were found, and only Atlantic herring exhibited a trend of faster growth in colder years. It's important to note that these species have more complicated migration strategies than the obligate inshore dwellers. It's possible that the cumulative effects of temperature in areas outside of Casco Bay are driving these differences in growth rates.

Sand lance are particularly weird; the species we are sampling is most likely to be *Ammodytes americanus*, which is typically found in coastal areas shallower than 20m and with sediment of an appropriate grain size for them to bury themselves in. Their level of seasonal migration is unknown, though they do go through a period of winter dormancy, where they bury themselves for extended periods of time. Casco Bay is not that far from the species' calculated center of gravity (about 270km north, at the northern tip of Nova Scotia). There is very little difference in growth rates we have calculated between cold and hot periods. It's possible that the population dynamics and growth of sand lance are currently being impacted more by habitat availability, food availability, and predation than by the temperature shifts we've seen.