Merrill Rudd

FSH 507 Project

June 9, 2016

How important is it to account for spatial variation in growth for data-limited fisheries?

**Introduction**

Over 80% of global catch occurs in fisheries without formal stock assessments (Costello et al. 2012), often due to the fact that data and capacity are limited. In capacity-limited situations, length measurements are often the easiest type of data to reliably collect. The length composition of the catch holds information on the size at which fish become selected to the gear, and the proportion of the catch in each length bin. Stock assessments that make use of length composition data make many assumptions about biological characteristics, such as growth, maturity, and mortality, in order to make inferences about how the length composition of the catch (i.e. portion of the population vulnerable to the gear) may have changed with fishing pressure (Hordyk et al. 2014a, 2014b). Other types of assessments make similar assumptions about biological characteristics and relate the mean length in the catch directly to fishing mortality (Gedamke and Hoenig 2006, Ault et al. 2008, Nadon et al. 2015).

Many stock assessment methods that rely on length composition or mean length data also assume that the populations are in equilibrium, i.e. that fishing mortality and recruitment have not changed significantly over a time period relative to the biological traits of the species. However, it is very possible this assumption is violated for many species harvested by small-scale fisheries, particularly in coral reef ecosystems for which this type of assessment is commonly applied. An extension of these length-based assessment methods has been developed to account for the time-variation in recruitment and fishing mortality using a mixed effects modeling approach (Rudd and Thorson *in prep*). Simulation testing has demonstrated that this method is unbiased in estimating spawning potential ratio (SPR; a common reference point for data-limited fisheries), when at least one year of length composition data is available and assumptions about biological characteristics can be made. Because this mixed effects model is an integrated assessment, as more years of length composition data become available, precision on estimates of SPR improves. The mixed effects approach on length composition data allows for the possibility that more processes than fishing mortality are acting on the population to explain the lengths in the catch that arise.

Another important process governing the length composition of the catch is the growth of individuals within the population. Due to the strong spatial heterogeneity in coral reef ecosystems, it is possible that growth of one species may be very different between reef areas (MacNeil and Connolly 2015). This is an important consideration for data-limited fisheries, since it is often uncertain from which reef fish are captured, samples of length composition may be low (i.e. difficulty measuring fish across the spatial extent of the fishery over time), and information on stock structure may be very limited. These factors commonly lead to analysts pooling length composition data together across landing sites for use in the stock assessment. This practice assumes that there is no variation in growth across sites. The objective of this project is to quantify bias and imprecision in estimates of stock status in the mixed effects model relying on length composition data when spatial variation in growth is ignored.

**Methods**

I used a simulation test with “control” and “test” operating models to compare the ability of a single estimation model to output unbiased and precise estimates of SPR, the ratio of potential lifetime egg production at the current level of fishing to that with no fishing. The population dynamics in the simulation are modeled after *Siganus sutor*, one of the top commercially important species in the coral reef fishery for which an assessment will be conducted as part of my thesis. The true values used to generate the true populations and data, input values into the estimation model, as well as which parameters are fixed or estimated in the estimation model, can be found in Table 1. I assumed that ten years of length composition data would be available (similar to the Kenyan coral reef fishery that I will be assessing in my thesis), and modeled twenty years because more than ten years of information on the population are contained within the length composition data due to growth and recruitment lag.

*“Control” operating model – no spatial variation*

The first “control” operating model does not include spatial variation in the growth process. The underlying population dynamics are age-structured:



(1)

Where *Na,t* is the abundance at age *a* over time *t, Rt* is recruitment over time, *M* is the annual natural mortality rate (fixed constant over time), *Ft* is the annual fishing mortality rate over time, *Sa*is the selectivity of the gear at age, and *amax* is the maximum age in the population.

Recruitment, or the number of age-0 fish, is specified into the operating model with lognormally-distributed process error:



 (2)

Where *μt* is the specified recruitment pattern, *εt* are the annual deviations in recruitment, and *σR* is the recruitment standard deviation. The specified recruitment pattern was fixed constant at 1 so that the true recruitment, *Rt* represents relative increases and decreases through the annual deviations only. Because the length composition data does not contain information on absolute size of the population, the assessment focuses only in relative terms (e.g. aiming to get an idea of relative abundance, SPR, or fishing mortality rate, not total biomass).

We explored three different patterns of fishing mortality over time when simulating data, including constant, endogenous, and a ramp-up pattern. The different patterns of fishing mortality hold varying degrees of information on the population parameters (Magnusson and Hilborn 2007), thus the comparison in performance across true fishery dynamics helps confirm that the assessment method can work in many situations. The constant fishing mortality was held at 0.25. The endogenous fishing mortality scenario uses an effort dynamics model from Thorson et al. (2013) that replicates a developing fishery, with an equilibrium fishing mortality rate of 0.25, initial fishing mortality of 0.05, and the *Frate* value of 0.2.The ramp-up pattern represents a situation where the fishing mortality increases from an initial value of 0.05, to a peak of 1.0, and then a “managed” state at 1/3 of the maximum fishing mortality rate. These time series of fishing mortality were generated separately and specified in the operating model to generate the underlying age-structured population and length composition data that arise from the given pattern of exploitation.

Growth follows a von Bertalanffy model:



(3)

Where *L*ais the predicted length-at-age, *L∞* is the asymptotic length (36.2 cm), *k* is the Brody growth coefficient, and *t0­* is the age when length is equal to 0. Weight at age is an allometric function of the length at age:

 (4)

Maturity-at-age follows a logistic function:



(5)

Where *ma* is the proportion of the population mature at each age, and *a50*­ is the age at 50% maturity. Selectivity also follows a logistic function:



(6)

Where *s­50­* is the age at 50% selectivity and *s95* is the age at 95% selectivity. Annual spawning biomass is a function of the weight, proportion mature, and abundance in each year. Catch follows the Baranov equation:

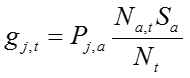
 (7)

Length measurements are then generated from the catch based on the probability of being in a length bin (1 cm bins) given age:



(8)

Where *Pj,a* is the probability of being in a length bin *j* given age, *ϕ* is the cumulative normal distribution, and *CVL* represents the variability in the age-length relationship. I then calculate the probability of being harvested in a given length bin each year (*gj,t*):



(9)

The length composition data is then generated using a multinomial distribution to get the predicted number of individuals capture in each length bin over time (*y*):

 (10)

Where *n* is the effective sample size of length measurement observations. I assumed 100 independent length measurements were available each year. I realize in writing this up this is likely too many, as in my real assessment I assumed the annual effective sample size to be the number of times the data collectors visited the landing sites in a year, since the length measurements on any given day would likely not be independent.

*“Test” operating model – spatial variation in growth process*

The equations used in the “test” operating model with spatial variation in the growth process are the same, except that the underlying age-structured dynamics occur by site, length composition data is generated by site, and the asymptotic length has 1D spatial variation along the coastline. The spatial variation in the growth process is governed by spatial variation in asymptotic length because asymptotic length is a function of both anabolism and catabolism, whereas the Brody growth coefficient, another option of a parameter that could have been varied by site, is only a function of anabolism.

Spatial variation in asymptotic length across the total 15 sites (*I*) was generated using a Gaussian random field with variation  of 0.01 and scale δ of 2 using the RandomFields package in R:

 (11)

Where *li* is the latitude at site *i* approximated along the Kenyan coastline, *X* is the random field, *L∞i* is the asymptotic length at site *i,* σ is the standard deviation of the site variability set at 1, and β is the slope of the spatial trend along the coastline set to 0.02. The vector of *L∞i* was scaled so that the mean asymptotic length across sites for each simulation was equal to the value with no spatial variation:

 (12)

If the simulated “true” asymptotic lengths at any site were less than the length at 50% maturity (20.3 cm, from the literature), a situation that does not make biological sense, the process was re-iterated until the asymptotic lengths across sites were all greater than the length at 50% maturity (Hicks and McClanahan 2012). Examples of simulated site variation in asymptotic length are shown in Figure 1. The same time series of fishing mortality and recruitment, which include process error, were specified into the operating model so that the same dynamics were governing the population at each site. The only thing that varied between sites was the asymptotic length, which would then have an impact on the rate at which the fish grow through their age structure and the length composition in the catch that arises from that process.

For each site, length composition data was generated in the same manner as in the “control” operating model. For the “test” operating model, however, I summed the number of individuals observed at each site in each length bin, to generate a “pooled” length composition of the number of individuals in each length bin across all sites (Figure 2).

*Estimation model*

The underlying population dynamics follow the same equations as the operating model. The parameters estimated as fixed effects are the age at 50% selectivity and the annual fishing mortality rates. Random effects on time-varying recruitment are estimated, but the parameters of the distribution from which the random effects arise are fixed due to the fact that 1) there is no information in the length composition data to determine the scale of the population, represented as the mean value of recruitment across the time series (therefore it is fixed at 1), and 2) it is very difficult to estimate the standard deviation of the recruitment variability even in more data-rich situations (therefore fixed at 0.6). The only equation added to the estimation mode that varies from the operating model is a smoother on the fishing mortality rate, such that the estimates of fishing mortality one year cannot vary too much from the estimates of fishing mortality in the last year:

 (13)

Where  is the standard deviation on the fishing mortality rate between years that is fixed at 0.3.

*Comparison*

At this point, I have explained 6 scenarios for data generation: 3 time series of fishing mortality patterns, with no spatial variation in growth across sites, and with spatial variation in growth across sites. I then generate 1,000 iterations of length composition data for each of the 6 scenarios, running each through the estimation model to determine the ability to estimate SPR:

 (14)

This calculation assumes that recruitment is equal to 1 and the abundance in the fished (Nf) and unfished (N0) states are calculated at the fishing mortality in the last year. SPR is commonly used as an output/reference point for other data-limited assessment methods that rely on length composition or mean length data (Hordyk et al. 2014b, Nadon et al. 2015). Therefore, it is easy to compare SPR to the other length-based assessment methods for data-limited fisheries.

**Results**

From 1,000 iterations of different patterns in spatial variability in asymptotic length, estimates of stock status (spawning potential ratio) are over-estimated when spatial variation is ignored and length composition data is pooled across sites (Figure 3). The precision in estimates of SPR across 1,000 iterations of generated data does not appear to decline when ignoring spatial variability in growth. These findings remain true across all patterns of fishing mortality over time. This means that the estimator is biased high when spatial variation in growth is ignored.

When asymptotic length varies relatively evenly around the true value of 36.2 cm, there are no issues with accuracy or precision of estimates of SPR compared to the scenario when there is no spatial variation in asymptotic length (Figure 4A). However, in scenarios that have a few sites with much higher asymptotic length than the others, fishing mortality tends to be under-estimated (Figure 4B). This leads to the positive bias in estimates of SPR by thinking the population is not fished as hard as it really is. I expect the opposite pattern holds true when there are outliers in asymptotic length lower than that in many of the other sites – it would seem that the population is fished harder than it actually is.

**Discussion**

Given these results, analysts of length composition data for stock assessment should be aware of potential good or poor zones for growth if length composition data are being pooled or the value of asymptotic length is being assumed constant for all sites. This could be a problem in areas like Kenya, where I will be using this stock assessment method for coral reef fisheries. Kenyan coral reef areas are managed for protection from fishing, protection with minimal fishing, and open access. Therefore, some sites may have much different growth than others depending on how long the area protected from fishing has been in place and/or enforced. Usually the level of fishing would not be associated with growth rates, but in the Kenyan case there was a study that found variation in fecundity associated with different management types for marbled parrotfish in the same area (Locham et al. 2015). This implies the possibility that biological characteristics can change as a result of ongoing fishing pressure, and that the growth processes would need to be modeled differently across areas for both physical, biological, and anthropogenic reasons.

It is possible that the trend in asymptotic length modeled in this study leads to more iterations having some sites with asymptotic length much higher than the average, as opposed to having any iterations with sites much lower than the average. If this is true, this would explain the positive bias we see in estimates of asymptotic length when the spatial growth process is ignored. In that case, there would be no real difference in accuracy or precision in the ability to estimate SPR when the spatial growth process is ignored. It is possible, however, that there would be a higher probability of bias in estimates of stock status using real data if some sites pooled within the length data have much higher or lower asymptotic length than others. From a manager’s perspective, positive bias in estimates of SPR mean that from the stock assessment, we would perceive the stock to be in a better state than it really is. This is concerning from a conservation perspective, in that it may lead to recommendations for future harvest/management that may lead to overfishing.

A possible next step is to ask the question, “is it more important to account for temporal variation in fishing mortality, temporal variation in recruitment, or spatial variation in growth for estimates of stock status?” If it is determined that there are no outliers in growth across sites, the results of this project imply that it may not be important to account for spatial variation across sites when dealing only with length composition data. However, if there are large and/or uneven differences in asymptotic length across sites, then biases are possible. In this case, it may be convenient to add an option to the estimation model that estimates random effects on asymptotic length across sites. This would make the model more flexible to consider the possibility that the growth rate at each site arises from a distribution. However, length composition data does not contain enough information to tease apart time-varying fishing mortality from time-varying recruitment from site-varying growth. In this case, strong assumptions would need to be made on the other processes. Part II of this study could look at the relative importance of these assumptions.

Overall, this project was useful in adding spatial complexity to my operating model upon which to test the estimation model that considers time-, but not space-variation. This will be helpful in convincing reviewers and other colleagues that the estimation model is expected to be unbiased. It will also give me more confidence that the estimation model would do well when tested against the operating model used within the DLMtool (Carruthers and Hordyk 2015). All of this testing will help my case when comparing the estimation model used in this project against other stock assessment methods currently used for coral reef fisheries and other applications where catch data is not available.

**References**

Ault, J.S., Smith, S.G., Luo, J., Monaco, M.E., and Appeldoorn, R.S. 2008. Length-based assessment of sustainability benchmarks for coral reef fishes in Puerto Rico. Environ. Conserv. **35**(3): 221. doi: 10.1017/S0376892908005043.

Carruthers, T., and Hordyk, A. 2015. DLMtool: Data-Limited Methods Toolkit.

Costello, C., Ovando, D., Hilborn, R., Gaines, S.D., Deschenes, O., and Lester, S.E. 2012. Status and solutions for the world’s unassessed fisheries. Science (80-. ). **338**: 517–520. Available from http://www.sciencemag.org/content/338/6106/517.short [accessed 3 April 2014].

Gedamke, T., and Hoenig, J.M. 2006. Estimating mortality from mean length data in nonequilibrium situations, with application to the assessment of goosefish. Trans. Am. Fish. Soc. **135**(2): 476–487. doi: 10.1577/T05-153.1.

Hicks, C.C., and McClanahan, T.R. 2012. Assessing gear modifications needed to optimize yields in a heavily exploited, multi-species, seagrass and coral reef fishery. PLoS One **7**(5): e36022. doi: 10.1371/journal.pone.0036022.

Hordyk, A., Ono, K., Sainsbury, K., Loneragan, N., and Prince, J. 2014a. Some explorations of the life history ratios to describe length composition, spawning-per-recruit, and the spawning potential ratio. ICES J. Mar. Sci.

Hordyk, A., Ono, K., Valencia, S., Loneragan, N., and Prince, J. 2014b. A novel length-based empirical estimation method of spawning potential ratio (SPR) and tests of its performance, for small-scale, data-poor fisheries. ICES J. Mar. Sci.

Locham, G.A., Kaunda-arara, B., Wakibia, J., and Muya, S. 2015. Phenotypic Divergence in the Reproductive Traits of Marbled Parrotfish Leptoscarus vaigiensis ( Quoy and Gaimard , 1824 ) on Variably Protected Reefs in Kenya. **13**(1): 69–80.

MacNeil, M., and Connolly, S. 2015. Multiscale patterns and processes in reef fish abundance. *In* Ecology of fishes on coral reefs. *Edited by* C. Mora. Cambridge University Press, Cambridge, UK. pp. 116–126.

Magnusson, A., and Hilborn, R. 2007. What makes fisheries data informative? Fish Fish. **8**(4): 337–358. doi: 10.1111/j.1467-2979.2007.00258.x.

Nadon, M.O., Ault, J.S., Williams, I.D., Smith, S.G., and DiNardo, G.T. 2015. Length-Based Assessment of Coral Reef Fish Populations in the Main and Northwestern Hawaiian Islands. PLoS One **10**(8): e0133960. doi: 10.1371/journal.pone.0133960.

Thorson, J.T., Minto, C., Minte-vera, C. V, Kleisner, K.M., and Longo, C. 2013. A new role for effort dynamics in the theory of harvested populations and data-poor stock assessment. **1844**(September): 1829–1844.

Table 1. Variables used in the operating and estimation model, their descriptions, their true values in the operating model where applicable, same as the input values in the estimation model where applicable, and whether the value is fixed or estimated in the estimation model.



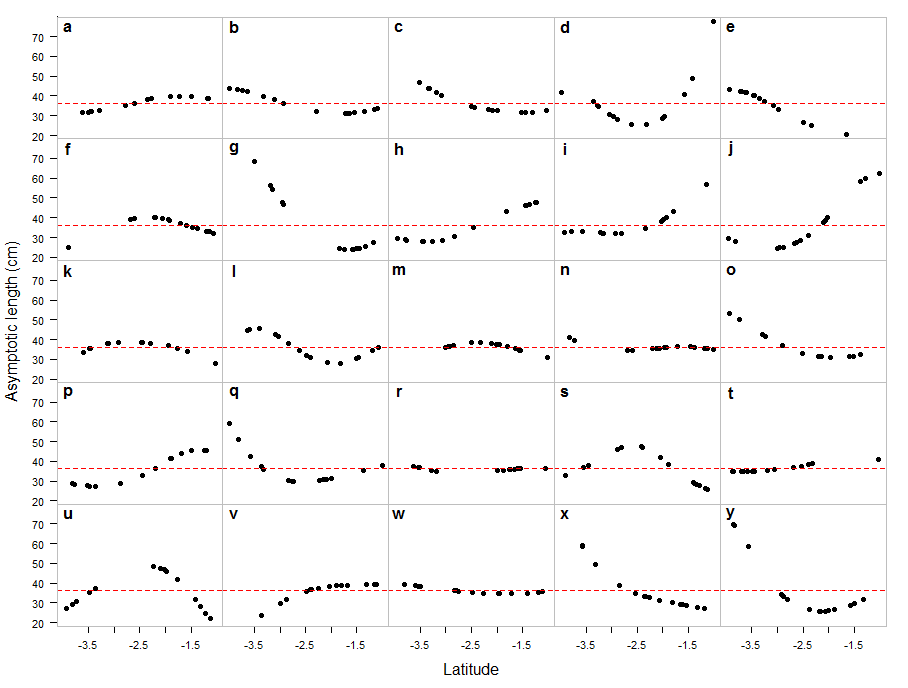
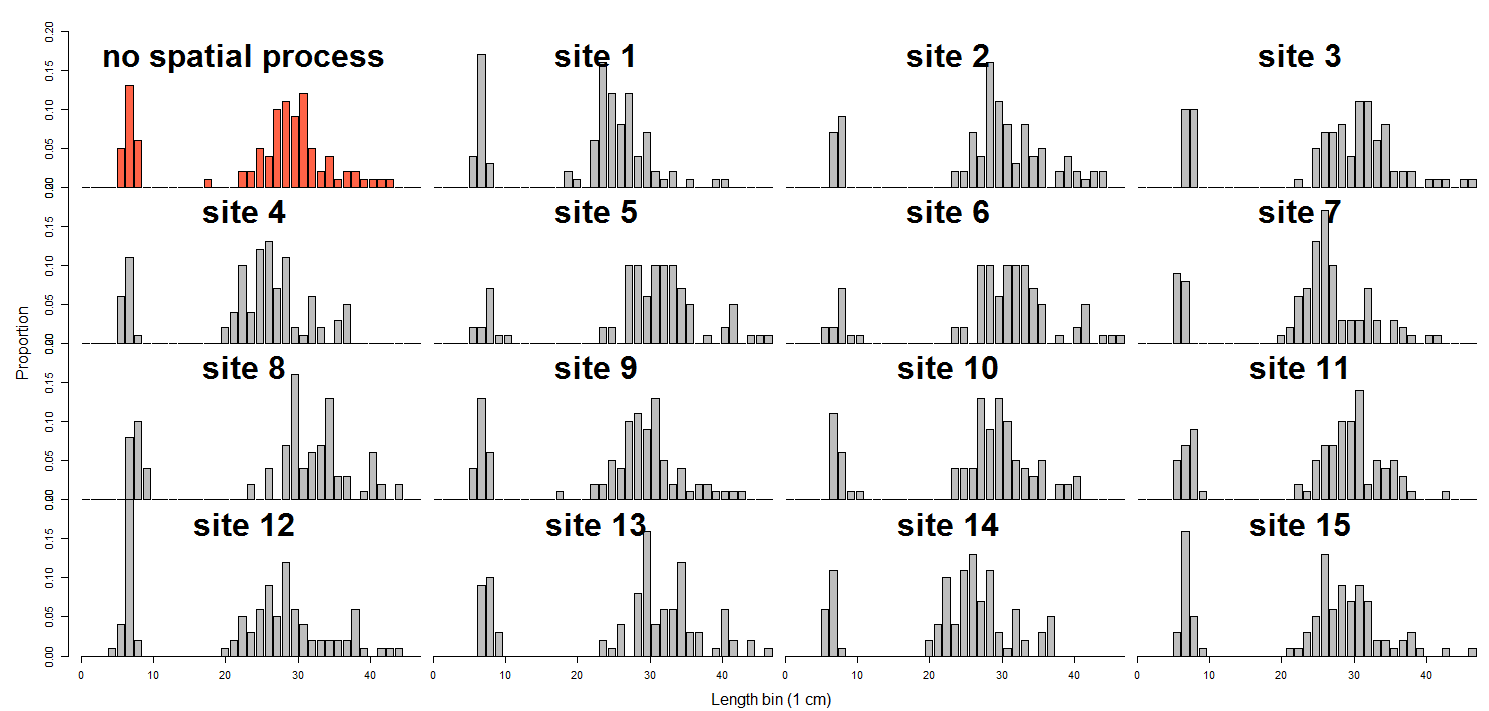


Figure 1. 25 examples of simulated asymptotic length across sites (black points), with the value of asymptotic length used for the scenario without spatial variation, and set as the mean in all scenarios, shown with the red dotted line.

A.



B.

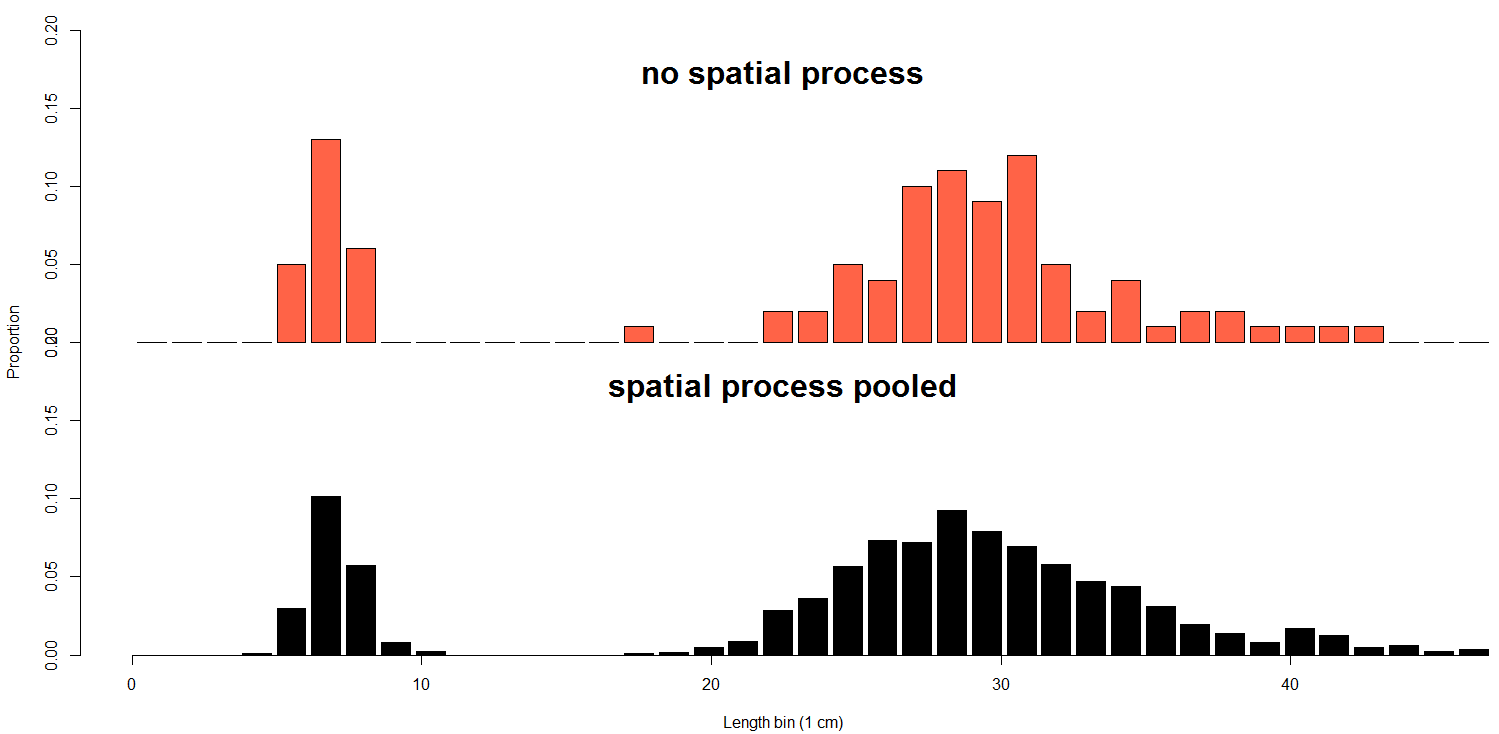


Figure 2. A. Example of length composition generated with no spatial variation in asymptotic length compared to that which was generated at each site and B. The length composition with no spatial process compared with the pooled length composition of all the length compositions at each site in panel A.

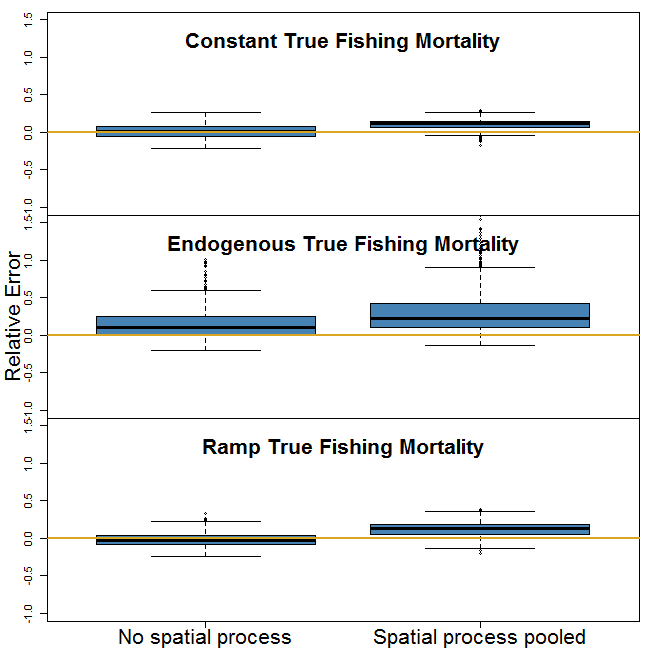
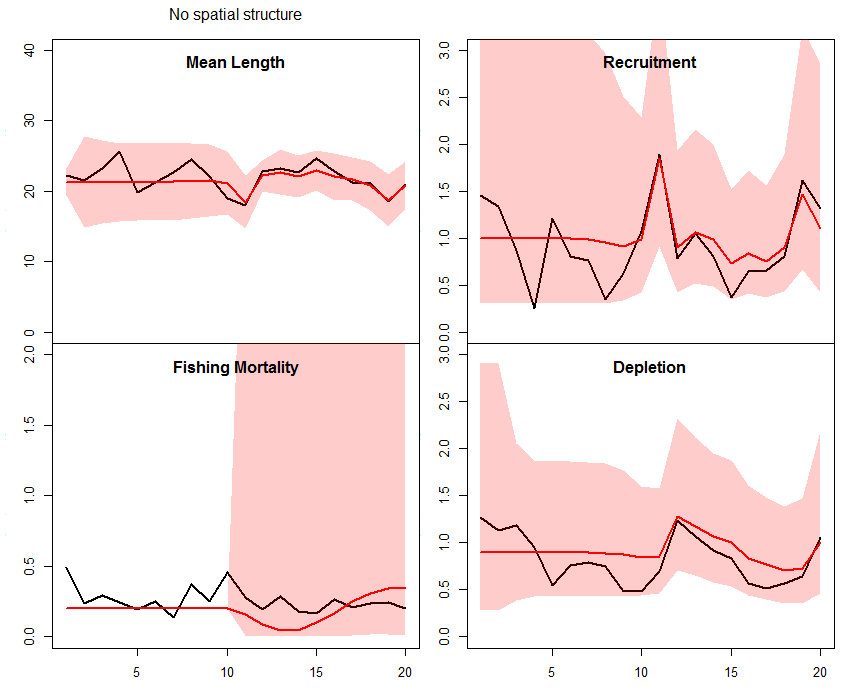
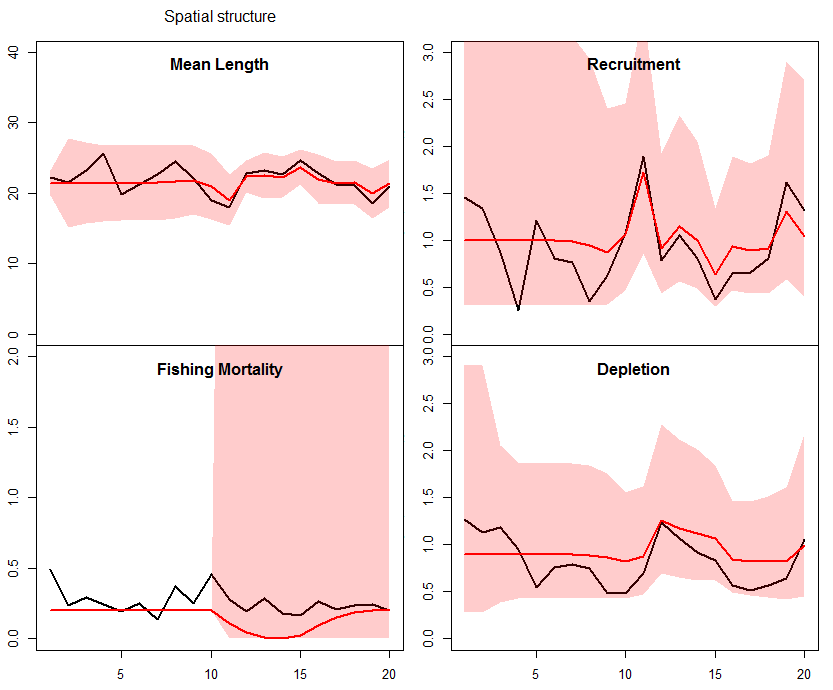


Figure 3. Relative error between true and estimated values of spawning potential ratio (SPR) when there is no spatial process in the operating model, and when there is a spatial process on growth but the data are pooled together.

1. Model fits from simulated data in Figure 1a.

1. Model fits from simulated data in Figure 1g.

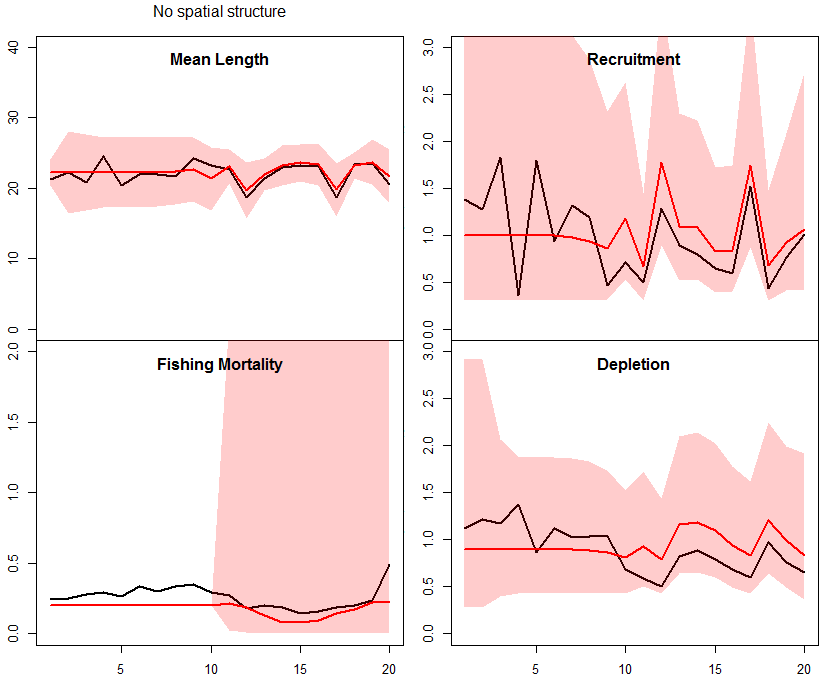
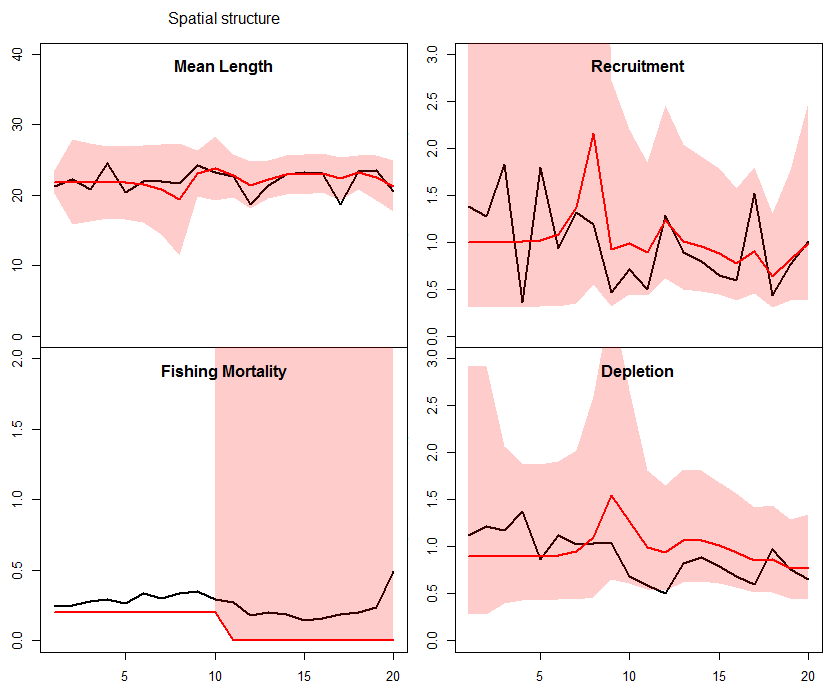
 

Figure 4. Model fits from single iterations of simulated data with no spatial variation in growth or spatial variation in asymptotic length. A. shows the results from the iteration in Figure 1a and B. shows the results from the iteration in Figure 1g. The black lines are the true values of mean length, recruitment, fishing mortality, and relative abundance (i.e. depletion), the red lines are the maximum likelihood estimates, and the red shading is the 95% confidence intervals.