Description of the Kuskokwim River Chinook Salmon Run Reconstruction and an Investigation of Data Weighting

A report to the Kuskokwim River Salmon Management Working Group

Ben Staton* and Matt Catalano, Auburn University

Lew Coggins, US Fish and Wildlife Service

Bill Bechtol, Association of Village Council Presidents, Bechtol Research,

Dan Gwinn, Biometric Research, LLC

*Corresponding Author: Ben Staton (bas0041@auburn.edu)

Executive Summary

- **This document is the result of a suite of independent analyses investigating the consequences of different ways of weighting weir and aerial survey data in the run reconstruction model.** All of the authors have statistical training and experience with models like the run reconstruction model and voluntarily took on this task.
- **The run reconstruction model is a tool to estimate annual Chinook salmon run abundance in the Kuskokwim River since 1976.** The model uses observed data from weirs, aerial surveys, harvest, and total abundance estimates based on a mark-recapture study to estimate a historical run abundance time series. We conclude that the model structure is generally consistent with broadly applied and accepted stock assessment modeling approaches.
- **We investigated the results of 5 different models.** Four of the models were different approaches to weighting weirs and aerial surveys, and the fifth incorporated the 2014 mark-recapture abundance estimate:
    - ADF&G base model: individual weights are estimated for each escapement monitoring project.
    - Pooled weights model: a common weight is estimated for all weir projects and one common weight is estimated for all aerial survey projects.
    - No 2014 weirs model: 2014 weir counts are excluded from the ADF&G base model.
    - No 2014 aerial surveys model: 2014 aerial surveys are excluded from the ADF&G base model
    - With MRC model: the ADF&G model including the 2014 mark-recapture estimate.
- **All of the models resulted in the same major conclusion that the 2014 run was lower than average and the drainage-wide escapement goal was met**.
- **The estimated weights for weirs were higher than for the aerial surveys**, which means that, all else being equal, the model's predictions will more closely match the weir data overall.
- **Retrospective analyses showed that the ADF&G base model run estimates varied more than the pooled weights model as more data were added**, but neither model exhibited worrisome retrospective patterns.
- **The ADF&G base model exhibited an undesirable tendency to perfectly fit to one of the weir time series in some years**. This behavior is extreme and warrants further investigation and consideration of alternate models because it suggests that in those cases the model treated one of the weirs as a perfect indicator of escapement. This extreme behavior occurred in two out of eight years in the retrospective analysis: 2007 and 2010. Both of these years were not years in which the model was used to make management decisions.
- **Future modeling efforts should continue to investigate alternate weighting schemes and model structures**. The pooled weights model we constructed represents a sensible first step in this direction, but there are likely many other potential model structures that could be considered.

Introduction

During the Kuskokwim River Salmon Management Working Group session of the Kuskokwim Area Interagency Meeting held in Bethel, AK in March 2015 there were several important inquiries regarding the run reconstruction model. These questions included:

- How does it work?
- Which data have the most influence (i.e., weight)?
- Are there other ways of weighting the data components, and have these been considered?

The latter two questions have come up before with biologists and biometricians and were discussed prior to the Interagency Meeting. For example, some of the sensitivity analyses presented in this report have been previously conducted by the model developers and others using older versions of the model.

There are two primary objectives of this report. First is to answer the above questions and make clear why this issue of weighting has recently been raised. The second objective is to elaborate on issues related to data weighting by conducting several analyses to demonstrate how the data weighting works. This report will summarize and present the major findings from these analyses for working group and agency consideration.

The run reconstruction model is a tool that biologists and managers use to estimate how many Chinook salmon have been coming back to the river annually since the 1970s. The model was developed by Brian Bue and other assessment scientists and was completed in 2012 (Bue et al. 2012). The original report can be accessed at: http://www.adfg.alaska.gov/FedAidpdfs/FDS12-49. There have been several small changes to that model in 2014, as noted in this memorandum from Hamazaki and Liller (2015): http://www.adfg.alaska.gov/static/fishing/PDFs/commercial/kuskokwim/2014RunReconMemo.pdf.

How does the Kuskokwim River Chinook Salmon Run Reconstruction work?

To estimate run abundance, the model needs two major components. First, it needs data on how many fish are in the river each year. These data come primarily from weirs, aerial surveys, harvest estimates, and mark-recapture studies. Second, it needs some way to relate those data to actual numbers of fish, since we can't possibly count all the fish each year. The model represents a rational approach to combining information from all of the data types and sources to estimate how many Chinook salmon have been returning annually to the Kuskokwim River. The general model formulation and its assumptions are broadly consistent with commonly applied stock assessment modeling practices.

Weirs and Aerial Surveys

The weirs and aerial surveys are treated as indices of escapement, which means that they are not complete counts, but instead represent relative changes in escapement over time in each tributary. The model predicts how many fish should be observed at each weir or in each aerial survey by assuming escapement to each tributary is proportional to the drainage-wide escapement. These proportions are assumed to be constant over time and are estimated by the model (i.e., the data is used to inform the model about the tributary proportions). The model then compares how many fish were actually counted in each tributary to these predictions. If the predictions differ greatly from the data, the model adjusts the run abundance and proportions so that the predicted counts are closer to the actual counts that were observed.

Mark-Recapture Studies

How do we deal with tributaries that do not have weirs or aerial surveys? The majority of Chinook salmon producing tributaries are not monitored and if we want to know the total escapement each year, we need a way to expand the weir and aerial survey counts to the whole drainage. A mark-recapture study can provide information for such an approach. The mark-recapture study was an experiment in which biologists captured and tagged fish at Kalskag, then "recaptured" the tagged fish upstream at weirs later in the run. By comparing the number of marked versus unmarked fish observed at weirs, the biologists were able to estimate how many fish passed upstream of the tagging site. The methods for the 2003-2007 mark-recapture study and the resulting estimates are provided in Schaberg et al. (2012). The mark recapture estimates, along with observed harvests and escapement estimates based on drainage-area for streams downstream of the tagging site, are used in the model to scale up the weir and aerial survey counts to a total drainage-wide abundance of fish for a given year. One way to think about this is that the weirs and aerial surveys inform the ups and downs of how many fish there are from year to year, and the mark-recapture estimates specify the actual total number of fish, which allows for the model to estimate how many fish there are in the whole drainage without actually counting them all every year.

Commercial Catch and Effort

There is also a component that uses the historical catch and effort from the commercial fishery. We all know that when there are lots fish in the river, commercial fishers don't need to spend as much time catching fish as they would if there were fewer fish in the river. In years in which many fish were caught with little effort, the model predicts that there were many fish in the river that year. In recent years, this component has had minimal impact on the model since there is no directed commercial fishery for Chinook salmon. The component is still included because there are data from the 1970s and 1980s when the commercial component was more extensive that can provide information on total run abundance in those years.

End Result

Of course there are statistical complexities surrounding how the model adjusts its predictions to match the data we observed. Essentially, the model adjusts the run abundance estimates by trial-and-error until it minimizes the differences between the model predictions and the observed data across all data sources. The resulting run abundance estimates are those that make the data we observed for the past approximately 40 years most likely to have occurred. Another way to think about this is that the model's run abundance estimates are those that best explain all of our observations.

Which data are most influential in the model?

This question has to deal with an issue known as "data weighting". Data weighting means that the model is influenced by some data more than others. Another way to say this is that when two datasets disagree (i.e., one says there are lots of fish, another suggests there are few fish), which dataset does the model follow more closely?

If the model weighted all datasets the same, for example weirs and aerial surveys, then the model would tend towards the average abundance suggested by all of them. However, if the weirs are weighted more heavily, then the model will follow the weir patterns in escapement trends more closely than patterns suggested by the aerial surveys.

Ideally, weirs and aerial surveys would agree perfectly all the time. However, this is not the case since the two methods of counting fish are very different. An aerial survey is a pure index of escapement, that is, its counted value is related to the true escapement in that tributary, but it is not an attempt to be a complete count. Weirs are nearly complete counts, however there are factors that prevent them from being entirely complete counts (e.g., it was not possible for the weir to be functional during the entirety of the run). Because of these missed counts, weirs are also indices, but they are more representative of the true number of fish escaping to that tributary than are aerial surveys. This might suggest that we should have more trust in what the weirs tell us than what the aerial surveys tell us, or in other words, place more weight on weirs.

As the model is currently formulated, it is "self-weighting." That is, the model uses the data to determine which data sources provide the most consistent information about total run size and places the most weight on them. When working properly, this is very desirable since it removes the subjectivity of biologists, managers, and stakeholders deciding on which data are best and by how much. However, it can also cause the model to behave counterintuitively. For example, it is possible for the model to heavily weight a very small aerial survey that does not really reflect what is happening in the whole drainage, just because the data collected in that tributary are less variable through time than other tributaries. This is problematic when data sources are contradictory. In this case, one must decide which data to most believe: the consistent but small aerial survey or a larger one that we think is more representative of the total escapement? These types of scenarios inevitably must be confronted in nearly all stock assessment models.

Why is data weighting an issue in 2014?

In 2014, there was substantial disagreement among the various assessment projects relative to the implications of drainage-wide escapement size. In particular, the Kogrukluk, Kwethluk, and Tuluksak River weirs, along with the Cheeneetnuk and Holokuk River aerial surveys, indicated that the overall run was much smaller than suggested by the Pitka Fork of the Salmon, Kipchuk, and Aniak River aerial surveys. The disagreement among the estimates based on these individual assessment projects can be easily seen in Figure 3 of the Hamazaki and Liller (2015) memo. While disagreement among individual assessment projects is not unexpected, the level of disagreement among the 2014 assessment projects was relatively severe and is partially responsible for the high uncertainty in the 2014 estimate. The coefficient of variation (a relative measure of estimator uncertainty) of the 2014 run abundance estimate was 15% and was the largest coefficient of variation since 1996. Additionally, the model potentially down-weighted the information from the Kogrukluk and Kwethluk weirs more than was appropriate based on the relative size of these systems and the greater accuracy of weir data relative to aerial survey data (Table 1). In contrast, the model possibly placed too much weight on the relatively small Tatlawiksuk River information and some of the higher, but potentially less accurate, aerial survey counts. While it is impossible to know what the perfect data weighting scheme should be, this model behavior is troubling if we assume that the weirs on the larger systems are a more reliable indicator of drainage-wide escapement trends. This issue prompted us to investigate the behavior of how the different data sources are weighted in the model, particularly for weirs and aerial surveys.

Table 1.

Project Type	Tributary	Weighting Value	Lower 95% CI	Upper 95% CI
Weir	Kwethluk	5.81	2.61	12.93
	Tuluksak	5.00	2.48	10.09
	George	12.44	5.14	30.10
	Kogrukluk	8.36	4.26	16.41
	Tatlawiksuk	23.58	7.72	72.00
	Takotna	7.85	3.42	18.02
Aerial Survey	Kwethluk	3.07	1.33	7.07
	Kisaralik	1.48	0.84	2.61
	Tuluksak	3.37	1.47	7.73
	Salmon (Aniak)	2.92	1.74	4.92
	Kipchuk	4.22	2.23	8.00
	Aniak	8.80	3.86	20.06
	Holokuk	1.71	0.82	3.55
	Oskawalik	2.18	1.16	4.12
	Holitna	4.42	2.16	9.04
	Cheeneetnuk	3.27	1.67	6.40
	Gagaryah	4.26	2.18	8.34
	Pitka	3.58	1.50	8.56
	Bear	6.70	3.11	14.43
	Salmon (Pitka)	4.78	2.60	8.77

How did we investigate the weighting behavior?

We have used two approaches to investigate this question, as outlined below. These approaches are standard in stock assessments to diagnose model behavior and are often required before considering an assessment complete.

Sensitivity Analyses

A sensitivity analysis is a way of testing what impact a component of the model has on the output (e.g., the abundance estimates). Essentially, we change something about the model formulation and see how much the output changes. If the output changes a lot compared with the original model, then we infer that the component we changed has a large influence in the model. This exercise is important to allow us to see what would happen if we make different assumptions about how the model works. The idea is that when we cannot be perfectly certain of the best way to formulate the model, we can try several different variations of the model to see if they paint a different picture regarding run abundance. With regard to the question of how the model should weight the aerial survey data versus the weirs, for example, we can force the model to weight weirs and aerial surveys in a variety of different ways, and see how much it matters in terms of the run abundance estimates. This approach allows us to explore a range of possible models rather than having to decide on a single model when we lack an objective way to decide.

Retrospective Analyses

A retrospective analysis is an exercise in which we go back in time, so to speak. We pretend that we were back in 2007 and ignore all of the data collected between then and now and run the model. Then we include data from 2008 and run the model, then 2009, 2010, and so on up until the present. By doing this, we can look at how the model results change as we accumulate more and more data. If there are conflicts in the data, or the data weightings change over time then we might expect the model estimates to change more as we accumulate data. We are particularly interested in how the model chooses which data sources to weight over time and how much the weightings change when we add new data. This can help shed light on how the model selects which data are most important (i.e., aerial surveys versus weirs), which could indicate whether there are potential errors in the model estimates.

Methods

Sensitivity Analyses

We formulated five models that differed mainly in how weirs and aerial surveys were weighted and also included a model that incorporated the preliminary 2014 mark-recapture drainage-wide abundance estimate. Table 2 describes the different model structures and combinations included in the sensitivity analysis.

Table 2.

General Name	Description
Separate weights	The model estimates separate weights for each weir and aerial survey. This is the ADF&G base model.
Pooled weights	The model estimates a single weight for all weirs and a single weight for all aerial surveys
Without 2014 Weir Data	Run the model without 2014 weir project data, separate weights for all projects
Without 2014 Aerial Data	Run the model without 2014 aerial survey data, separate weights for all projects
With 2014 MRC (Mark-recapture)	Run the model with 2014 weir and aerial survey data and include the 2014 mark-recapture estimate (using the separated weights model)

Each of these scenarios represents a separate run of the model and each one was chosen strategically to illustrate how the weirs and aerial surveys impacted the model estimates. The first two scenarios were different model structures: the first allowed each individual project (i.e., individual weirs and aerial surveys) to have a different weight whereas the second forced all weirs to have the same weight and all aerial surveys to have the same weight. This approach was useful because if the model tended to track a particular aerial survey or weir count very closely at the expense of the others for no rational reason, then it may make sense to consider another weighting scheme such as pooling the weights by project type (i.e., weirs versus aerial surveys). This approach would prevent the model from picking one particular survey and following it too closely. We call this the "pooled weights model" for comparison with the ADF&G base model, which we call the "separate weights model."

A more extreme data weighting scheme that is often employed in sensitivity analyses is to assign zero weight to a particular dataset to assess the effect of its complete removal. With this approach, we can bound the extremes suggested by each type of data. We used this approach by removing (i.e., assigning a zero weight) the 2014 weir data and look at how much larger the aerial surveys suggest the 2014 run was. We then did the same for the aerial surveys. Finally, the last model scenario included the 2014 mark-recapture estimate, which is an independent estimate of drainage-wide abundance. Its inclusion allowed us to investigate which data were consistent with this estimate. Please note that the 2014 mark-recapture drainage-wide abundance estimate provided by ADF&G is preliminary and could change depending on the outcome of continued internal review by staff biometricians.

Retrospective Analysis

We conducted two separate retrospective analyses: one was conducted using the separated weights model and one used the pooled weight model. We conducted a separate retrospective analysis for each of these models to investigate the influence of the model formulation (i.e., how it dealt with weights: separate or pooled) on (1) which projects had the largest estimated weight overall and (2) whether or how much the weights changed as more data accumulated. We started the retrospective analysis in 2007 as this was the last year of the original mark-recapture studies (Schaberg et al. 2012). Starting after all of the original mark-recapture studies were completed allowed for all of the scaling information provided by those projects to be included in all of the retrospective model runs. Thus, any differences in the model estimates across subsequent model runs (i.e., as additional years of data are included) could be attributed mostly to how the model dealt with weir and aerial survey counts in those years.

All models were run using the model code provided in Hamazaki and Liller (2015). We made minor changes when warranted by the particular model scenario (e.g., removing particular data, pooling weights, etc.).

Results

When conducting these analyses, we discovered that we could not exactly reproduce the estimates reported by Hamazaki and Liller (2015). A detailed investigation of this phenomenon indicated that the parameter bounds in the ADF&G model (separate weights model) were too narrow to allow the model to freely estimate all of the parameters. Parameter bounds are simply upper and lower caps on parameter values (e.g., run abundance). They are commonly applied in stock assessments to prevent the models from inadvertently choosing implausible parameter values during the statistical trial-and-error estimation process. In the Kuskokwim Chinook model, the parameter bounds for the weights were not high enough such that one of the model weight estimates was pushed up against the bound, which means that the true estimate was likely outside the bound and suggests that the model was overly constrained by the bound. That is, the model was not allowed to estimate parameter values that fully satisfied the data. For our analyses, we loosened this constraint, which resulted in changes of 1-2% for the 2014 run estimate. For this reason, the estimates in this document do not exactly match those presented in the memo by Hamazaki and Liller (2015), but the difference is small.

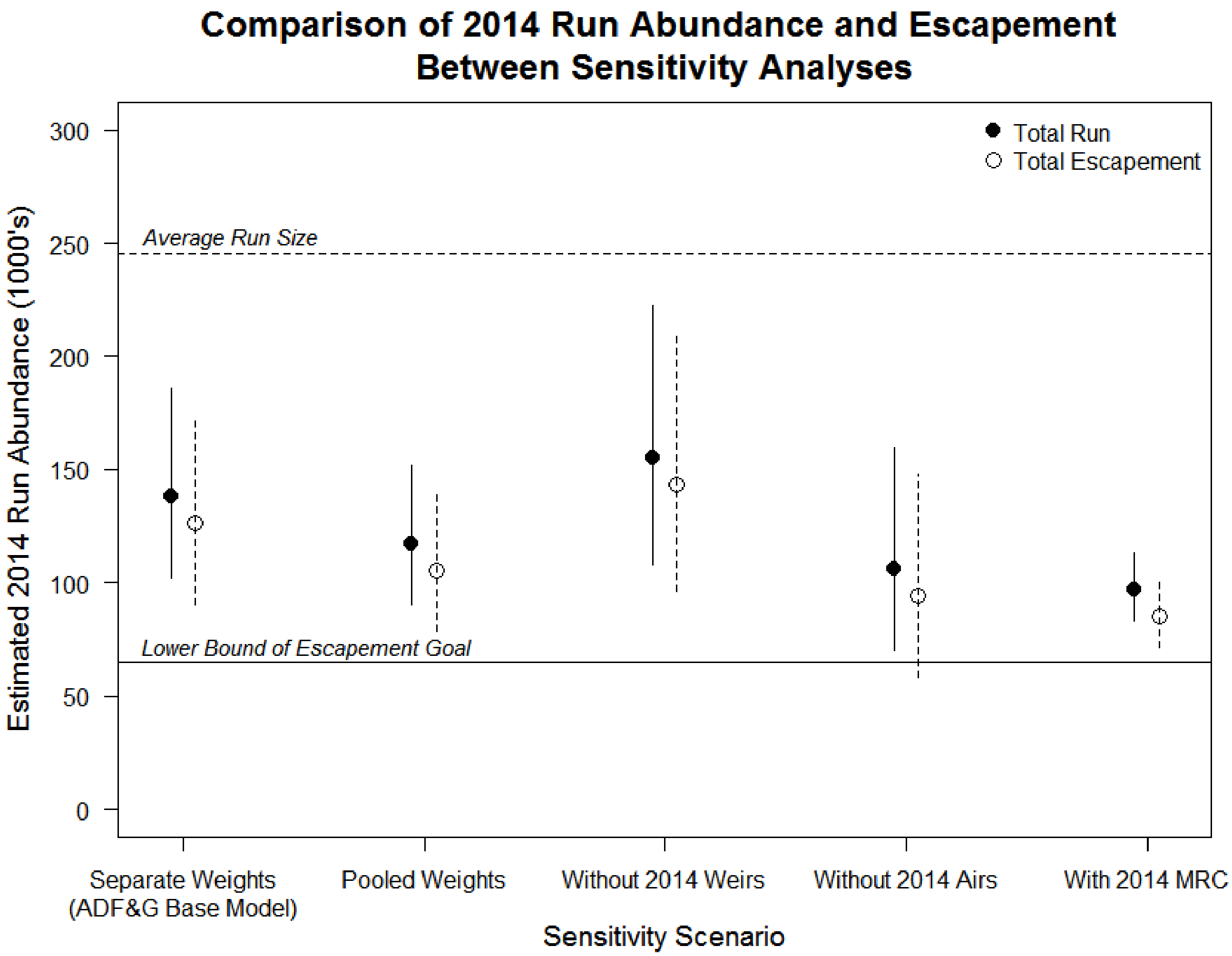
Additionally, when comparing our estimates to those presented in Hamazaki and Liller (2015), we discovered an error in how the estimates were reported. The order of the over-dispersion parameters (or weights, as they are referred to in this document) for the aerial surveys was inadvertently shuffled in the memo (see Table 2 in Hamazaki and Liller [2015]). For this reason, the weights that we report will differ from the memo. We have informed ADF&G of these issues.

Sensitivity Analyses

The primary measure we used to compare the different scenarios was the 2014 run abundance estimated by each model, as shown in Table 3 and Figure 1, below.

Table 3.	**Run Abundance**			**Escapement**		
Scenario	**Mean**	**Lower 95% CI**	**Upper 95% CI**	**Mean**	**Lower 95% CI**	**Upper 95% CI**
ADF&G Base Model (Separate Weights)	137,932	102,364	185,857	126,170	90,602	174,095
Pooled Weights	117,411	90,673	152,035	105,649	78,911	140,273
Without 2014 Weir Projects	154,962	108,046	222,250	143,200	96,284	210,488
Without 2014 Aerial Surveys	105,842	70,023	159,985	94,080	58,261	148,223
With 2014 MRC Estimate	97,087	83,423	112,989	85,325	71,661	101,227

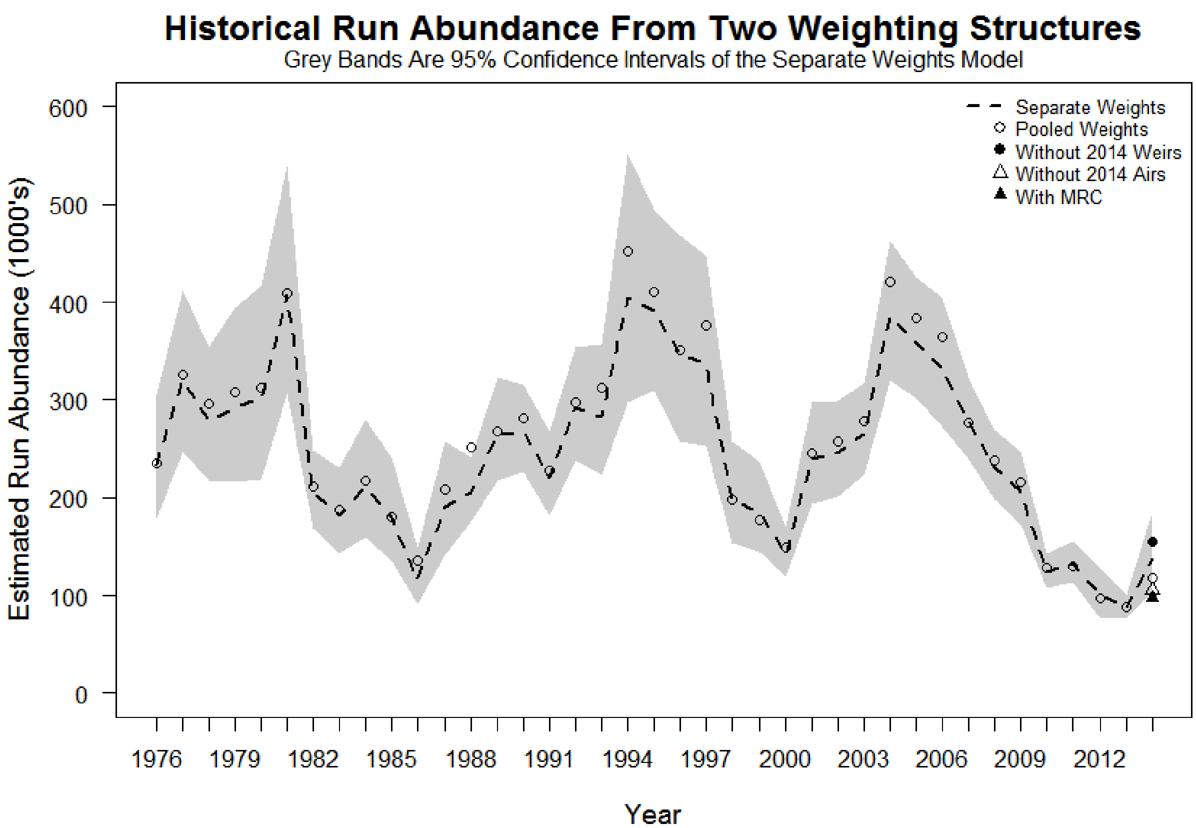
Figure 1.



The model structure clearly influenced the 2014 run estimate, but relative to the larger picture, these differences were small. The error bars are 95% confidence intervals, which is an estimate of uncertainty and represents how confident we are in the abundance estimate. The 95% confidence intervals on escapement were calculated by subtracting the total harvest estimate from the upper and lower bounds of the total run confidence interval, which ignores the uncertainty in the 2014 harvest estimate (which is considered to be minimal in this model). When we pooled the weights across project type (one weight for all the weirs and one weight for all the aerial surveys), the estimate was 15% smaller than when we estimated a different weight for each project. This finding resulted from the pooled model placing more weight on weirs that suggested a smaller run than did the aerial surveys in 2014. This model behavior was further demonstrated when we examined the next two points in Figure 1. When we excluded the 2014 weir data, the model predicted a higher 2014 run abundance and when we totally excluded the 2014 aerial survey data, the model predicted a lower 2014 run abundance. Finally, the model that considered the preliminary abundance estimate from the 2014 mark-recapture project produced the lowest estimated 2014 run abundance. Considering estimates across this set of sensitivity analyses provided a relatively complete characterization of the uncertainty in the run reconstruction among a wide set of options related to model structure and data considered. The important thing to keep in mind is that no matter which scenario you look at, the main conclusion is the same: the 2014 run was smaller than average and the drainage-wide escapement goal was met.

However, these findings pertain just to 2014. Since the model reconstructed the run abundance time series going back to 1976, it is important to look at how changing the model structure affected the whole time series. Figure 2 shows what happened when we compared the separated weights model (which is the ADF&G base model) to the pooled weights model:

Figure 2.

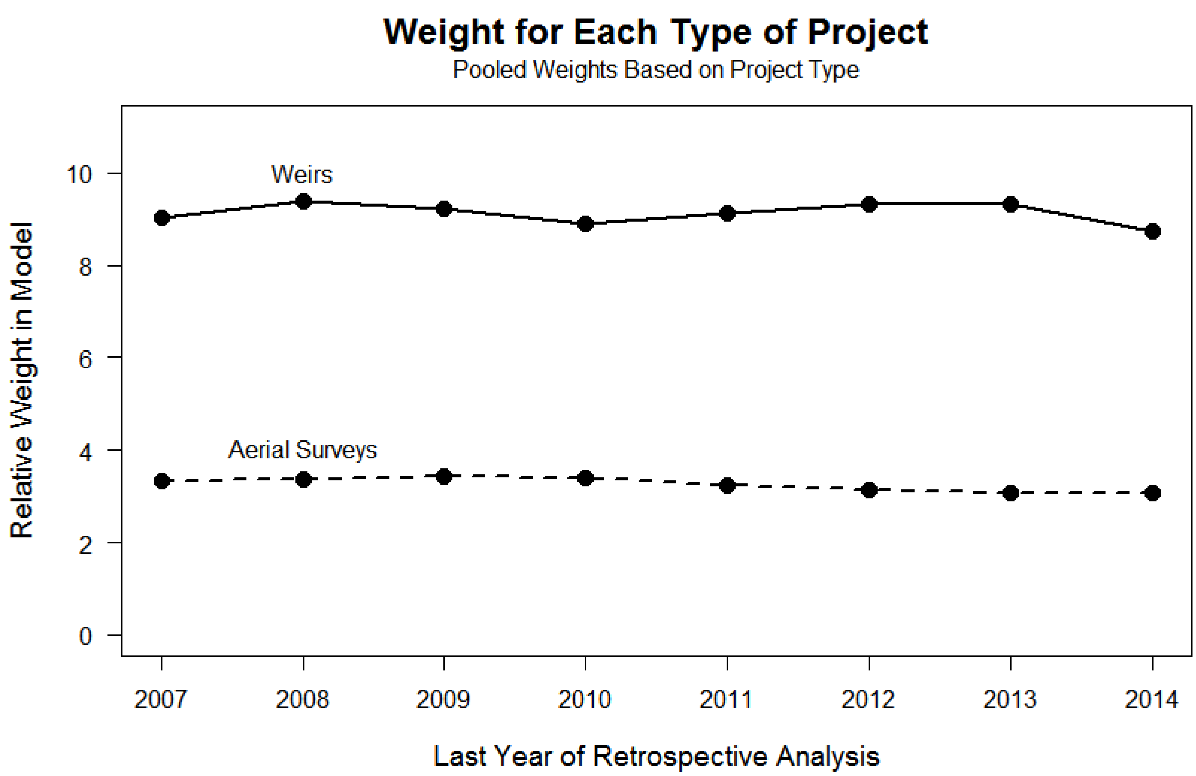


The run estimates from the pooled weights model fell within the 95% confidence intervals of the separated weights model nearly every year, with the exception of 1988. Based on this finding, we can conclude that the run estimates from these two models are similar. Several of the estimates from the other scenarios fall within this confidence interval as well. The only scenario that did not result in a 2014 run estimate within the confidence intervals of the ADF&G base scenario model (separated weights model) was the model with the 2014 mark recapture estimate included (solid triangle), but it was very close to being within the interval.

Retrospective Analyses

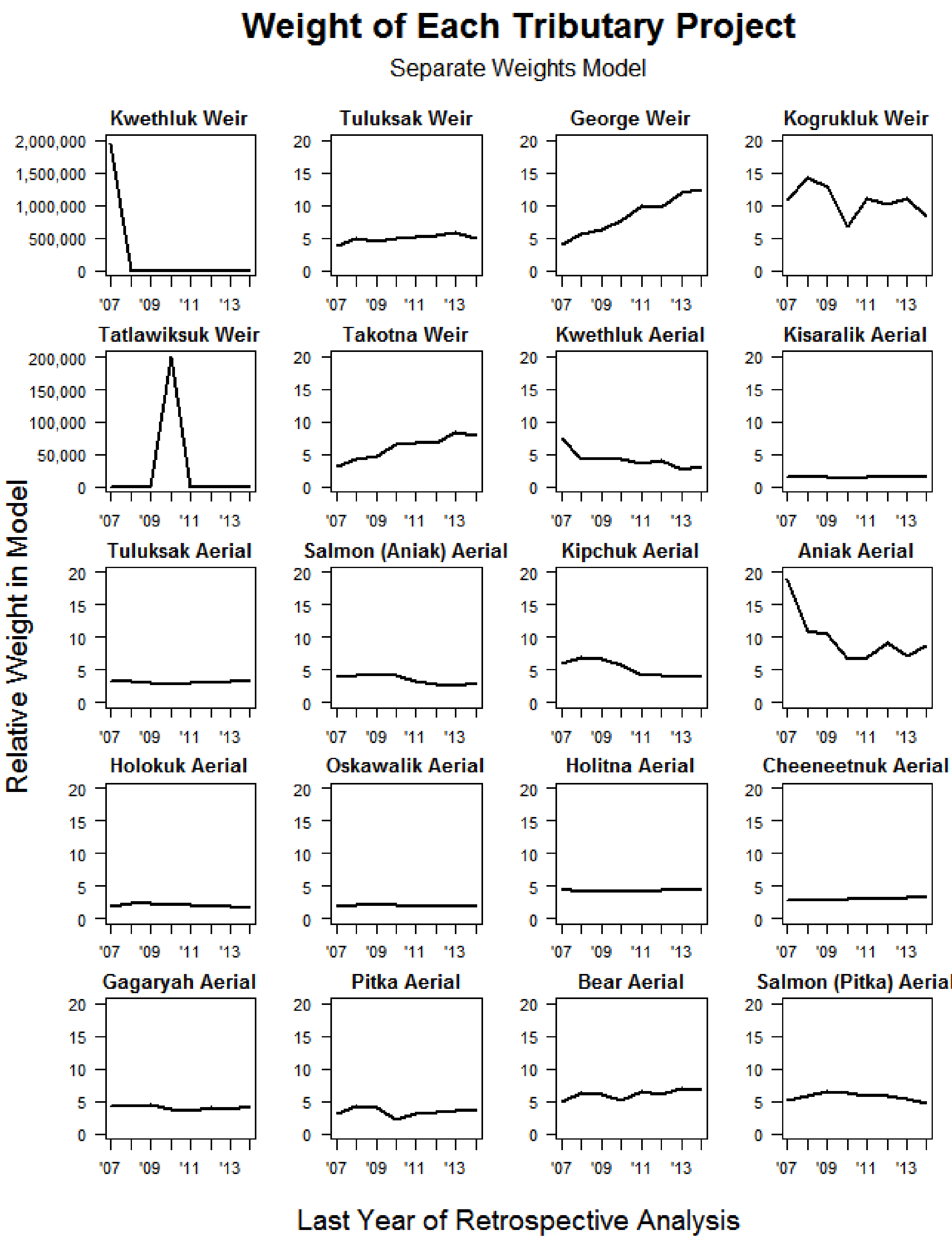
Figure 3 depicts how the weights for the two project types changed overtime as more data were added using the pooled weights model.

Figure 3.



Weir data were weighted more heavily than aerial survey data in the pooled weights model. The estimated weights from this model were consistent as more data were added. In contrast, estimated data weightings from the separate weights model varied drastically for some datasets (larger numbers mean it had more influence on the model; note the different magnitude of the axes for certain projects, Figure 4).

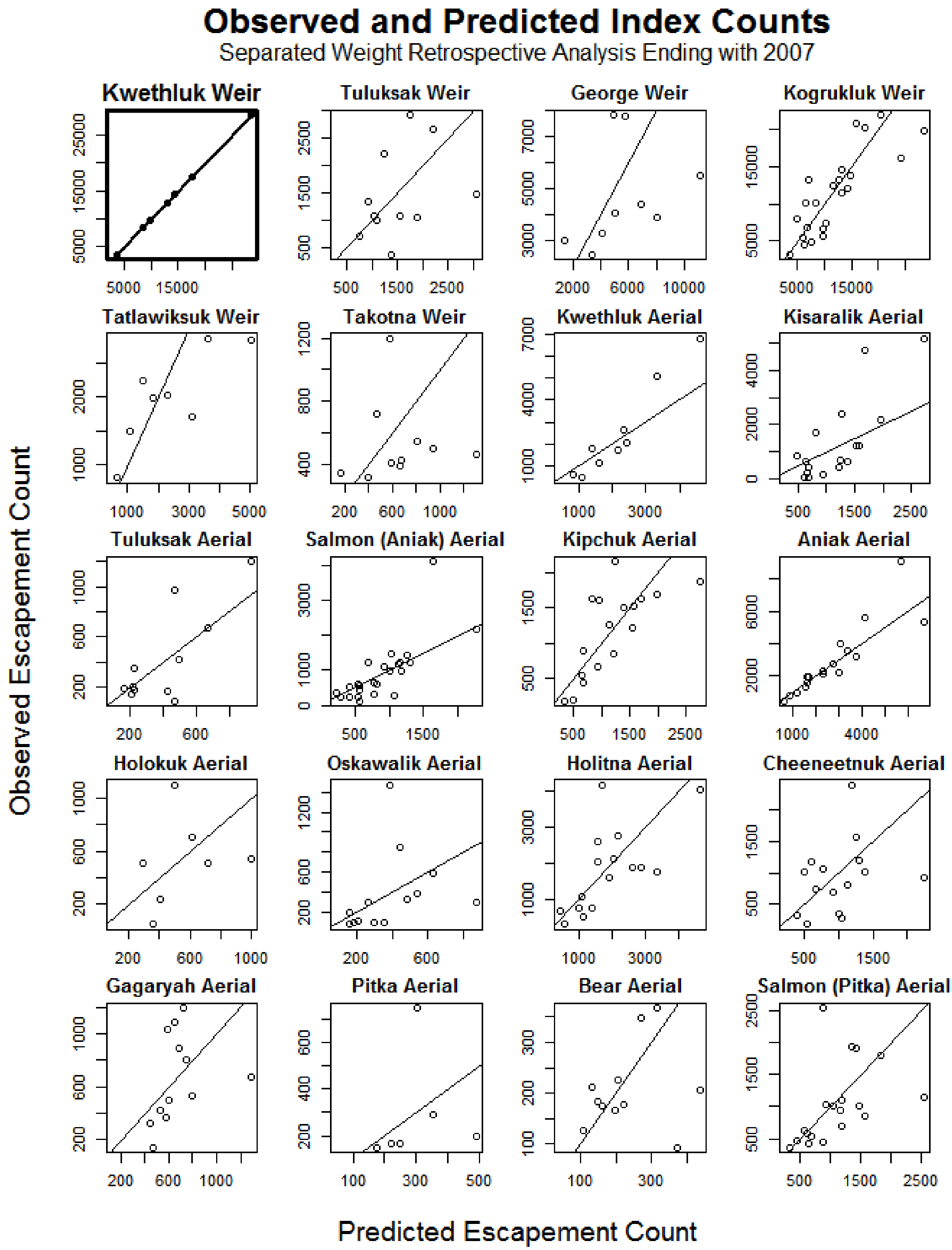
Figure 4.



Weight of Each Tributary Project
Separate Weights Model
Kwethluk Weir
Tuluksak Weir
George Weir
Kogrukluk Weir
Tatlawiksuk Weir
Takotna Weir
Kwethluk Aerial
Kisaralik Aerial
Tuluksak Aerial
Salmon (Aniak) Aerial
Kipchuk Aerial
Aniak Aerial
Holokuk Aerial
Oskawalik Aerial
Holitna Aerial
Cheeneetnuk Aerial
Gagaryah Aerial
Pitka Aerial
Bear Aerial
Salmon (Pitka) Aerial
2,000,000
1,500,000
1,000,000
500,000
200,000
150,000
100,000
50,000
0
5
10
15
20
'07
'09
'11
'13
Relative Weight in Model
Last Year of Retrospective Analysis

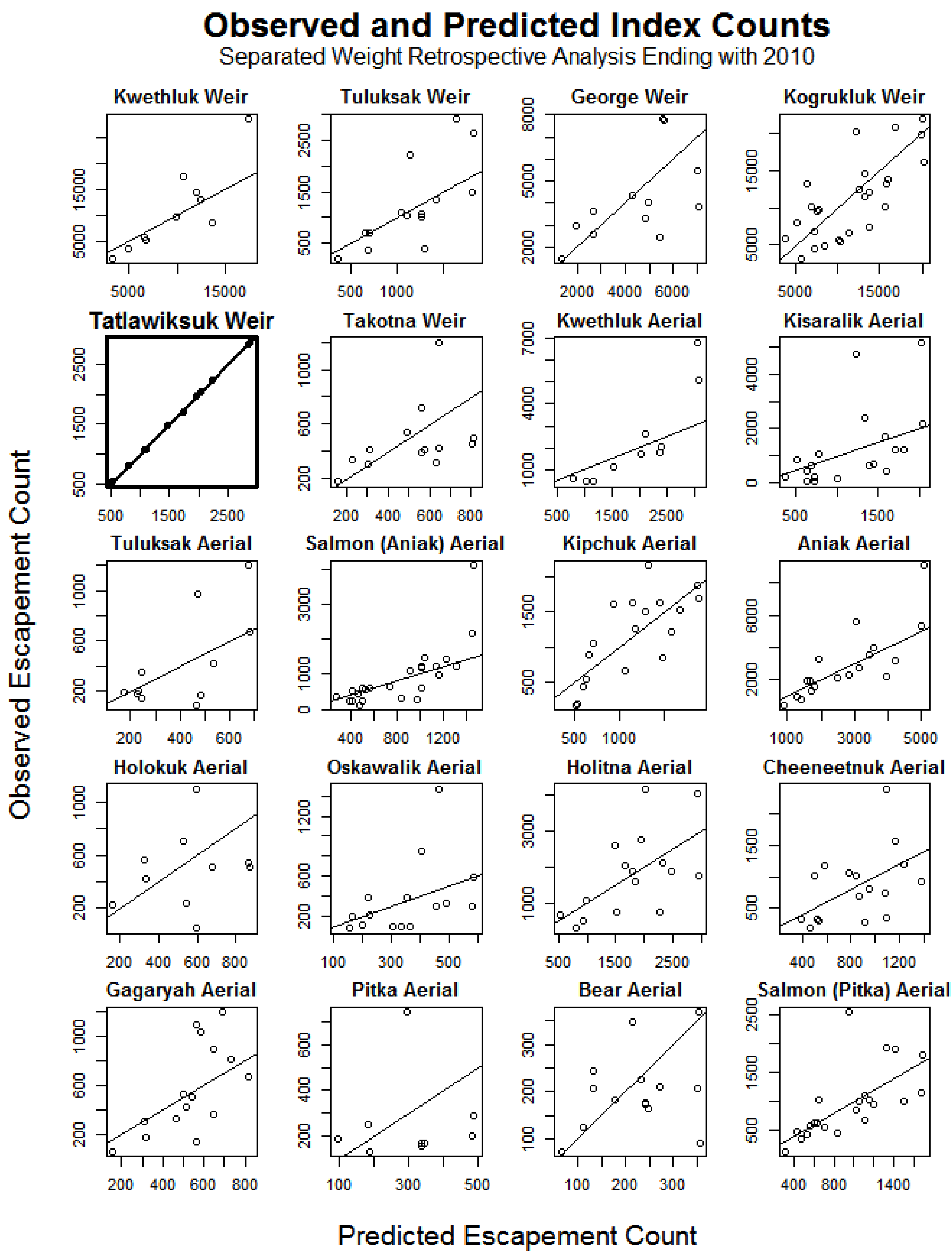
In particular, the model estimated extremely large weights for the Kwethluk weir in 2007 and Tatlawiksuk weir in 2010, for reasons that are not obvious. We investigated this behavior further by plotting the model predictions versus the observed data points for those years. Figure 5 shows observed and predicted escapement counts for the retrospective model run using data through 2007 when the Kwethluk weir received an extremely large weight. Figure 6 shows the same information but for the retrospective model run using data through 2010 when the Tatlawiksuk weir received an extremely large weight. The line on each graph depicts a model prediction that is exactly the same as the observed data, so points that fall closer to the line indicate that the model followed that dataset very closely for that year. Figures 5 and 6 indicate that the model fit the Kwethluk weir and Tatlawiksuk weir data perfectly in 2007 and 2010, respectively, because all of the data points fall exactly on the line. We were unable to identify the reason that the model adjusted its estimates so that they perfectly fit the observed data for those weirs in those model runs.

Figure 5.



Observed and Predicted Index Counts
Separated Weight Retrospective Analysis Ending with 2007
Observed Escapement Count
Predicted Escapement Count
Kwethluk Weir
Tuluksak Weir
George Weir
Kogrukluk Weir
Tatlawiksuk Weir
Takotna Weir
Kwethluk Aerial
Kisaralik Aerial
Tuluksak Aerial
Salmon (Aniak) Aerial
Kipchuk Aerial
Aniak Aerial
Holokuk Aerial
Oskawalik Aerial
Holitna Aerial
Cheeneetnuk Aerial
Gagaryah Aerial
Pitka Aerial
Bear Aerial
Salmon (Pitka) Aerial

Figure 6.

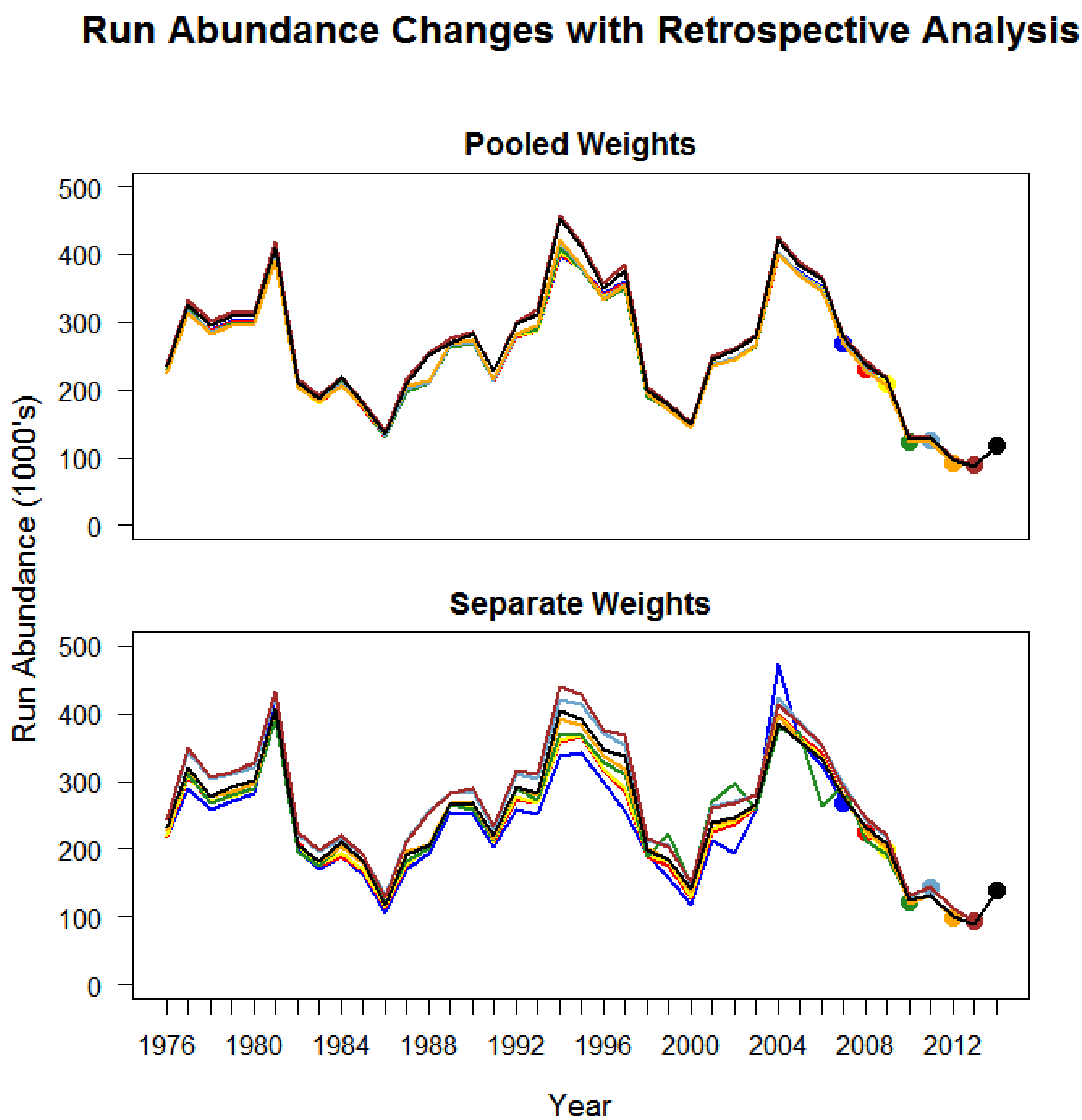


Observed and Predicted Index Counts
Separated Weight Retrospective Analysis Ending with 2010
Kwethluk Weir
Tuluksak Weir
George Weir
Kogrukluk Weir
Tatlawiksuk Weir
Takotna Weir
Kwethluk Aerial
Kisaralik Aerial
Tuluksak Aerial
Salmon (Aniak) Aerial
Kipchuk Aerial
Aniak Aerial
Holokuk Aerial
Oskawalik Aerial
Holitna Aerial
Cheeneetnuk Aerial
Gagaryah Aerial
Pitka Aerial
Bear Aerial
Salmon (Pitka) Aerial
Observed Escapement Count
Predicted Escapement Count

The separated weights model behavior of choosing one dataset and following it precisely at the expense of all other information is extreme and abnormal, especially when the chosen dataset changed as more years of data were considered. One way to think about this model behavior is that in 2007 for example, the model treated the Kwethluk weir as a perfect representation of escapement for the entire drainage, which strains the credibility of the separated model. The extreme weighting behavior in the 2007 and 2010 retrospective model runs indicates that there is currently insufficient information in the data to allow consistent estimation of a stable and sensible weighting scheme under the separated weights model. The retrospective analysis revealed that if this model had been used to produce estimates following either the 2007 or 2010 sampling seasons, it is likely that the analyst would have chosen some other method that resulted in a more justifiable set of data source weights.

In addition to the weights under the pooled and separated model structures, we also investigated how the estimated run abundance time series changed under these scenarios as we added more years of data, as shown in Figure 7.

Figure 7.



The run abundance time series from the separated weights model fluctuated more as we added data than did the pooled weights model. Although the overall trend was similar between models, the peak years in the early 1990s fluctuated by 10-20% (nearly 100,000 fish) between retrospective runs, whereas in the pooled model they fluctuated by 5-10% (~50,000 fish). Neither model demonstrated systematic retrospective bias that is commonly observed in many stock assessments. An example of a systematic bias would be if the abundance estimates consistently trended downward across the board with the inclusion of each additional year of data. Although both of these models behave reasonably well in this respect, the pool weights model clearly has less retrospective variation in the abundance estimates and does not suffer from the extreme tendency to perfectly fit one of the escapement surveys in some years.

Discussion

These analyses shed light on how the run reconstruction model assigns weights to individual datasets. In general, weirs were weighted more heavily than aerial surveys, which makes sense because we should expect the aerial surveys to have larger sampling error. When we look retrospectively, abundance estimates and data weights from the separated weights model fluctuated more as more data were added when compared to the pooled weights model. We have not yet figured out a way to quantify what a weight of about 9 for weirs versus a weight of about 3 for aerial surveys exactly means (how much more do weirs influence the model than aerial surveys?), but it is sufficient to say that an individual weir carries more weight in the model than does an aerial survey with an equivalent number of observations.

While there is still some question of whether it is better to separate the weights by tributary or to pool them by project type, our analyses suggest that a pooled weight model should be seriously considered. It is important to note that the pooled weight model is only one approach for dealing with the extreme and unsuitable "perfect-fitting" behavior of the separated weights model. Other weighting schemes that we did not investigate could potentially deal with this issue better than the pooled weights model. For example, it would be beneficial to explore whether tributary projects that count more of the total escapement should be weighted more heavily (tributaries with more fish get weighted more heavily). For the analyses presented in this document, our intent was to highlight the weighting behavior of the current model, not to investigate which (if any) alternate model is better. These are statistical issues and should be discussed and addressed eventually, but they are beyond the scope of this document.

One option for dealing with differences among model formulations would be to run multiple models each year and then average their results. Such a strategy would help to address uncertainty in which model is best. For example, there may be three models: one that separates the weighting by each tributary project, one that pools the weights by project type, and one that weights data sources by the contribution of that tributary to the total escapement. The results could then be averaged across these models, and this average could be used for the run abundance estimate. Under this approach, it would need to be decided whether each of the models would receive equal treatment, or if there are some formulations that are considered more plausible than others. There are well-established methods that have been developed to do these kinds of computations, if this approach were to be pursued in the future.

Acknowledgements

This work would not have been possible without the cooperation of ADF&G. In particular, we would like to thank Zach Liller and Toshihide "Hamachan" Hamazaki for their interest and support in these analyses. ADF&G was involved in the preliminary discussion that led to these analyses and provided model code, data, and the preliminary 2014 mark-recapture estimate. We would like to thank them for their support and assistance in these analyses. We, as independent analysts, and ADF&G collectively agree that analyses of this sort are critical in advancing our understanding and that they lead to overall better assessment techniques. While this document is the result of a set of independent analyses, it would not exist without the support of ADF&G and we look forward to continued collaboration in the future.

Literature Cited

Bue, B.G., K.L. Schaberg, Z.W. Liller, and D.B. Molyneaux. 2012. Estimates of the historic run and escapement for the Chinook salmon stock returning to the Kuskokwim River, 1976-2011. Alaska Departement of Fish and Game, Report to the Board of Fisheries, Anchorage. Available online at: http://www.adfg.alaska.gov/FedAidpdfs/FDS12-49

Hamazaki, T., and Z.W. Liller. 2015. 2014 Kuskokwim River Chinook Salmon Run Reconstruction and Model Revisions. Memorandum from the Alaska Department of Fish and Game, Division of Commercial Fisheries. Available online at: http://www.adfg.alaska.gov/static/fishing/PDFs/commercial/kuskokwim/2014RunReconMemo.pdf

Schaberg, K.L., Z.W. Liller, D.B. Molyneaux, B.G. Bue, and L. Stuby. 2012. Estimates of total annual return of Chinook salmon to the Kuskokwim River, 2002-2007. Alaska Department of Fish and Game, Fishery Data Series No. 12-36, Anchorage. Available online at: http://www.adfg.alaska.gov/FedAidPDFs/FDS12-36.pdf