Regional Information Report 5J12-23

Western Alaska Salmon Stock Identification Program Technical Document 16: Examining Prior Sensitivity using the Chum Salmon Baseline

by

Christopher Habicht,

William D. Templin,

and

James R. Jasper

November 2012

Alaska Department of Fish and Game Divisions of Sport Fish and Commercial Fisheries

Symbols and Abbreviations

The following symbols and abbreviations, and others approved for the Système International d'Unités (SI), are used without definition in the following reports by the Divisions of Sport Fish and of Commercial Fisheries: Fishery Manuscripts, Fishery Data Series Reports, Fishery Management Reports, Special Publications and the Division of Commercial Fisheries Regional Reports. All others, including deviations from definitions listed below, are noted in the text at first mention, as well as in the titles or footnotes of tables, and in figure or figure captions.

Weights and measures (metric)

centimeter	cm
deciliter	dL
gram	g
hectare	ha
kilogram	kg
kilometer	km
liter	L
meter	m
milliliter	mL
millimeter	mm

Weights and measures (English)

cubic feet per second	ft^3/s
foot	ft
gallon	gal
inch	in
mile	mi
nautical mile	nmi
ounce	oz
pound	lb
quart	qt
yard	yd

Time and temperature

day	d
degrees Celsius	°C
degrees Fahrenheit	°F
degrees kelvin	K
hour	h
minute	min
second	s

Physics and chemistry

all atomic symbols	
alternating current	AC
ampere	A
calorie	cal
direct current	DC
hertz	Hz
horsepower	hp
hydrogen ion activity (negative log of)	pH
parts per million	ppm
parts per thousand	ppt, ‰
volts	V
watts	W

General

Alaska Administrative Code	AAC
all commonly accepted abbreviations	e.g., Mr., Mrs., AM, PM, etc.
all commonly accepted professional titles	e.g., Dr., Ph.D., R.N., etc.
at	@
compass directions:	
east	E
north	N
south	S
west	W
copyright	©
corporate suffixes:	
Company	Co.
Corporation	Corp.
Incorporated	Inc.
Limited	Ltd.
District of Columbia	D.C.
et alii (and others)	et al.
et cetera (and so forth)	etc.
exempli gratia (for example)	e.g.
Federal Information Code	FIC
id est (that is)	i.e.
latitude or longitude	lat. or long.
monetary symbols (U.S.)	$, ¢
months (tables and figures): first three letters	Jan,...,Dec
registered trademark	®
trademark	™
United States (adjective)	U.S.
United States of America (noun)	USA
U.S.C.	United States Code
U.S. state	use two-letter abbreviations (e.g., AK, WA)

Mathematics, statistics

all standard mathematical signs, symbols and abbreviations	
alternate hypothesis	H_A
base of natural logarithm	*e*
catch per unit effort	CPUE
coefficient of variation	CV
common test statistics	(F, t, χ^2, etc.)
confidence interval	CI
correlation coefficient (multiple)	R
correlation coefficient (simple)	r
covariance	cov
degree (angular)	°
degrees of freedom	df
expected value	*E*
greater than	>
greater than or equal to	≥
harvest per unit effort	HPUE
less than	<
less than or equal to	≤
logarithm (natural)	ln
logarithm (base 10)	log
logarithm (specify base)	log_2, etc.
minute (angular)	'
not significant	NS
null hypothesis	H_O
percent	%
probability	P
probability of a type I error (rejection of the null hypothesis when true)	α
probability of a type II error (acceptance of the null hypothesis when false)	β
second (angular)	"
standard deviation	SD
standard error	SE
variance	
population	Var
sample	var

REGIONAL INFORMATION REPORT 5J12-23

WESTERN ALASKA SALMON STOCK IDENTIFICATION PROGRAM TECHNICAL DOCUMENT 16: EXAMINING PRIOR SENSITIVITY USING THE CHUM SALMON BASELINE

by
Christopher Habicht, William D. Templin, and James R. Jasper
Alaska Department of Fish and Game, Division of Commercial Fisheries, Gene Conservation Laboratory, Anchorage

Alaska Department of Fish and Game
Division of Sport Fish, Research and Technical Services
333 Raspberry Road, Anchorage, Alaska, 99518-1565

November 2012

The Regional Information Report Series was established in 1987 and was redefined in 2006 to meet the Division of Commercial Fisheries regional need for publishing and archiving information such as project operational plans, area management plans, budgetary information, staff comments and opinions to Board of Fisheries proposals, interim or preliminary data and grant agency reports, special meeting or minor workshop results and other regional information not generally reported elsewhere. Reports in this series may contain raw data and preliminary results. Reports in this series receive varying degrees of regional, biometric and editorial review; information in this series may be subsequently finalized and published in a different department reporting series or in the formal literature. Please contact the author or the Division of Commercial Fisheries if in doubt of the level of review or preliminary nature of the data reported. Regional Information Reports are available through the Alaska State Library and on the Internet at http://www.adfg.alaska.gov/sf/publications/.

Christopher Habicht, William D. Templin, James R. Jasper
Alaska Department of Fish and Game, Division of Commercial Fisheries, Gene Conservation Laboratory
333 Raspberry Road, Anchorage, Alaska, 99518-1565 USA

This document should be cited as:
Habicht, C., W. D. Templin, and J. R. Jasper. 2012. Western Alaska Salmon Stock Identification Program Technical Document 16: Prior sensitivity using the chum salmon baseline. Alaska Department of Fish and Game, Division of Commercial Fisheries, Regional Information Report 5J12-23, Anchorage.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Figure **Page**

ABSTRACT

Uncertainty about the magnitude, frequency, location, and timing of nonlocal harvest of sockeye salmon *Oncorhynchus nerka* and chum salmon *Oncorhynchus keta* in Western Alaska fisheries was the impetus for the Western Alaska Salmon Stock Identification Program (WASSIP). The project was designed to use genetic data in mixed stock analysis (MSA) to reduce this uncertainty. During a joint meeting of the Advisory Panel/Technical Committee to evaluate how best to establish priors (prior distributions) for both sockeye and chum salmon in Bayesian models used in MSA, the decision was made to use (a) internally-derived priors based on results from associated fishery strata and (b) a sequential-prior approach for remaining priors. The Advisory Panel requested a sensitivity analysis to examine the effect of different priors on the direction and magnitude of bias and magnitude of error in stock composition estimates. This document provides the results from 2 sets of sensitivity analyses run by the Gene Conservation Lab; the first used a mixture made up of fish from the Coastal Western Alaska reporting group and the second made up of fish from southern Alaska Peninsula reporting group. Four methods were used to derive priors: 1) regional-level method, 2) first and second sequential priors following the regional-level method, 3) uniform-binary method, and 4) first and second sequential priors following the uniform-binary method. Discrepancies from the truth were similar for priors based on the uniform-binary and the first sequential originating with the regional uniform methods. The smallest discrepancies from the truth were for priors based on the second sequential originating with the regional uniform prior method, or the first and second sequential priors originating with the binary uniform method. Informative priors, such as first and second sequential priors, provide the largest relative decrease in misallocations to reporting groups that are not represented in the mixture. For example, in Coastal Western Alaska, misallocation to the Northern District, Alaska Peninsula decreased from 2.96% to 0.17%, a 94% relative decrease when an informative prior was used. The effect of the initial prior through the sequential-prior process was quickly lost and we anticipate that the effect of the priors would be minimal in the initial strata and lost after the first sequential analysis because the associated fishery mixture estimates are likely to be more similar to the mixture under analysis than the regional-level or the binary uniform priors. These results support expending effort to develop appropriate priors and provide support for obtaining priors using methods analogous to those proposed for WASSIP, where information from other associated strata are used to inform the prior.

Key words: Western Alaska Salmon Stock Identification Program, WASSIP, mixed stock analysis, Bayesian analysis, initial prior, sockeye salmon, chum salmon, *Oncorhynchus nerka*, *Oncorhynchus keta*

INTRODUCTION

During the joint Advisory Panel (AP)/Technical Committee (TC) meeting held in Anchorage on September 21 and 22, 2011, Gene Conservation Laboratory presented options for establishing priors for both sockeye salmon (*Oncorhynchus nerka*) and chum salmon (*Oncorhynchus keta*) that are required for analyzing fishery mixtures using Bayesian methods. By the end of the meeting, we had consensus from the AP, pending final TC approval, to use a combination of internally-derived priors based on results from associated fishery strata for the first set of strata and a sequential-prior approach for the remaining priors (Appendix C in Jasper et al. 2012). However, during the discussions leading to this decision, the AP requested a sensitivity analysis to examine the effect of different priors on the direction and magnitude of bias and magnitude of error in stock composition estimates. Here we provide the results from this sensitivity analysis.

METHODS

Two test sets of chum salmon were used to test the sensitivity of estimates to the choice of priors, a set from Coastal Western Alaska (CWAK) and a set from Southern District, Alaska Peninsula, hereinafter referred to as South Peninsula. For each set, 400 fish were selected from populations assigned to the respective reporting group. These individuals were removed from the baseline and used as mixtures to test sensitivity to the choice of priors. Testing followed the methods used for the 100% proof tests outlined in Dann et al. (2012) except for the priors and sample sizes. The prior *sample size* was set to 1 fish.

In the first set of analyses, the first prior used was the regional-uniform prior (described in Jasper et al. 2012), where the prior for each region is set at the same weight (weight of each region equals 1 divided by the number of regions) and the priors within regions are distributed evenly across all the populations within that region (weight of each population equals weight of the region divided by the number of populations within that region). We then used sequential priors, first using the results from the regional uniform prior as the prior for the second analysis, and then using the results from this second analysis for the prior in the third analysis (Figure 1 and 2).

The second set of analyses used the uniform binary method described in the presentation to the AP/TC on September 21, 2011 (Appendix C in Jasper et al. 2012) followed by sequential priors. This uniform binary prior is used for the initial prior and is based on expert opinion. We used the expert opinion recommendations from the Department presented in Appendix C in Jasper et al. (2012). In this prior, of the G total number of reporting groups, $G^{(IN)}$ groups are deemed likely to contribute to a mixture and are tagged *IN* while $G^{(OUT)}$ groups are deemed unlikely to contribute significantly and are tagged *OUT*. The prior parameter value (α) assigned to the group proportions for each of these sets of reporting groups is: $\alpha_g = 0.01$ for $g \in G^{(OUT)}$ and $\alpha_{g'} = \frac{(1-0.01\times G^{(OUT)})}{G^{(IN)}}$ for $g' \in G^{(IN)}$. For CWAK the reporting groups tagged *IN* were: Asia, Kotzebue, CWAK, and upper Yukon River. For the South Peninsula, all reporting groups were tagged *IN*, so the analysis was identical to the regional-uniform prior. We then used sequential priors, using the results from the uniform-binary prior as the prior for the second analysis, and then using the results from this second analysis as the prior in the third analysis.

RESULTS

Results are shown in Figures 1 and 2. Regional-uniform priors provided the most downward biases for correct allocations and upward biases for incorrect allocations. Biases were similar for priors based on the binary uniform (D) and from the first sequential prior originating from the regional uniform (B). The smallest discrepancies from the truth were for priors based on the second sequential prior originating with the regional uniform (C) or the first and second sequential priors originating with the binary uniform (E and F). Misallocations were more pronounced in the CWAK tests (Figure 1) than in the South Peninsula tests (Figure 2). In the South Peninsula tests, the results from the 2 test sets were identical because the uniform-regional prior and the binary-uniform prior methods provided identical weights to all the reporting groups, simply because all stocks were deemed possibly present by the AP in this fishery.

DISCUSSION

These results support expending effort to develop appropriate priors. Informative priors provide the largest relative decrease in misallocations to reporting groups that are not represented in the mixture. For example, in CWAK, misallocation to the Northern District, Alaska Peninsula decreased from 2.96% to 0.17%, a 94% relative decrease when an informative prior was used. The higher misallocation was obtained using the regional-uniform prior, while the lower value was obtained using methods analogous to those proposed for WASSIP (Appendix C in Jasper et al. 2012), where information from other associated strata are used to inform the prior.

The effect of the prior through the sequential-prior process was quickly lost. In the CWAK tests, the effect of the initial prior was lost after the second sequential analysis (discrepancy < 0.1%), whereas for South Peninsula tests, the effects were gone after the first sequential analysis (discrepancy < 0.1%). By using the approach outlined in Appendix C of Jasper et al. 2012, we anticipate that the effect of the initial priors would be minimal in the initial strata and lost after the first sequential analysis because the associated fishery mixture estimates are likely to be more similar to the mixture under analysis than the binary uniform prior.

The consistent misallocations to the Upper Yukon River in tests of CWAK (Figure 1) were likely due to the artifact that one of the collections in the baseline (Jim River) was misassigned to the CWAK reporting group and should have been assigned to the Upper Yukon River group. This population is genetically similar to other Upper Yukon River collections and was the farthest upstream collection in the Yukon River assigned to the CWAK reporting group. This population has been reassigned into the Upper Yukon River reporting group for future analyses, so these apparent misallocations should become smaller if the analysis were to be repeated using the revised reporting groups.

The methods outlined in Appendix C of Jasper et al. (2012) should produce priors that substantially reduce discrepancies from the truth when allocating mixtures to reporting groups for chum salmon compared with the regional uniform prior. We anticipate an improvement for sockeye salmon as well, although it may not be as pronounced because sockeye salmon have deeper genetic differentiation than chum salmon. On the other hand, we have more reporting groups for sockeye salmon and small misallocations to many reporting groups will add to significant numbers of misallocated fish. Therefore, it seems prudent to invest in methods to minimize biases and errors by incorporating the most appropriate prior information.

QUESTIONS FOR TECHNICAL COMMITTEE

1) Given that this analysis was not designed to provide a comprehensive examination of the sensitivity to priors of estimating of stock composition estimates, do these methods and results provide enough information to conclude that an informative prior is better than an uninformative prior?
2) Are these methods appropriate to test the hypothesis that stock composition estimates are sensitive to the prior, at least in some cases?

TECHNICAL COMMITTEE REVIEW AND COMMENTS

This comment came by e-mail from Dr. Robin Waples on 10/4/11 with affirmation by Dr. Weir

This short document delivers what was proposed and shows that, indeed, informed priors can be very important in some cases. As expected, their importance is less when true stock differences are large and diminishes quickly under the sequential method. However, the strong effects for A1 indicate this issue is worth pursuing more.

ACKNOWLEDGEMENTS

The Technical Document series served as a record of communication between the Alaska Department of Fish and Game Commercial Fisheries Division and the Western Alaska Salmon Stock Identification Program (WASSIP) Technical Committee during the implementation of the program. The authors would like to thank the WASSIP Technical Committee and Advisory Panel for their constructive input on each of the documents throughout the project. The authors would also like to thank Erica Chenoweth who coordinated and prepared the Technical Document series for publication and Publication Specialists Amy Carroll and Joanne MacClellan for implementing the series into Regional Information Reports

REFERENCES CITED

Dann, T. H., A. Barclay and C. Habicht. 2012. Western Alaska Salmon Stock Identification Program Technical Document 5: Status of the SNP baseline for sockeye salmon. Alaska Department of Fish and Game, Division of Commercial Fisheries, Regional Information Report 5J12-10, Anchorage.

Jasper, J., S. Turner, and C. Habicht. 2012. Western Alaska Salmon Stock Identification Program Technical Document 13: Selection of a prior for mixed stock analysis. Alaska Department of Fish and Game, Division of Commercial Fisheries, Regional Information Report 5J12-20, Anchorage.

FIGURES

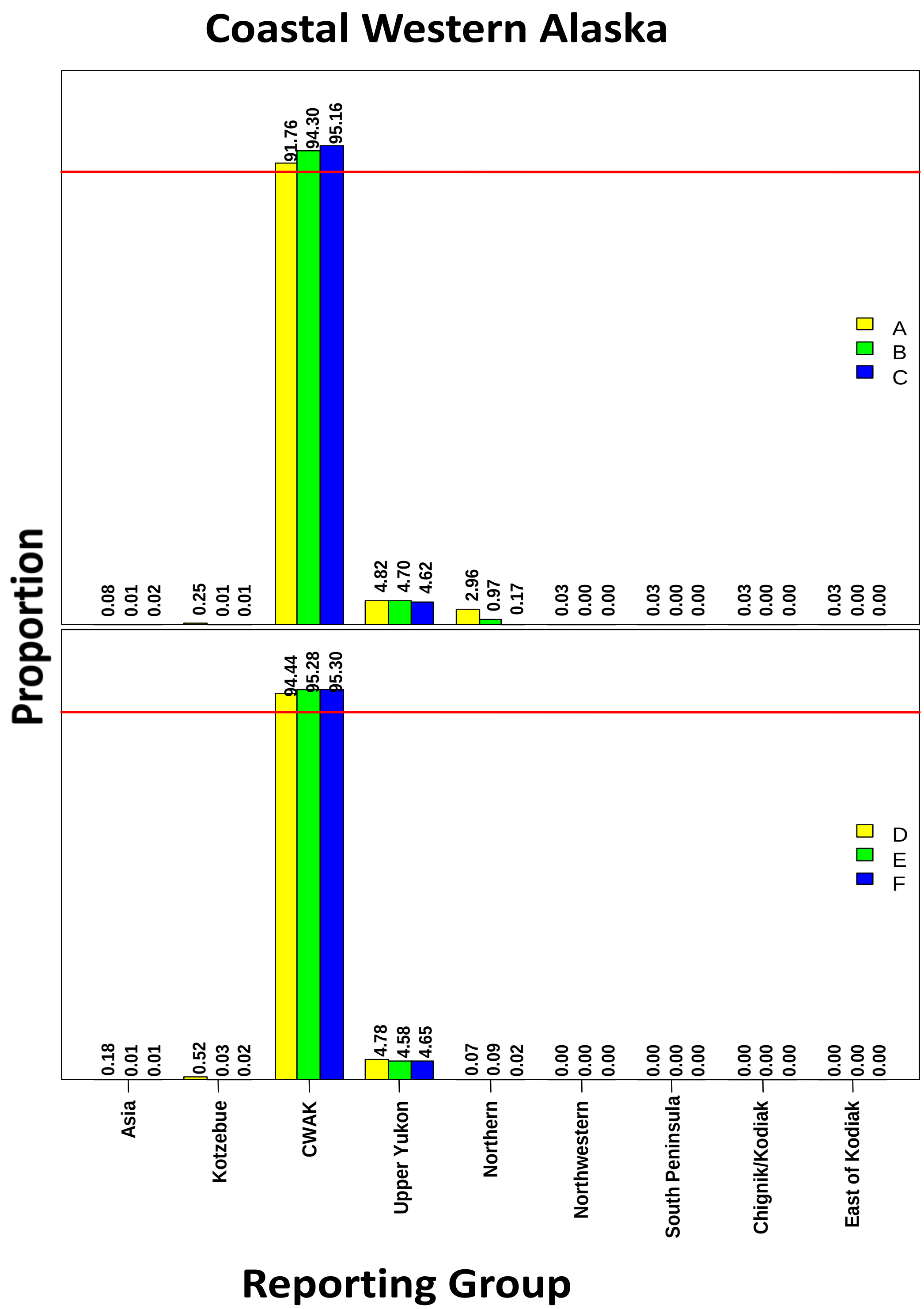

Figure 1.– Sensitivity analysis of priors for a mixture of chum salmon from baseline populations of the Coastal Western Alaska reporting group: A) regional-level prior, B and C) sequential priors following the regional-level prior, D) uniform-binary prior, E and F) sequential priors following the uniform-binary prior. The red horizontal line is at 90%.

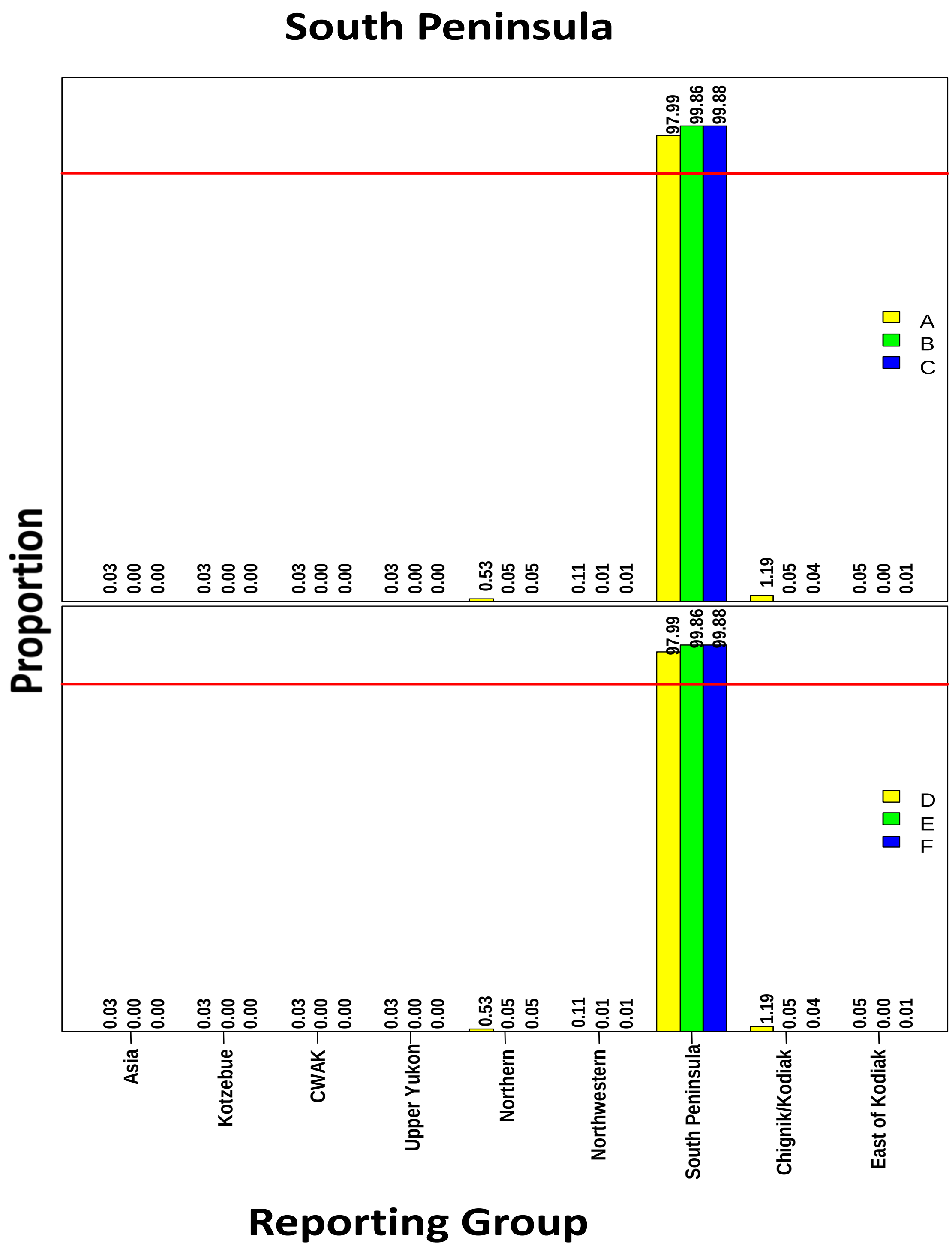

Figure 2.– Sensitivity analysis of priors for a mixture of chum salmon from baseline populations of the South Peninsula reporting group: A) regional-level prior, B and C) sequential priors following the regional-level prior, D) uniform-binary prior, E and F) sequential priors following the uniform-binary prior. The red horizontal line is at 90%.