



Fisheries and Oceans
Canada

Pêches et Océans
Canada

Ecosystems and
Oceans Science

Sciences des écosystèmes
et des océans

Canadian Science Advisory Secretariat (CSAS)

Proceedings Series 2021/037

Pacific Region

Proceedings of the Pacific Regional Peer Review on Run Timing and Diversion Rate Models for Fraser River Sockeye

**October 27-28, 2015
Nanaimo, British Columbia**

**Chairperson: Peter Chandler
Editor: Shelee Hamilton**

Fisheries and Oceans Canada
Science Branch
3190 Hammond Bay Road
Nanaimo, BC V9T 6N7

Foreword

The purpose of these Proceedings is to document the activities and key discussions of the meeting. The Proceedings may include research recommendations, uncertainties, and the rationale for decisions made during the meeting. Proceedings may also document when data, analyses or interpretations were reviewed and rejected on scientific grounds, including the reason(s) for rejection. As such, interpretations and opinions presented in this report individually may be factually incorrect or misleading, but are included to record as faithfully as possible what was considered at the meeting. No statements are to be taken as reflecting the conclusions of the meeting unless they are clearly identified as such. Moreover, further review may result in a change of conclusions where additional information was identified as relevant to the topics being considered, but not available in the timeframe of the meeting. In the rare case when there are formal dissenting views, these are also archived as Annexes to the Proceedings.

Published by:

Fisheries and Oceans Canada
Canadian Science Advisory Secretariat
200 Kent Street
Ottawa ON K1A 0E6

<http://www.dfo-mpo.gc.ca/csas-sccs/>
csas-sccs@dfo-mpo.gc.ca



© Her Majesty the Queen in Right of Canada, 2021
ISSN 1701-1280
ISBN 978-0-660-40227-7 Cat. No. Fs70-4/2021-037E-PDF

Correct citation for this publication:

DFO. 2021. Proceedings of the Pacific Regional Peer Review on Run Timing and Diversion Rate Models for Fraser River Sockeye; October 27-28, 2015. DFO Can. Sci. Advis. Sec. Proceed. Ser. 2021/037.

Aussi disponible en français :

MPO. 2021. Compte rendu de l'examen par les pairs de la région du Pacifique sur les modèles de période de montaison et du taux de déviation pour le saumon rouge du fleuve Fraser; le 27 et 28 octobre 2015. Secr. can. de consult. sci. du MPO. Compte rendu. 2021/037.

TABLE OF CONTENTS

SUMMARY	iv
INTRODUCTION	1
REVIEW	2
PRESENTATION OF WORKING PAPER	2
PRESENTATIONS OF WRITTEN REVIEWS.....	3
REVIEWER 1: MIKE LAPOINTE, PACIFIC SALMON COMMISSION	3
REVIEWER 2: ROBERT CONRAD, NORTHWEST INDIAN FISHERIES COMMISSION	3
REVIEWER 3: DAVE BLACKBOURN, RETIRED DFO RESEARCH SCIENTIST	4
GENERAL DISCUSSION	4
SCIENCE ADVISORY REPORT DISCUSSION	6
SOURCES OF UNCERTAINTY	6
RESULTS AND CONCLUSIONS.....	7
ADDITIONAL ADVICE TO MANAGEMENT	7
FUTURE WORK	7
NEXT STEPS.....	7
CONCLUSIONS.....	7
RECOMMENDATIONS & ADVICE	8
ACKNOWLEDGEMENTS	8
REFERENCES	9
APPENDIX A: TERMS OF REFERENCE.....	10
RUN TIMING AND DIVERSION RATE MODELS FOR FRASER RIVER SOCKEYE	10
APPENDIX B: AGENDA	12
APPENDIX C: PARTICIPANTS	14
APPENDIX D: WORKING PAPER REVIEWS.....	15
ROBERT CONRAD, NORTHWEST INDIAN FISHERIES COMMISSION U.S. CHAIR FRASER TECHNICAL COMMITTEE.....	15
DAVE BLACKBOURN, DFO RESEARCH SCIENTIST, RETIRED	17
MIKE LAPOINTE, PACIFIC SALMON COMMISSION.....	20

SUMMARY

These Proceedings summarize the relevant discussions and key conclusions that resulted from a Fisheries and Oceans Canada (DFO), Canadian Science Advisory Secretariat (CSAS) Regional Peer Review meeting on 27-28 October 2015 at the Pacific Biological Station in Nanaimo, B.C. A working paper assessing the performance of run timing and diversion rate forecast models for Fraser River Sockeye was presented for peer review.

In-person and web-based participation included Fisheries and Oceans Canada Science and Fisheries and Aquaculture Management Sectors staff; and external participants from First Nations organizations, the commercial and fishing sectors, and environmental non-governmental organizations.

The conclusions and advice resulting from this review will be provided in the form of a Science Advisory Report to DFO Fisheries Management, Pacific Salmon Commission and the Canada-US bilateral Fraser River Panel. The Science Advisory Report and supporting Research Document will be made publicly available on the [Canadian Science Advisory Secretariat](#) website.

INTRODUCTION

A Fisheries and Oceans Canada (DFO) Canadian Science Advisory Secretariat (CSAS), Regional Peer Review (RPR) meeting was held on October 27-28, 2015 at the Pacific Biological Station in Nanaimo to review an evaluation of models used to forecast Fraser River sockeye return timing and diversion rates.

The Terms of Reference (TOR) for the science review (Appendix A) were developed in response to a request for advice from DFO Fisheries Management. Notifications of the science review and conditions for participation were sent to representatives with relevant expertise from First Nations, Pacific Salmon Commission, US Fraser Technical Committee, commercial and recreational fishing sectors, environmental non-governmental organizations and academia.

The following working paper (WP) was prepared and made available to meeting participants prior to the meeting:

Evaluating Models to Forecast Fraser Sockeye Return Timing and Diversion Rate by Michael Folkes, Richard Thomson, and Roy Hourston (CSAP WP2013-SAL07).

The meeting Chair, Peter Chandler, welcomed participants both in the meeting room and online, reviewed the role of CSAS in the provision of peer-reviewed advice, and gave a general overview of the CSAS process. The Chair discussed the role of participants, the purpose of the various RPR publications (Science Advisory Report, Proceedings, and Research Document), and the definition and process around achieving consensus decisions and advice. Everyone was invited to participate fully in the discussion and to contribute knowledge to the process, with the goal of delivering scientifically defensible conclusions and advice. It was confirmed with participants that all had received copies of the Terms of Reference, working papers, and draft SAR.

The Chair reviewed the Agenda (Appendix B) and the Terms of Reference for the meeting, highlighting the objectives and identifying the Rapporteur for the review. The Chair then reviewed the ground rules and process for exchange, reminding participants that the meeting was a science review and not a consultation. The room was equipped with microphones to allow remote participation by web-based attendees, and in-person attendees were reminded to address comments and questions so they could be heard by those online.

Members were reminded that everyone at the meeting had equal standing as participants and that they were expected to contribute to the review process if they had information or questions relevant to the paper being discussed. In total, 27 people participated in the RPR (Appendix C). Shelee Hamilton was identified as the Rapporteur for the meeting.

Participants were informed that Bob Conrad (Northwest Indian Fisheries Commission), Mike Lapointe (Pacific Salmon Commission), and Dave Blackburn (DFO, retired) had been asked before the meeting to provide detailed written reviews for the working paper to assist everyone attending the peer-review meeting. Participants were provided with copies of the written reviews prior to the meeting (Appendix D).

The conclusions and advice resulting from this review will be provided in the form of a Science Advisory Report to Fisheries and Aquaculture management to inform salmon fishery planning for Fraser River sockeye. The Science Advisory Report and supporting Research Document will be made publicly available on the [Canadian Science Advisory Secretariat](#) (CSAS) website.

REVIEW

Working Paper: Evaluating Models to Forecast Fraser Sockeye Return Timing and Diversion Rate by Michael Folkes, Richard Thomson, and Roy Hourston. WP2013-SAL07

Rapporteur: Shelee Hamilton

Presenter(s): Michael Folkes

PRESENTATION OF WORKING PAPER

Management of the Fraser River sockeye fishery includes a pre-season planning component that relies on the forecast of variables such as adult migration run timing to local waters, and their migration route around Vancouver Island. This paper explored new statistical models that relate migratory patterns of returning adult Fraser River sockeye to potential environmental correlates. The authors presented the results from several software tools that they had developed to search North Pacific oceanic data for time series that were relevant to the migratory patterns of Fraser River sockeye salmon. Statistical models were used to examine the relationships between these time series and changes in Fraser River sockeye migration data, including performance testing of forecast precision, accuracy, and robustness to changes in the time series. Statistical models with high performance rankings will likely be suitable candidates to produce annual forecasts of Fraser sockeye migratory patterns that can be applied to both pre-season fishery planning models and (as Bayesian priors) to in-season run size estimation models. The performance analysis was broken into two approaches: retrospective and jackknifing. The results from each of the three analyses (Early Stuart timing, Chilko timing, and Fraser sockeye northern diversion rate) were described.

The generalized additive model (GAM) gave biologically unrealistic relationships between independent and dependent variables. For Early Stuart timing, all naïve models were consistently below the median rank and tended to be the worst performers. Multiple regression models were better than naïve models, and the top 50 models were all multivariate. The retrospective and jackknife top ten models were based on multivariate regression (non-North East Pacific Salmon Tracking and Research (NEPSTAR)). Offshore temperature and wind stress were shown to be secondary influences. NEPSTAR is the new physical oceanographic model and was used to provide near real-time estimates of current velocity in forecast models.

Unlike the Early Stuart results, all 14 NEPSTAR-mlr models for the Chilko were top ranked by retrospective performance. All naïve models ranked worse than the median rank. None of the models based on shore station sea surface salinity or Pacific decadal oscillation index met the threshold to be considered. A substantial number of the multivariate models included current velocity data based on the Ocean Surface Current Analysis Real Time (OSCAR) data series. Models based on three or more variables produce the lowest root mean squared error.

Forecast models of northern diversion rate (ND) are predominantly based on sea surface temperature (SST). When any month between September to May of the year prior to the adult return year indicates El Nino conditions based on the BEST index, there is significantly higher northern diversion than otherwise. This is not true for El Nino events from two years prior to return. A re-evaluation of the historical model relating ND to Fraser discharge and Tofino sea level reaffirmed that this relationship is not valid. Within the retrospective results naïve models ranked worse than the median rank. While, within the jackknife results, some naïve models ranked superior to median their performance was not adequate to warrant further consideration. Statistical models based on geomagnetic data did not meet the initial criteria of $R^2 > 0.5$, and therefore were not considered in the performance analysis. Offshore SST has a strong influence on northern diversion forecast models. Models based on the Pacific decadal oscillation index,

shore station sea surface temperature and shore station sea surface salinity did not meet the threshold and were not included in the performance analyses. The current velocity variables dominated the contribution to the NEPSTAR-mlrs. While one OSCAR based model passed the initial filter some of the neighbouring cells fell just below the minimum requirement.

Naïve models had the greatest uncertainty estimated by mean absolute error and root mean squared error. Models based on three or more variables maintained consistent ranking in both retrospective and jackknife analyses. Based on the ability of multivariate models to better forecast extreme events models with less than three variables were excluded from the final model selection step. It was accepted that similarly ranked models could have differing values of bias and uncertainty.

Tolerance curves, based on results of the performance analysis, are presented as the final step in model selection. These plots are intended to be an objective tool relating model forecast uncertainty and likelihood of that uncertainty. Thus model selection can be based on a manager's tolerance of uncertainty.

A sequential Bonferroni test was applied during the initial filtering of all single variable models. This approach applies a significance probability level that is more stringent given the large number of statistical models being concurrently evaluated.

PRESENTATIONS OF WRITTEN REVIEWS

REVIEWER 1: MIKE LAPOINTE, PACIFIC SALMON COMMISSION

The reviewer began by providing background information. If there is a high diversion rate, most fish go through Canadian waters, less through the U.S. It is important to know what fraction of the run you have seen has passed the reference point (i.e. the daily abundance versus the date fish passed through the Juan de Fuca Strait); knowing run timing helps to better calculate the estimate of total return. The reviewer then discussed the working paper. While the statement of purpose was to explore new statistical models, the paper actually evaluated current models in the quest to find a "silver bullet". The methods were appropriate but there are a number of variables that are model predictions themselves, which may be a concern. Also of possible concern is that almost all of it is based on test fishing data, which may be biased and less precise. The paper was very thorough but it was hard to understand what years were or weren't included in the models. The reviewer agreed that multicollinearity was not a concern as the authors demonstrated its absence. In conclusion, the paper was appropriately cautious but the reviewer questioned how it handles bias versus precision.

Issues that need clarification or improvement, identified by the reviewer, included:

- Add a table or modify an existing one to show which years are included and excluded in the analyses and models.
- Whether the top models meet the criteria Quinn (2005) uses as a conceptual test for hypothesis to be consistent predictors.

REVIEWER 2: ROBERT CONRAD, NORTHWEST INDIAN FISHERIES COMMISSION

The reviewer had two concerns with the working paper:

1. Multicollinearity in models with more than three covariates; there is an inverse relationship between the number of covariates and the number of useful years.

-
2. Early dismissal of geomagnetics covariate. It is a unique data set and has been useful in the past. Would like to have seen data in multiple linear regression models.

The reviewer had two comments:

1. A good job of showing data from El-Niño years, but could consider a more direct analysis and inclusion with the models.
2. Discussion of influence of stocks to the northern diversion rate very pertinent to US management.

The author responded saying the multicollinearity issue was dealt with in earlier conversations. With respect to model over-fitting, the retrospective evaluation should have exposed an over-fitted model. The best geomagnetic-based models didn't meet the $R^2 > 0.5$ selection criteria, and would have fallen short using Bonferroni-adjusted P-values as well. There is a need for a metric to determine outliers still under consideration. There is a need to know the priorities of fishery managers who may need a different model strategy or assumption. There is an unexamined assumption that the fish data are accurate. The impact on Northern diversion from stocks other than Fraser River sockeye was beyond the scope of the work. A participant suggested that a time-window approach could help identify extreme events (outliers). However, as pointed out by the author, single variable models are derived from variables with varying length, complicating the application of a common time window. In the case of predicting extreme events, the calculated root mean square error is sensitive to extreme events and reflects these occurrences.

REVIEWER 3: DAVE BLACKBOURN, RETIRED DFO RESEARCH SCIENTIST

The reviewer felt the paper was well written and thorough. He liked the extra explanation that was given on tables and figures. The conclusions and recommendations were optimistic given there is no detailed knowledge of the distribution or behaviour of any non-maturing Fraser sockeye in the North Pacific. The reviewer considered there was more information available in the history of the use of Kains Island SST data, and would like to see coastal current data used more in the report. There were no questions.

GENERAL DISCUSSION

The authors discussed comments that were raised with respect to the data and methods used. There was some concern that the data used were the output from other models. The authors clarified that these data are considered as estimates. The concern that the years of data available affect its utility is dealt with in the working paper. The issue of modelling data used in modelling similar issues was discussed and lead to an emphasis to keep multicollinearity to a minimum. The input variables are from common areas, so similar results from the models are expected. Non-NEPSTAR models were averaged over a gridded area to avoid risks of spurious correlations, and therefore more commonality among models can be expected; importantly NEPSTAR models were not averaged. This issue does not arise using NEPSTAR multivariate models. The authors debated how to limit the number of models tested, and ultimately decided to use trade off curves of the RMSE metric curve of probability and uncertainty. The mean raw error (MRE) column in the tables has values of 0.19, or 0.4 days which is considered minimal. The authors acknowledged the collinearity risk of using non-NEPSTAR models.

A reviewer stated that the text referring to Figures 2 and 6 should mention that the figures probably mostly represent fish from the Late (Shushwap L.) and Mid-Summer (Stuart L.) sub-populations. Early Stuart and Early Summer sockeye may have a slightly different ocean distribution from that shown.

A reviewer mentioned that previous studies ignore geomagnetics because:

1. Age data could tell if last year was the driver or if the first year was. Multiple ages come back at same time.
2. Juveniles go out one way, adults come back two ways.

One of the authors said that with respect to multicollinearity, the data sets are independent. While the Princeton Ocean Model had lower resolution more recent versions have higher resolution. It is expected that as the ocean currents change so may migration patterns. More recent years have a higher Northern Diversion rate, which may be influenced by the data available for comparison; test fishery data versus other catch methods. The effects of El-Niño years should be emphasized and may need further analysis. Future work may examine the underlying processes that correlate with changes in timing and migration (for example, is Kains Island SST an effective proxy for what is happening in the ocean that affecting fish?)

There was a discussion on the relevance of Area 20 migration timing. There have been recent observations where fish are holding at the mouth of the Fraser River; it may be too simplistic to assume it takes five days for fish to travel from Area 20 to the Fraser River. There are various summary estimates based on complex datasets, but there is rarely sufficient metadata and documentation to give insight to origins of the data, or any changes over the time period of data collection. It is a widespread problem that deserves more attention.

There was a discussion on the clarity of the methods section of the paper. Some reviewers wanted to see “trade-off plots” explained more thoroughly, and to clarify the meaning of the tolerance curves. It was suggested that the authors take a slice of the graph and do a trade-off with a bias metric, and highlight performance versus precision. Similarly it was suggested that more clarity about bias with respect to metrics would be helpful. The analysis assumes errors are normally distributed, and these may deviate at extremes. Figures 50 and 51 are different from the other plots in that they use all of the data. The basis of the performance metrics, and the relevance of bias, require consideration when making management decisions.

There was a question about the likelihood that the datasets will change, and whether the results are relevant for future conditions. The authors’ response was that the datasets are unlikely to be revised; if anything the resolution would improve. A larger concern is with non-stationary data, or data that are lost and cannot be retrieved in the future.

There was a brief discussion on using average or median naïve comparison. The authors showed that time series averages and the naïve models performed poorly, and it was unlikely that better information would be gained by their use. There was a concern that using the bootstrap method on the top 20 models might be favouring certain models, especially those with non-NEPSTAR input.

It was agreed that future work needs to consider differentiation and de-correlation of data (i.e. pseudo-replication from auto-correlated data) to avoid the tendency of a confirmation bias. The authors recommended getting forecast from the top 10 models and bootstrapping them together to get a common/combined forecast. Other ways to deal with this issue were discussed; one practical method would be to map strong correlations between non-NEPSTAR (SST and Oscar currents) and dependent variables from single variable analyses. If there are similarities, use the average over the larger region instead of individual cells; then combine the fine scale grids as an input variable. It was noted that non-NEPSTAR single variables used a grid to reduce the risk of spurious values. The method to select one single variable model to use in bootstrapping was discussed. When comparing multivariate models the variables may be from a similar area, but they are not necessarily the same variables, not all variables are shared between models.

The fact that the top six models share common variables raised suggestions that future work could consider how this impacts the model performance ratings for managers.

The outputs from models are considered as estimates, without a representation of error. The Princeton Ocean Model (POM) has been used extensively for the last 30 years, and the output is used without direct comparison to measured data due to the paucity of measured data.

The risk of model over-fitting was discussed. Based on a review of the literature the authors applied a rule of thumb for some models. The authors use of a retrospective analysis determined over-fitting was unlikely to be an issue. Authors were asked to add an explanation that the rules were developed after analysis but are still valid, and to address confusion about consistency around rules of thumb. Authors were also asked to elaborate on their use of the three variable limit, and why it was applied to non-NEPSTAR-mlr models but not NEPSTAR-mlr models. Additional comments included adding reference information to the sources of the data in table 2 (for example the years used for NEPSTAR data), and further discussion on the uncertainties with the output from current velocity models.

The authors were asked to provide more discussion of the diversion and timing estimates, and potential biases. It was considered necessary to more fully examine the dependent variable data (Northern Diversion and run timing) provided by the Pacific Salmon Commission including a description of how the data were collected, their meaning, and their associated uncertainties, biases and confidence levels.

There was consensus to accept the paper as presented with revisions as discussed during the meeting.

SCIENCE ADVISORY REPORT DISCUSSION

SOURCES OF UNCERTAINTY

- Dependent data (Northern Diversion and timing) as provided by the Pacific Salmon Commission are used without accuracy or precision information, and this uncertainty is not considered explicitly in this assessment.
- Accuracy and precision. There is a need to know what information is important to managers in order to determine which model is most suited to provide the best estimate, for example a model superior at forecasting whether the timing will be late or early may not do well at forecasting whether there will be a significant northern diversion. Performance measures were considered that quantify the accuracy and precision of all models equally. The influence of outliers depends on which models are used. In the absence of specific direction as to the relative importance of performance measures (precision and accuracy), the authors considered each of these factors equally in the ranking performance of alternative models. Results are presented in a probability framework.
- Future evaluation includes pre-analysis of spatial correlation of environmental variables to avoid confounding the interpretation of model predictions.
- Expert interpretation of the forecast is required to identify caveats associated with the forecast, and the ability to assign confidence limits.
- Knowledge gaps were addressed in the “future work” section, and are summarized in the Science Advice Report in the “Sources of Uncertainty” section.

RESULTS AND CONCLUSIONS

A list of top ranked models based on performance metrics was generated. However, there is still a need to develop a framework for managers that reflects how the models can be used to address various issues. The framework is based on the tolerance to uncertainty identified by management tolerance criteria. The decision of model choice is a trade-off between the tolerance criteria, resources, time, data availability, and the certainty required.

An analytical framework was developed and accepted to evaluate the performance of statistical models that relate the run timing (Early Stuart, Chilko) and northern diversion of Fraser River sockeye salmon with environmental factors including sea surface temperature, sea surface salinity, geomagnetic, near-surface ocean currents, and surface wind stress. TOR#1

The practical aspects of running some models raised some concerns. While some input variables are readily available others models are constrained by data availability. Some input information, like ocean modelling, is not presently within DFO's in-house capacity and requires contractor support. After acquiring the input data it may take weeks to run the model. Fundamentally, to be run on an annual basis NEPSTAR requires financial commitment. Non-NEPSTAR forecast models are based on free data so require in-house implementation. Most of the top ranked models are dependent on data that are currently publicly available but sourced from external agencies (e.g. NOAA).

ADDITIONAL ADVICE TO MANAGEMENT

It is possible to provide a table showing model ranks, northern diversion, early Stewart and Chilko timing (with tolerance curves), and the data resources (but not costs) required to generate this information.

FUTURE WORK

There is a need for a pre-analysis of the spatial correlation of environmental variables to avoid confounding model predictions. This is important to do before the pre-season forecasts. While de-collinearity was reduced by using step-wise regression there remains the need to reflect these effects in the model uncertainties.

There is potential for a new Research Document to examine the relationship between run timing and biophysical conceptual models. This analysis need not be limited to Fraser River stocks.

There was consensus that information on dependent data (northern diversion and timing) provided by the Pacific Salmon Commission needs to include more comprehensive documentation.

NEXT STEPS

There was a preference from managers for advice whether the run will be significantly early or late as opposed to off by a day or two. There is also the need for "rare event modelling", and how existing models can identify these events.

CONCLUSIONS

- The working paper was accepted with revisions.
- DFO Science will use this advice to provide annual forecasts of run timing and northern diversion.

-
- Over 150,000 models were evaluated to predict the northern diversion and run timing of Early Stuart and Chilko based on environmental variables.
 - The environmental factors assessed included single predictor variables and combinations of variables as input to multiple linear regression models. These were examined to determine which variables, geographical regions, time lags, and time-averaging periods revealed statistically significant relationships.
 - Many models provide similar results given the similar environmental factors used as independent variables.
 - An analytical framework was developed and accepted to evaluate the performance of statistical models that relate the run timing (Early Stuart, Chilko) and northern diversion of Fraser River sockeye salmon with environmental factors including sea surface temperature, sea surface salinity, geomagnetics, near-surface ocean currents, and surface wind stress (Objective #1).
 - The performance of the models are presented as (probability based) tolerance plots to guide managers and analysts in the selection of a model(s) used to forecast northern diversion and run timing (Objective #3).
 - A list of top ranked models was identified based on performance metrics (Objective #4).
 - Most of the top ranked models are dependent on data that are currently publicly available but sourced from external agencies (e.g. NOAA). Others are dependent on contractors (NEPSTAR) and require an annual financial commitment. These factors could constrain the ensemble of models that can be used in the forecast and may exclude the use of the top performing models. It should also be noted that each model requires varying input from DFO staff that has not been considered in this process (Objective #2).
 - Management tolerance criteria are required to identify an ensemble of top performing models to be used to produce the forecast.
 - The best-choice model to use is a trade-off between the time, data availability, financial cost, accuracy and precision of forecast.
 - This approach has advanced our ability to forecast northern diversion and run timing through the use of many environmental variables and the power of multiple model ensembles.

RECOMMENDATIONS & ADVICE

- It is recommended that future work examine the overlap between statistical analysis and biophysical models, using stocks in addition to Fraser River stocks, to provide an improved understanding of fish behavior.
- Information on dependent data (northern diversion and timing) provided by the Pacific Salmon Commission should include more comprehensive documentation.
- It is recommended that future work better prescribe model selection criteria, examine alternate models, and include rare event modelling.

ACKNOWLEDGEMENTS

The Chair would like to thank the reviewers, Mike Lapointe, Robert Conrad, and Dave Blackburn for their expertise and valuable review of the working paper, and all of the

participants for their constructive engagement in the science review process at this meeting. Thank you to Shelee Hamilton for being a rapporteur as well as writing the proceedings, and the CSAS office for assistance coordinating the meeting and producing final reports.

REFERENCES

Quinn, TP. 2005. The Behavior and Ecology of Pacific Salmon and Trout. University of Washington Press, Seattle. 378 p.

APPENDIX A: TERMS OF REFERENCE

RUN TIMING AND DIVERSION RATE MODELS FOR FRASER RIVER SOCKEYE

Regional Peer Review Process - Pacific Region

October 27-28, 2015

Nanaimo, BC

Chairperson: Peter Chandler

Context

Pre-season forecasts of adult Fraser Sockeye Salmon (*Oncorhynchus nerka*) marine run timing and diversion rate (the proportion of fish migrating through Johnstone Strait versus Juan de Fuca and Johnstone Straits combined) are essential for planning fisheries and are a Canadian responsibility under the Pacific Salmon Treaty (PST). Fisheries and Oceans Canada (DFO) Fisheries Management annually requests Science Branch to provide pre-season forecasts for marine run timing and diversion rate of Fraser River Sockeye Salmon stocks.

The performance of both statistical models currently used to generate the run timing and diversion rate forecasts has degraded during the last decade. Additionally, the oceanic variables used in the current timing model are themselves derived from an oceanographic model that is not domestically maintained and there is a risk that these data may not be available in the future. Recently, regionally developed and supported oceanographic models have been incorporated into new run timing and diversion rate models in attempts to improve performance of these models.

The objective of this review is to assess the performance of newly developed run timing and diversion rate models, including those that utilize near real-time oceanographic data. DFO Science will utilize advice arising from this Canadian Science Advisory Secretariat (CSAS) Regional Peer Review Process to provide Fisheries Management, Pacific Salmon Commission and the Canada-US bilateral Fraser River Panel with annual forecasts of Fraser River Sockeye Salmon run timing and diversion rate.

Objectives

The following working paper will be reviewed and provide the basis for discussion and advice on the specific objectives outlined below.

Michael Folkes, Richard Thomson, Roy Hourston. Evaluating Models to Forecast Fraser Sockeye Return Timing and Diversion Rate. CSAP Working Paper 2013SAL07

The specific objectives of this review are to:

1. Review statistical models developed to forecast the run timing of Early Stuart R. and Chilko R. Sockeye Salmon (four models considering six variables), and the diversion rate of the combined return Fraser Sockeye Salmon stocks (four models considering seven variables).

Models include naive time series models, single variable regression, generalized linear models, and generalized additive models.

The input oceanographic variables include sea surface temperature, salinity, and sea currents. The sea current data were independently derived from the NEPSTAR (Thomson et al. 2013), OSCAR (Bonjean and Lagerloef 2002), and OSCURS (Ingraham and Miyahara 1988) models.

The influence of a geomagnetic variable will also be evaluated for the diversion forecasts.

2. Describe the data inputs and characteristics (e.g. source and date available) of the different models.
3. Present relevant performance metrics related to bias and precision using both jackknife and retrospective approaches for each model.
4. Synthesize and compare performance between models across evaluation approaches and metrics.

Expected Publications

- Science Advisory Report
- Research Document
- Proceedings

Expected Participation

- Fisheries and Oceans Canada (DFO) Science Branch and Fisheries Management Branch
- Pacific Salmon Commission staff
- Canada-US bilateral Fraser River Panel members
- Fraser River Technical Committee members
- Province of BC
- Commercial and recreational fishing interests
- First Nations
- Non-government organizations
- Academia

References Cited

Bonjean, F., and G.S.E. Lagerloef, 2002: Diagnostic Model and Analysis of the Surface Currents in the Tropical Pacific Ocean, *Journal of Physical Oceanography*, Vol. 32, No. 10, pages 2938-2954.

Ingraham, J., and M.K. Miyahara. 1988. Ocean surface current simulations in the north Pacific ocean and Bering sea (OSCURS-numerical model). U.S. Dep. Commer. NOAA Tech. Memo. NMFS F/NWC-130: 155 p.

Thomson, R., R. Hourston, and S. Tinis. 2013. OSCURS for the 21st Century: Northeast Pacific Salmon Tracking and Research (NEPSTAR) Project, Year 3 Interim Report. Annual report submitted to the Pacific Salmon Commission. 37p. (request copy from csap@dfompo.gc.ca)

APPENDIX B: AGENDA

Canadian Science Advisory Secretariat
Centre for Science Advice Pacific

Regional Peer Review Meeting (RPR)

Evaluating Models to Forecast Fraser Sockeye Return Timing and Diversion Rate

27 – 28 October, 2015

Pacific Biological Station, Nanaimo

Chair: Peter Chandler

DAY 1 - Tuesday, October 27th

Time	Subject	Presenter
0900	Introductions Review Agenda & Housekeeping CSAS Overview and Procedures	Chair
0915	Review Terms of Reference	Chair
0930	Presentation of Working Paper	Authors
1030	Break	
1050	Overview Written Reviews	Chair + Reviewers & Authors
12:00	Lunch Break	
1300	Identification of Key Issues for Group Discussion	Group
1330	Discussion & Resolution of Technical Issues	RPR Participants
1445	Break	
1500	Discussion & Resolution of Results & Conclusions	RPR Participants
1645	Check in on progress and confirmation of topics for discussion on Day 2	RPR Participants
1700	Adjourn for the Day	

DAY 2 - Wednesday, October 28th

Time	Subject	Presenter
0900	Introductions Review Agenda & Housekeeping Review Status of Day 1	Chair
0915	Discussion & Resolution of Technical Issues (Continued from Day 1)	RPR Participants
1030	Break	
1045	Discussion and Resolution of Working Paper Conclusions	
1130	Develop Consensus on Paper Acceptability & Agreed-upon Revisions	RPR Participants
1200	Lunch Break	
1300	<i>Science Advisory Report (SAR)</i> Develop consensus on the following for inclusion: <ul style="list-style-type: none">• Sources of Uncertainty• Results & Conclusions• Additional advice to Management (as warranted)	RPR Participants
1430	Break	
1445	<i>Science Advisory Report (SAR)</i> (Continued)	RPR Participants
1630	Next Steps – Chair to review <ul style="list-style-type: none">• SAR review/approval process and timelines• Research Document & Proceedings timelines• Other follow-up or commitments (<i>as necessary</i>)	Chair
1645	Other Business arising from the review	Chair & Participants
1700	Adjourn meeting	

APPENDIX C: PARTICIPANTS

Last Name	First Name	Affiliation
Blackbourn	Dave	DFO Science, Retired
Campbell	Kelsey	Joint Technical Working Group (JTWG)
Chandler	Peter	DFO Ocean Sciences
Conrad	Bob	US Fraser Technical Committee Co-chair
Debertin	Allan	DFO Science
Folkes	Michael	DFO Science
Folkes	Shelee	DFO Science
Hargreaves	Marilyn	DFO Centre for Science Advice Pacific
Holt	Carrie	DFO Science
Hourston	Roy	DFO Ocean Sciences
Huang	Ann-Marie	DFO Fisheries Management
Hyatt	Kim	DFO Science
Irvine	Jim	DFO Science
Jantz	Les	DFO Fisheries Management
Lapointe	Mike	Pacific Salmon Commission
MacDonald	Bronwyn	DFO Science
Mundy	Peggy	National Marine Fisheries Service
Nicklin	Pete	Upper Fraser Fisheries Conservation Alliance/JTWG
Parken	Chuck	DFO Science
Patten	Bruce	DFO Science
Patterson	David	DFO Science
Pechter	Beth	DFO Science
Robinson	Kendra	DFO Science, Contractor
Scroggie	Jamie	DFO Fisheries Management Fraser
Staley	Mike	Fraser River Aboriginal Fisheries Sec./JTWG
Thomson	Richard	DFO Ocean Sciences
Tompkins	Arlene	DFO Science

APPENDIX D: WORKING PAPER REVIEWS

ROBERT CONRAD, NORTHWEST INDIAN FISHERIES COMMISSION U.S. CHAIR FRASER TECHNICAL COMMITTEE

This was a very well written and thorough report. In general, the documentation of methods and reporting of results were detailed and thorough. It provides a framework that should improve our ability to forecast Early Stuart and Chilko timing, and northern diversion rate, relative to current models.

A summary of my major concerns/comments follows:

- I have concerns about multi-collinearity and over-fitting with models using more than three covariates. Many of the highly ranked NEPSTAR models have more than three covariates. There is discussion of the dangers of over-fitting and guideline "rules of thumb" presented in both the Methods (page 51) and Conclusions and Recommendations section (page 148) but then the rules were generally not followed.
- I wish the geomagnetics had not been dismissed so early in the model selection process for northern diversion rate forecasts. This is a unique set of data that has shown promise in earlier research and is uniquely different from all the other environmental covariates being considered. Regardless of not meeting the initial screening criteria, it would have informative as to their value in forecasting diversion to include the geomagnetics as possible covariates during the mlr model process using the environmental covariates (maybe both NEPSTAR and non-NEPSTAR models).
- From a Panel perspective, models that accurately signal the outlier years (e.g., exceptionally early timing like Early Stuart in 2008, exceptionally late timing like Early Stuart and Chilko in 2005) are important. For the "top-rated" models, I would like to have seen a more in-depth evaluation (comparison of performance) of models that do well predicting these exceptions. Specifically, how do top-rated models that predict outliers well compare to top-rated models that don't predict outliers well for just the non-outlier years? This could be important information for the Panel when making a decision on models. If the two groups of models perform similarly for the non-outlier years then the choice is most likely the models that predict outliers well. However, if the models that predict outliers well do "worse" at predicting non-outlier years then there is a policy decision needed by the Panel on which risk they are most comfortable with (incurring more error during a rare event or more error during non-outlier years). This is obviously limited to the jackknife evaluations.
- The discussion of the influence of different stocks to the overall northern diversion rate being forecast in the Conclusions and Recommendations section (page 147) is very pertinent to U.S. management. Because the diversion rate is typically lower early in the season and increases throughout the season, and U.S. fishing opportunity is essentially not available when diversion is high (>80%), knowing what the diversion rate might be during the first half of the return is as critical as forecasting what the total diversion rate might be. From a U.S. perspective, having a forecast of expected northern diversion for that portion of the run returning prior to August 1 (e.g.) would be very valuable. I know the number of years with daily estimates of northern diversion is limited but I think there is enough for a retrospective analysis.

More detailed comments/questions follow with editorial comments provided at the end.

1. Page 45, Table 2: A "Temporal Resolution" column would be useful for this table so that the time resolution differences between some of the environmental variables is clear (daily,

weekly, monthly). Also, might be clearer to call the label under Forecast "Diversion Only" instead of just "Diversion".

2. Page 46, line 733: I would have liked to see a bit more of a discussion here on why "no statistical models based on El Niño events are included in the forecasting component of this document" since timing for Early Stuart and Chilko returns, and northern diversion rate, during El Niño years (based on the BEST index) appears to be different than for "other" years. Maybe this could be explored as a qualitative 0 (non-El Niño event) and 1 (El Niño event) model covariate.
3. Page 48, line 818: It would be useful to know the number of variables considered under each "type", e.g., (shore station SST and SSS (XX), NOAA OI SST (XX), current velocity, etc.).
4. Page 49, line 861: It is not clear to me how the Holm-Bonferroni adjustment was applied for single variable models (it seems more applicable to the multivariate regression models). Please explain.
5. Page 51, line 923: "non-NEPSTAR-MLR models were limited to three variables" while no such limitation was placed on NEPSTAR-MLR models. This seems rather arbitrary and I would have preferred the same three variable limit be placed on the NEPSTAR-MLR models. As I will refer to later, I have concerns about multi-collinearity and over-fitting with models using more than three of the NEPSTAR environmental covariates.
6. Page 51, lines 934-937: Not sure why these "rules of thumb" were presented since they are generally ignored for some of the NEPSTAR models presented later as belonging to the "top 10 performing models". Per my comment above, I would have preferred models relying on a more limited number of environmental covariates, which the rules of thumb would support. Given there were ≤ 30 data points, these rules would indicated models with 3 or fewer independent variables are desired.

Number of top ten models with 4 or more NEPSTAR covariates:

- Early Stuart timing - 1 (jackknife);
 - Chilko timing - 8 (retrospective), 5 (jackknife)
 - Northern Diversion - 6 (retrospective), 7 (jackknife)
1. Page 55: why is absolute value mean raw error (AMRE) even listed since it is commonly not used in most salmon forecast evaluations? This PM could be removed from all the pair plots in the Results section.
 2. Page 67: I find it worrying the four SST covariates from Bristol Bay in June of the return year were in models considered as "the top 50" for early Stuart timing. This just indicates how random correlation can occur when you examine such an a large set of covariates. This raises questions as to how many other of the "top 50" models include covariates with this same issue.
 3. However, I appreciate the discussion of these four variables in the Discussion section (lines 1984-1910).
 4. Page 85, tables 6 and 7: Because of the large number of models examined the model labels are somewhat cryptic. For these summary tables of the "best" performing models it would be very helpful to have a description of the covariates in the models presented and an indication of the model type (SLR, MLR, SCAM) for each of the models listed. E.g., mlr7 and mlr188, or nepstar13 and nepstar14, do not help me much; I realize I can go back in the

appendix tables and dig this out but that is challenging because they are not listed in any order that facilitates this. Having the descriptions presented maybe as a footnote to each table would be more useful.

5. Page 86, lines 1479-1482: I thought SCAM models were to be considered when the GAM models supplied "biologically unrealistic statistical fits". It seems that there should be some mention of the SCAM model here where all OI SST GAM models are summarily dismissed.
6. Page 86, line 1488: I think "Chilko" is meant here not "Early Stuart".
7. Page 86, line 1490: I think "Chilko" is meant here not "Early Stuart".
8. Page 92, line 1564: Unlike the authors, I find a correlation between model covariates of 0.79 ($R^2 = 62\%$) somewhat of a concern and raises the possibility of multi-collinearity issues. Even the models with correlation between model covariates > 0.60 trouble me.
9. Page 103, tables 8 and 9: Same comments as in #9 above.
10. Page 106, Geomagnetism: I wish the geomagnetism had not been dismissed so early in the model selection process. This is a unique set of data that has shown promise in earlier research and is uniquely different from all the other environmental covariates being considered. Regardless of not meeting the initial screening criteria, it would have informative as to their value in forecasting diversion to include the geomagnetism as possible covariates during the mlr model process using the environmental covariates.
11. Page 126, tables 10 and 11: Same comments as in #9 above.

Editorial Comments:

- Page 15, line 262: first word is "estimate", I believe it should be "forecast".
- Page 20, line 378: "...displayed lower than average..."
- Page 49, lines 855-856: This is an awkward sentence.
- Page 56, line 1120: a random "15" in this sentence. Maybe a footnote reference?
- Page 62, line 1240: not sure what "glsplpm" refers to?
- Page 100, line 1602: This sentence is missing something, doesn't make sense as is "Appreciably, associating..."
- Page 115, line 1754: missing closing paren after "Figure 67".
- Page 128, lines 1844-1846: Seems to be fragments from two different sentences here.

DAVE BLACKBOURN, DFO RESEARCH SCIENTIST, RETIRED

General Comments

I think the paper as a whole is a superbly written and astonishingly thorough exposition of the subject(s). As a general framework of the problems associated with these forecasts, and their solutions, I cannot see how it could be much improved. I particularly appreciate the extra explanations on the Figure and Table headings. I took Sections 4 and 5 largely as read, though found them very interesting.

However, I disagree with some of the statements in Sections 2 and 3, and found Section 6 occasionally confusing. Also, I thought Sections 7 and 8 interesting, but very optimistic.

Detailed Comments

p.5 Fig.1. What is the reference for this Figure? Official timing data now begins in 1980.

p.6 and Figs.2 and 6. Nearly all Fraser sockeye tags were recovered from fisheries in the 60's and in 1985, and probably mostly represented fish from the Late (Shushwap L.) and Mid-Summer (Stuart L.) sub-populations. Early Stuart and Early Summer sockeye *may* have a slightly different ocean distribution from that shown. We have no detailed knowledge of the distribution or behaviour of any non-maturing Fraser sockeye in the North Pacific (and see Section 6 –Discussion).

p.9 para 4. I thought that OSCURS data were first used a little earlier than 1998, but as my references are currently unavailable, I won't argue!

p.17 para 2. I understand the caution of the PSC staff for not making available their relatively short data set of the ND rate of early-timed sockeye. The data will certainly be eagerly studied when released, particularly in the United States.

p. 17 paras 3 and 4. I completely disagree with the 'history' of the use of Kains Id. SST as written here, and strongly recommend that it be rewritten. A perfectly adequate description of the use of those simple data as a forecast tool for Fraser Sockeye Diversion Rate was first published in 1984 (IPSFC Ann. Rept. 1983) accompanied by a graph and statistics of the relationship for various periods. This publication had a wide distribution and an avid readership among those interested in Fraser salmon. The 1984 publication was itself referenced in two books written in 1991 –(1) by J. Roos and (2) by Groot and Margolis. Thus, I contend that Kains Id. SST was used continuously, singly, or in combination, as a forecast tool for ND from about 1983 to 2014, or for over 30 years. That's a fairly long time for any fisheries management tool.

Pp 18 and 19. The discussion of the Vancouver Island Coastal Current and how it might be related to Kains Id. SST, and to ND is excellent, and I am surprised that this topic was not directly developed any further in the paper.

p. 23 para 1. In some unusual years, high ND does not follow from a northern landfall. In one or two years in the mid-1990's Fraser sockeye seemed to be blocked from entering Johnstone Strait.

p. 32 Fig. 23 The 2011 and 2012 cycle line series of ND look more continuous than the other two cycle lines (Summer Run dominance), in which there seems to be a more abrupt transition to higher average ND after about 1977.

P.39 The use of NEPSTAR POM data will obviously require expert assistance –see Sections 7 and 8.

p. 40 Fig. 28. I could not easily understand this figure heading.

pp. 45 and 46. Magnetism. A crucial unspoken assumption of this hypothesis is that Fraser sockeye *all* spend only two years at sea. Models including magnetism were not appraised in the performance analyses, but the topic may come up in Committee discussion.

p. 47. Prior work by M.F. on GAM fitting. What happened?

p. 47. Multicollinearity. Obviously this is important. Is it enough to say that "most series have low to no correlation—"?

pp 48/49. and p.109. Kains Id. SST does not quite meet the ND selection criterion of $R^2 = 0.50$ (0.49). Otherwise, its use makes great sense in terms of salmon migration!

p.51 etc. I think that this is a wonderful description of Performance Analysis and Model Selection – which of course leads me to ask more naïve questions! For example, on p.53 in para.2, I am still not sure if ‘non-stationarity’ (in ND) is a major source of ‘instability’ or not.

Results

p. 74. This discussion of the very late timing in 2005 might be better placed in the Discussion Section.

Figs.43, 44, 54, 55, 68 and 69. I found the forecasts shown here, and summarized from the two analytical methods, to be perhaps the most useful part of the paper .

p.109. ‘ND forecast-----last decade?’ –or last 30 years? See earlier comment.

p.109.para 5. I wish we had some idea of the biological significance of this apparently important variable.

p.126.Table 10. The heading is wrong and should not include ‘days’ or ‘ordinal date’.

p.128. para3. Suggest a change of wording from “restricting our”---- to “allowing our—to have broad spatial—“.

p.128 para.4 and pp.130, 131. See my earlier comment about p.6, and the fact that Early Stuart sockeye positions may not be well represented by tagging studies.

pp.130,131. The NEPSTAR-MLR models certainly seem to have more realistic biological significance for return timing than the non-NEPSTAR models, although there are different sets of years involved in the analyses.

p.132 I could not disentangle the arguments about the various forecast models for ND, but it is difficult for me to imagine a direct role for an ocean current variable from the second year, rather than the last year at sea (non-NEPSTAR models).

This whole section might benefit from more linkage with the Figures which show the direction of modelled ocean currents as well as their location.

6.6 and 8.0. Model Selection –and---Recommendations. The Authors outline a complex, lengthy, ideally iterative process whereby decision makers move from model selection to the assessment of risk in the outcome of forecasts for some Timing and ND. This in itself sounds an enormous task, especially when added to a similar process already underway for population – specific returning sockeye numbers.

If levels of future funding and staffing or the availability of some data sets and models prove to be less than desirable, particularly pre-season, I imagine that the PSC would appreciate that someone is also thinking of a slightly less complex Plan B, involving only some of the data and methods in this report.

8.1 Future Work

I hesitate to agree with more work than contained herein, but it would be very interesting for Biologists and Managers to see analyses made of Timing for late-returning stocks and of stock-specific ND rates.

MIKE LAPOINTE, PACIFIC SALMON COMMISSION

Statement of Purpose

The purpose “exploring new statistical models ... is both clearly stated and justified in the document. This is a document focused on exploration and evaluation of statistical models and therefore in my view should be reviewed in this context.

That said, while the document advances the science of statistical models in the forecasting of timing and diversion, the capacity of the paper to advance our abilities to generate accurate and robust forecasts of timing and diversion into the future is hampered by our ignorance of the mechanistic underpinnings for these statistical relationships that link changes in the environment to changes in fish behavior. In this context, I found myself wondering whether this paper was a ***hypothesis generating or testing exercise***? This is not intended to be a criticism of the authors, nor the considerable effort and excellent work reflected in this paper. Nor am I suggesting that the authors are guilty of ignoring a large body of data that exists on mechanisms. However, it is intended to be a strong call to action to ask what is needed to make significant advances on this issue. I will offer some less than half baked ideas about potential next steps at the additional areas of research section at end of my review, but among them will be a suggestion for a potential think tank, so that we can get a few better ideas closer to the oven.

Do data and methods in support of conclusions

The data and the methods are supportive of virtually all of the proposed operation model suggestions, though I do disagree with one item in the operational plan that pertains to quantification of uncertainty in the suite of best models (more details below in my comments on the discussion section.

I think the data and methods are appropriate with a few exceptions.

1. There are a number of X variables that are themselves model predictions. This left me wondering the extent to which these models that generated the X variables were validated against observed data. There also appear to be a few forms of modelled data: (a) data assimilation and model forecasts. Is there any reason to think of these model predicted values as less “trustworthy” than observed data? How might these predicted data lead us astray?
2. The authors should be aware of the strength and weakness in the data sets the models are predicting. I have outlined these below.

Data limitations

Suffice to say, data sets fit and used for retrospective and jackknife analyses in this paper are focused on a period when both timing and diversion rate come from test fisheries (see below). There is a further “split” in the data with respect to stock ID that was applied; pre-2001? is scale based, post-2000 largely microsatellite DNA. For Early Stuart, this distinction is less clear and early in the season, even now, it is not unusual to use scale based stock ID as a cost savings measure. Mixtures are simpler at that time of year, but important to note that some of marine migration is scale based. In the river (where reconstruction data come from) virtually all stock ID is DNA based, but low in-river catches can reduce sample sizes early in the season when Early Stuart fish are co-migrating with Chilliwack sockeye.

Timing

Chilko timing likely better estimated than Early Stuart. For both stocks timing prior to 1995 largely driven by commercial harvests. Post 1995(exception 1997), timing largely derived from

test fisheries. There are tradeoffs between accuracy and precision of these sources. Commercial harvest may have poor spatial temporal resolution (e.g. 6 days migration associated with JS harvest), but abundance (catch) well known. Test fishery data have much better spatial-temporal resolution, but are imprecise on a daily basis. Early Stuart data have the added source of error associated with low overall abundances – which creates additional imprecision. I think most of the Early Stuart time series post 1996 comes from Marine area data, but would need to confirm 1990's period in particular.

Diversification rate

Similar to timing, diversification rate is estimated largely from commercial data pre-1995. This makes the estimates in those earlier years much more accurate (e.g. see McKinnell 1997). Post 1995 (except 1997), the estimates are largely derived from test fisheries – which makes them considerably less precision and potentially biased, though the impact of this potential bias is likely greatest in situation where the ND is close to 50%. Recently, an area swept method for adjusting the relative catchability in purse seine test fisheries on each route has been applied retroactively back to 1987. This should improve the consistency of the abundance information derived from test fishery data, but not necessarily the accuracy. Because of the large number of daily CPUE data available in most years, the Central Limit Theorem applies – errors associated with the imprecision in daily estimates should average out in the cumulative CPUE used to estimate annual diversification rate. However, it is possible that the catchability on each route may vary systematically in some years – and deviate from the ratio associated with the area swept method applied to purse seines.

The methods appear to be appropriate, though I suppose a discussion of the rationale for the linear functional form as opposed to alternate forms could help. I suspect the fact that SCAM models did not prevail amongst the top models offers some support, but I note that the authors appear to allude to threshold type models (for E. Stuart) in the discussion.

Explanations of the data and methods explained in sufficient detail to properly evaluate the conclusions?

I generally found the methods section to be extremely thorough to the point that it is: (1) a good manual on model fitting, selection, diagnostics and validation, and (2) exhausting to read and follow. That said there are a few areas where I think the paper could be improved:

1. I found it hard to find what years were being included and excluded in which analyses and models. In particular, I think a table (or adding to an existing one) showing the years available for each Y and X variate would be helpful. I recognized that answering the question of which years in which models is a bit tricky, because of the fact that some lags eliminate some years, etc. but these impacts are mostly at the ends of the time series. I don't think the years included issue is a problem with the analysis, just hard to track.
2. Some commentary on potential influences of using predicted data for X variates mentioned above would be useful.
3. The authors have convinced me that **multicollinearity** is unlikely to be a problem in the classical sense, of incorrectly attributing more causal mechanisms to one variable relative to another within any one model. I furthermore do not believe there is a significant issue with respect to variables included in stepwise regression. I am less convinced though with respect to the Bonferroni (see 5 below)
4. The paper could benefit from an example of how the Holmes Bonferroni is used to establish the threshold alpha value against which the P values are being tested. Does each model in the hundreds of thousands tested have the same P value? What is meant by sequential in

this context? It seems to imply that the Bonferonni alpha varies depending on when the sequence a model is being tested.

5. There are obviously multiple comparisons in all the analyses (like 160,000-290,000 models). Bonferonni correction is the classical way to adjust alpha values in this circumstance, **but does the Holmes Bonferroni adjust for the fact that the data in the models in not independent? Is this an issue that readers should be concerned about? If not why not??** What I am concerned about is that I think the probability of incorrectly falsely concluding there is a significant relationship between two models that share correlated X variables (cross correlation across models, not within models) is not independent and I am just not sure how the Bonferroni correction that was applied deals with this situation. The question is whether there should be a further correction to account for the lack of independence of variables included in the multiple models being compared. This potential issue is outside my expertise, so I will leave its resolution to a qualified statistician.

Is advice for managers presented in a useable for and does that advice reflect uncertainty in the data and analyses.

The conclusions and recommendations section is appropriately cautious about offering a cookbook set of advice to be used by managers. There a number of caveats provided, most of them very good in my option. That said there is a suggestion that manager's use the Tolerance framework to identify a potential suite of "top" models and I would offer some cautions about that approach. I think it is good in that it puts the ball in the managers court – "If you can tell us your risk tolerance for error, we can identify the suite of models that will deliver that level of precision." But I think managers are also interested in how far off in terms of bias a particular set of models might be. As such I think there is a need for some guidance in this regard. In any given year will all the "top" models point in the same direction ?(i.e. be equally biased?). One way to visualize whether this might be an issue is to plot the predictions from the top 10 models on plots similar to the **Forecast plots Figs. 54 & 55 (p. 101 & 102; Chilko; and Figs. 43 & 44, p. 83 & 84. E. Stuart) –**

What do these plots look like if you plot the predictions for the 10 best models in each case – are these models scattered about the observation or consistent in each year – are they telling you something different – speaks to how to generate ensemble model – which set of models would be least biased on average.

The caution about likely non-stationarity is a good one and will always plague statistically derived forecast that lack fundamental mechanistic underpinnings. But I don't think this exercise has generated a suite of "independently derived forecasts" simply because the X data used in various models are correlated. In otherward, I am not convinced that this exercise has identified models that are necessarily less vulnerable to non-stationarity.

Lastly the procedure described in section 7 "Proposed Operational modelling implementation Scheme" proposed a method (resampling; third bullet) that will generate overly optimistic estimate of precision (intervals too narrow) because of the correlation amongst models in the X variable selected. Need to finesse a way around this – perhaps 3 or 4 "best" models that do not share X variates – or follow some other methods for selecting models to include similar to "ensemble" approach?

Additional areas of research

Back to basics.

Quinn's (2005) criteria that have to be met for hypotheses to be consistent predictions of homing ND. *Do all best models meet this conceptual test?*

Linking of timing and diversion.

Quinn's (2005) figure 2-10 for and this paper's Fig. 10 for Chilko and total diversion and Blackburn's 1987 displacement model – warmer SST, more northerly landfall, later timing and higher diversion rate. How to quantify and take advantage of this – stock specific diversion rate??

PSC staff have the capacity to, and “are in the process of”, generating both temporal patterns in diversion rate and diversion rate by stock information. We also have timing data sets for more stocks than Early Stuart and Chilko. How could/should such data be leveraged to further work on this area.

Model temporal and geographic dependency

These subsections are worthy of further examination for understanding next steps – some interesting patterns here.

I kept looking the maps for evidence in space and time consistent with the dogma of counter clockwise movement in GOA.

What do results suggest are the most important variables – are signs of coefficients consistent with likely impact i.e. SST positive northward current negative, eastward current positive?

What is behind the differences in relative importance/performance between NEPSTAR and non-NEPSTAR models across the timing and ND data sets?

Early Stuart relations are perplexing, Chilko makes more sense – need to look at other later times stocks???

I think a review of these results in some sort of workshop format might be helpful in further reducing the number of models to a plausible set that “makes sense” with the caveat that our knowledge is limited.

More detailed comments

These comments were provided to the authors separately. Some are on a edited pdf. Others in a list format.