NPAFC
Doc. 1717
Rev.

Evaluating Models to Forecast Salmon Dynamics

by

M.J.P. Folkes[1], R.E. Thomson[2], and R.A.S. Hourston[2]

[1] Fisheries and Oceans Canada, Pacific Biological Station
3190 Hammond Bay Road, Nanaimo, B.C., Canada V9T 6N7

[2] Fisheries and Oceans Canada, Institute of Ocean Sciences
9860 West Saanich Road, Sidney, B.C., Canada V8L 5T5

Submitted to the

NORTH PACIFIC ANADROMOUS FISH COMMISSION

by

Canada

April 2017

THIS PAPER MAY BE CITED IN THE FOLLOWING MANNER:
Folkes, M.J.P., R.E. Thomson, and R.A.S. Hourston. 2017. Evaluating models to forecast salmon dynamics. NPAFC Doc. 1717. 11 pp. Fisheries and Oceans Canada, Pacific Biological Station, and Fisheries and Oceans Canada, Institute of Ocean Sciences (Available at http://www.npafc.org).

Keywords: Sockeye Salmon, British Columbia, forecast, run timing, diversion rate

Abstract

Management of the Fraser River sockeye fishery includes a pre-season planning component that relies on the forecast of three variables that represent characteristics of the returning adult run: recruitment, migration timing to local waters, and migration entry route relative to Vancouver Island (as defined by the Northern Diversion Rate). In this paper, we evaluate the two components related to forecasting the homing migration of adult Fraser sockeye (return timing and diversion rate). We summarize key findings of this model search, performance analysis, and selection process.

Introduction

For at least two decades DFO Science staff has provided the Pacific Salmon Commission (PSC) with pre-season forecasts of Fraser sockeye migratory patterns that are to be used in pre-season fishery planning. The forecasts are based on statistical relationships between these migratory patterns and environmental variables that are assumed to play a role on the migratory behaviour of returning adult Fraser sockeye.

In the early period of migration forecasting, the return timing forecasts of the Early Stuart sockeye stock were moderately accurate, but their effectiveness has substantially declined in recent years. The Chilko stock timing forecast error has varied greatly throughout its history. It is now common to have forecasts that differ markedly from the post-season estimates.

Since 1981 there have been substantial improvements in the resolution of directly measured oceanic/environmental variables that are published and publicly accessible (Bonjean (2002), Reynolds (2002)). Satellite technology has allowed for better spatial and temporal resolution of oceanic variables, and near-real time access to these data is possible. Models to estimate both off shore and coastal current velocity have substantially improved during the last two decades, resulting in a much better representation of marine conditions during critical, seasonal transition periods. Finally, the software tools required to search these large data sets for robust statistical models has become freely available within the last decade (R Core Team. 2013). These changes inspired a review of potential statistical models and variables to forecast return timing and diversion (Folkes et al, 2017), which is summarized in this document.

We present an overview of the results from several software tools, developed by the authors, to search a collection of North Pacific oceanic time series for biologically relevant relationships to the migratory patterns of Fraser sockeye salmon. The relationships (estimated as statistical models) are evaluated by performance testing to appraise 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.

Data

Dependent Data (Timing and Northern Diversion)

Marine return timing of Fraser river sockeye salmon is now defined as the date when 50% of a stock has passed through a common point along the migration route en-route to their natal freshwater system for spawning. Data series of marine return timing estimates commenced in 1951 (Chilko stock) and 1953 (Early Stuart stock) and are generated by staff of the PSC. The timing data have undergone revision several times during recent decades (Jim Cave, PSC, 2011 Pers. Comm). However, estimates were always intended to reflect the date of peak timing referenced to outer Juan de Fuca Strait (see Gilhousen (1960)).

Since the 1970s Early Stuart run timing has appeared to trend toward later dates (Figure 1), but this change is not statistically significant. The variance in the timing signal was relatively low during the 1970s and 1980s, but has become much greater since the 1990s. To match the period of the environmental data series used as covariates, the marine timing series was limited to 1983–2012. During this period, there was no significant trend in timing. Chilko timing since the 1960s shows a statistically significant trend in timing to later dates (Figure 2). The slope averages 2.4 days later per decade. However, there is no significant trend in timing when considering just the years to be used in the present evaluation (1983–2012).

The northern diversion (ND) rate is also estimated by staff of the PSC. Similar to methods used for timing, run-reconstruction techniques were also used historically to estimate ND. Annual estimates of the ND are calculated as the ratio of the annual total abundance of Fraser sockeye migrating via Johnstone Strait (Figure 3) divided by the annual total abundance of Fraser sockeye migrating via both approaches, where the reconstructed daily abundances are summed along all days in each year.

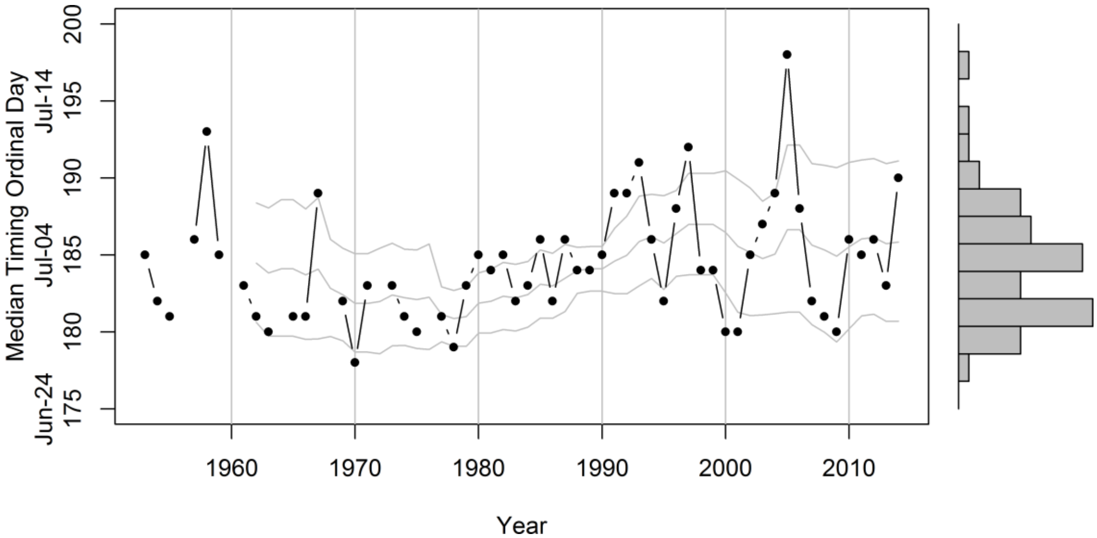

Figure 1: Time series of post-season estimated Early Stuart median arrival date to DFO statistical area 20. The y-axis has both calendar and ordinal date for comparison with other data and plots. The grey lines are the ten year running averages of median and standard deviation (SD), which were calculated in log-space. The histogram is scaled to density (fraction of the total numbers of occurrences).

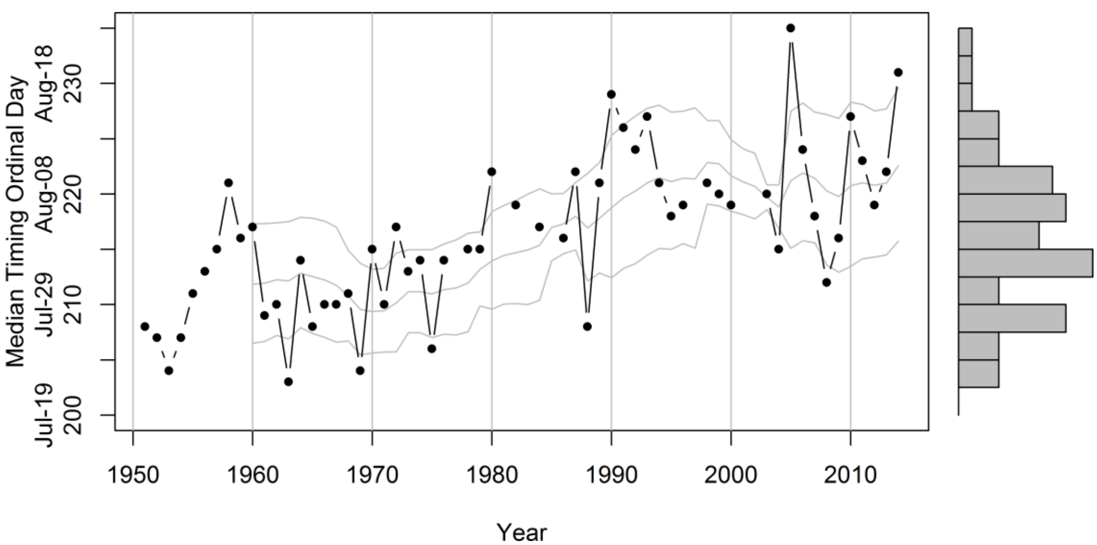

Figure 2: Time series of post-season estimated Chilko median arrival date to DFO statistical area 20. The y-axis has both calendar and ordinal date for comparison with other data and plots. The grey lines are the ten year running averages of median and standard deviation (SD), which were calculated in log-space. The histogram is scaled to density (fraction of the total numbers of occurrences).

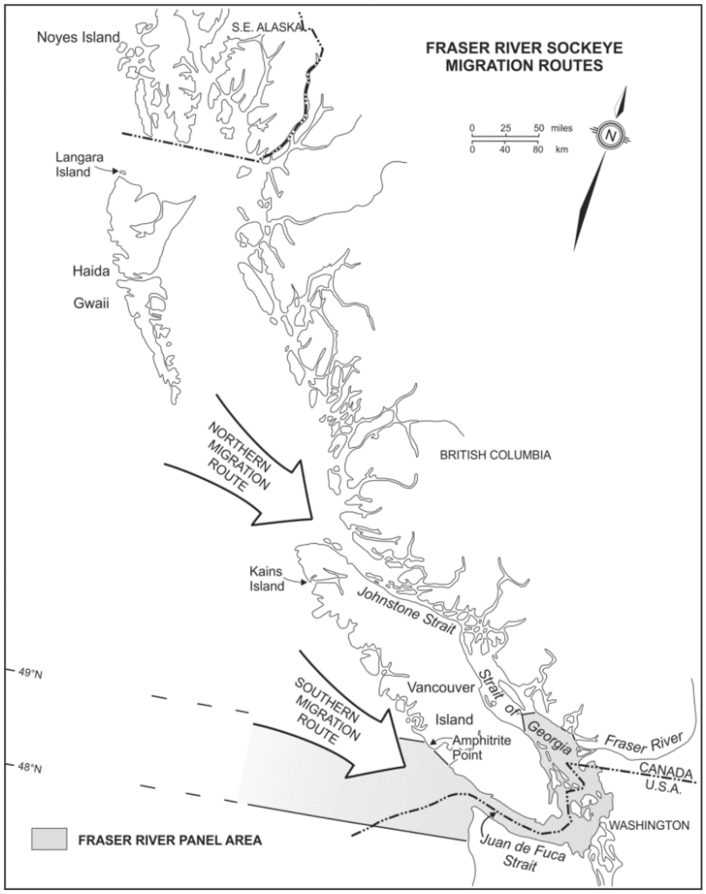

Figure 3: B.C. coast map depicting the two routes taken by adult Fraser sockeye when returning to the Fraser River. The proportion of total returns via the northern route is considered the ND rate. Image courtesy of the PSC.

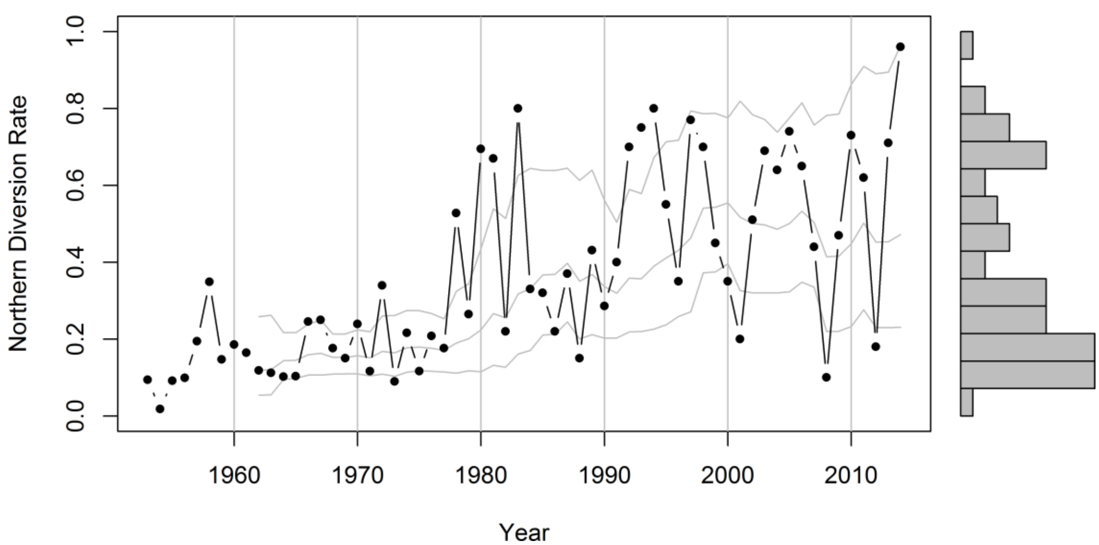

Figure 4: Time series of post-season estimates of the ND rate. The grey lines are the ten year running averages of median and SD, which were calculated in log-space. The histogram is scaled to density (fraction of the total numbers of occurrences).

Independent Data (environmental time series)

Eight distinct types of environmental data are gathered for exploration in this work: El Niño indices, Fraser river discharge, relative sea level, sea surface temperature (SST), sea surface salinity (SSS), wind stress, ocean current velocity, and earth magnetic field estimates. The derivation of these time series is documented in Folkes et al (2017).

Methods

Model Search

We use statistical models to forecast both return timing and northern diversion (ND) rate based on statistical relationships between these variables (the dependent data), and environmental/biological (independent) data. Thus, in the search for the best possible forecasting tool we have two components to evaluate: the type of statistical model and the data used in that model. Four different types of statistical models have been considered: naïve (based on time series average: TSA or time series median: TSMd), single and multiple variable linear regression (lm & mlr), generalized additive models (GAM; Wood 2006), and shape constrained additive models (SCAM; Pya, 2015).

Multiple linear regression models were developed using a stepwise regression approach, considering only the variables from single variables regressions that were statistically significant (based on a sequential Bonferroni adjustment), with R^2 >0.5, and a sample size of 17 or more data pairs. AICc values were used as an index of model improvement. All data sets considered in construction of the mlr models were limited to years shared in common. This constraint allows for application of the AICc. Additionally, the mlr models were constrained to a limit of three variables.

Model Selection By Performance Analysis

Qualifying models were then appraised by two performance evaluation methods: retrospective and jackknife analyses. Performance measures (PM) were calculated to estimate model bias (e.g. mean raw error: MRE) and precision (e.g. root mean square error: RMSE, & mean absolute error: MAE). Additional measure were appraised but are not discussed in this document. Within each performance evaluation method, the models were ranked by each PM. Mean model rank was then estimated as the mean of all PM ranks, by model.

Results

In all performance analyses all naive models ranked worse than the median rank value. The poor ranking of all naive models is mostly due to their greater forecast uncertainty as estimated by MAE and RMSE--relative to the environmental models. However, naive models that are based on a statistic of the complete time series will also be biased if the time series is trending. We see this result looking at the MRE of both the TSA and TSMd models of Chilko timing and the ND rate. Both the Chilko timing and ND time series are trending to increasing values. This confirms that naive models should not be considered in any attempt to forecast Fraser sockeye return timing or ND rate (or any time series that is trending).

Forecast models based on three variables consistently showed greater forecasting performance (ranking), compared to one and two variable models---under both retrospective and jackknife testing. This appears mainly due to the inability of models based on one or two variables to capture extreme events, including the anomalous timing of 2005.

The relative performance of a model in both the retrospective and jackknife analyses appears to be strongly based on model structure and the ability to forecast extreme value years. Models based on either one or two variables that ranked in the top 100 retrospective results usually fail to maintain similar ranking in the jackknife (example in figure 5). The retrospective PMs are calculated from forecasts of six years 2007--2012, while the jackknife PMs are calculated from forecasts of all available years over 1996--2012. Using the Early Stuart as an example, the range of observed timing values over 1996--2012 is relatively small. Most years (15 of 17) are within five days of the 1996--2012 median date. A model that tends to forecast close to the time series median (i.e., relies on environmental variables that are insensitive to observed timing changes) could forecast acceptably well during this period of the time series. However, when tested against years with extreme values (1997, 2005), the simple (one or two variable) models' inability to forecast outside the standard deviation of the time series becomes apparent.

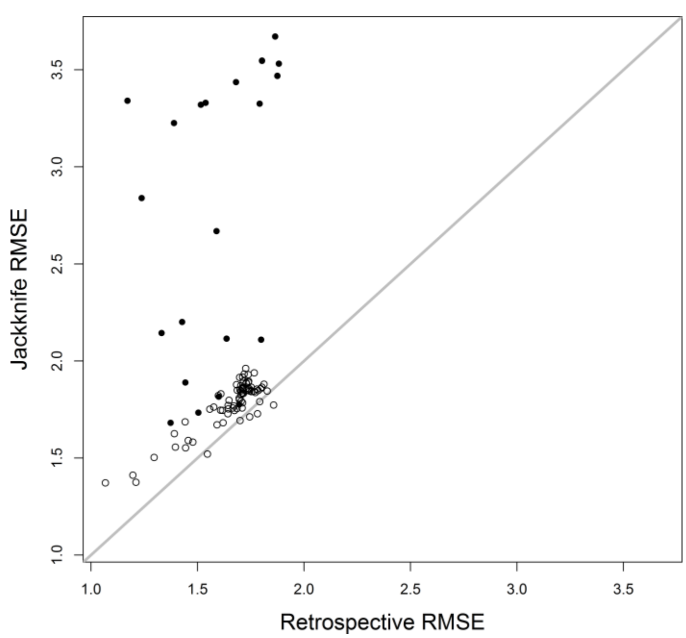

Figure 5: Comparison of model-specfic RMSE values from retrospective (x-axis) to those from jackknife results (y-axis) for top 100 models (Early Stuart timing). Solid points represent models comprising one or two variables, while open points represent models with three or more variables. The diagonal line

represents a slope of one, such that a point on (or near) the line represents a model with similar rank in retrospective and jackknife evaluation.

Discussion

Having interpreted why we see differences in model ranking between performance tests, we are now challenged with resolving the best candidate models---despite obtaining different results between the two performance tests. Ranks were recalculated for models comprising three variables since they generally performed better than models with one or two variables. The slope of the rank line within a Rank:Order plot (example Figure 6) indicates how rapidly model performance is declining between neighbouring models. A steep slope indicates a rapid decline in model performance, while slopes close to horizontal suggests neighbouring models are comparable in performance. Changes to the slope of the average rank line delineate a substantial shift in performance between neighbouring models. A sudden, positive increase in the slope (at an inflection point) of the average rank-line would suggest a threshold beyond which there is a rapid decline in model performance. However, a rapid increase of slope is apparent in just two of the Rank:Order plots: Early Stuart timing (Figure 6) and ND rate (not included here). All other analyses show rank-lines with diminishing (but always positive) slopes, i.e., successive models are not substantially inferior to their predecessors. While the Rank:Order plots do help give a general overview of top ranking model groups, this approach cannot be consistently applied to all evaluations for selection of top candidate models.

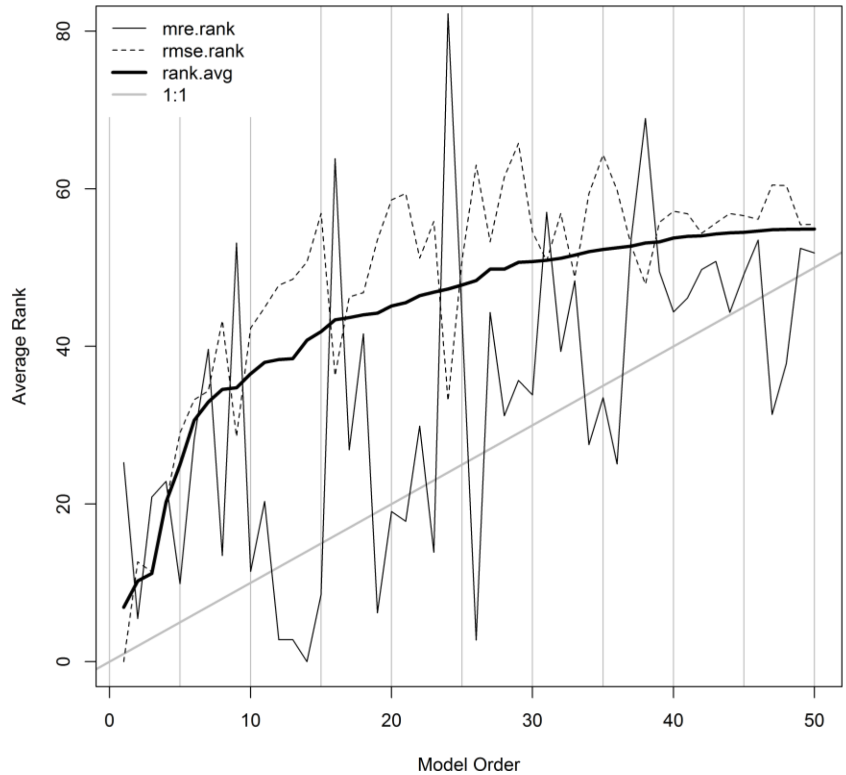

Figure 6: Rank:order plot. MRE rank, RMSE rank, and average rank for the top 50 Early Stuart timing forecast models comprising three or more variables, based on retrospective analysis. The diagonal grey line has a slope of one, which allows for comparison to the average rank line.

An alternate approach to selection from top models could be based on user tolerance of model uncertainty. User tolerance can be evaluated using graphs relating the probability of an event occurring and the range of outcome for that event. In this evaluation, the range of outcome for an event equates to the uncertainty of a forecast model. The model uncertainty (confidence limit) is estimated from the RMSE, which is available from both retrospective and jackknife evaluation (Haeseker 2005, 2008). The RMSE equates to the forecast model standard deviation, which has a probability of approximately 0.68. The frequency distributions of forecast errors (from the jackknife results) are approximately normal in shape and do not include any extreme outliers. If our assumption of normal error distribution is correct, we can calculate probability values from a range of confidence limits with little concern of poor estimation at the extremes of the probability distribution due to skewness.

We include an example tolerance plot, based on contours, indicating the relationship between uncertainty of a forecast model and the probability of that uncertainty (Figure 7).

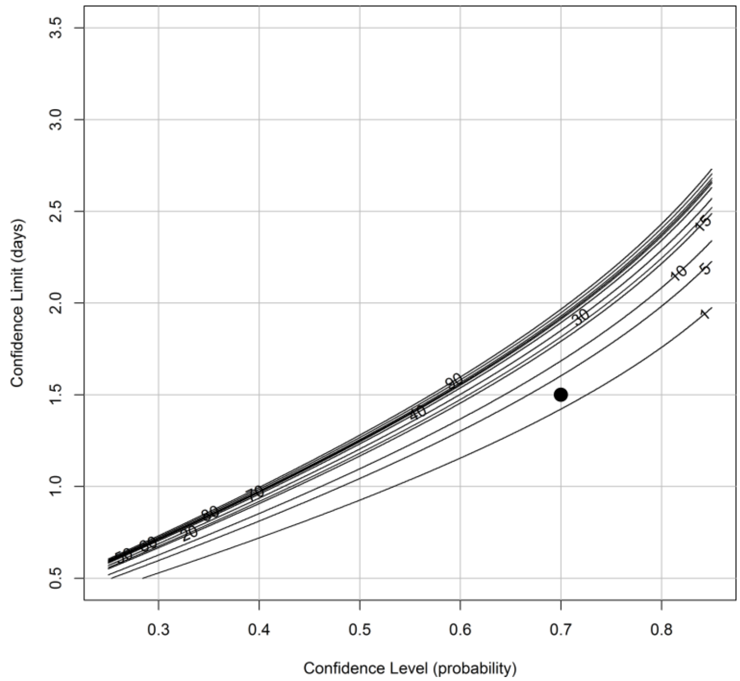

Figure 7: Contour plots relating model confidence level (i.e., likelihood, x-axis), and model confidence limit (i.e., range of error in days, y-axis), for Early Stuart timing forecast models. The confidence interval is twice the confidence limit. For example, we could say there is a 50% likelihood (confidence level) that a model, on average, will forecast timing within a two day confidence limit. Thus, given a specific likelihood (50%), the forecast fits into a range (confidence interval) of four days. The number on each contour line defines how many models qualify under these trade-off conditions. The large point on this plot defines the 70% confidence level with a confidence limit of 1.5 days. There are between 1-4 models that fulfil these criteria. Forecast certainty (a qualitative trait represented by both confidence limit and confidence level) is relaxed with increasing values on the y-axis and decreasing x-axis values. The isopleths could define lines of common forecast certainty. We refer to these as tolerance curves.

Given tolerance criteria, e.g., 0.70 probability of a forecast being within 1.5 days, we note in Figure 7 that this tolerance (identified by the large circular data symbol) lies between the contour lines representing one and five models. Thus, we can conclude that there are between one to four specific models that likely fulfil this tolerance criteria. The tolerance curves we present are based solely on model uncertainty, without consideration of the possible role of model forecast bias. However, forecast model bias that is based on the MRE estimate, is extremely small---considering just top 25 models from jackknife analysis. We suggest that due to the minimal role of bias in these top models, it need not be considered in the tolerance curves.

Conclusion

We have presented an approach to evaluating statistical models to forecast fish dynamics (represented here as marine timing and diversion rate). It marks an improvement over historical methods to forecast Fraser sockeye migratory behaviour. This evaluation takes advantage of ocean data sources that are updated frequently enough to allow their use in pre-season forecasts.

Additionally, we have presented an objective tool for forecast model selection given the trade-off between probability and forecast uncertainty. We are not recommending the use of a particular collection of forecast models as their selection, given the terms above, is fundamentally a subjective exercise determined by the risk tolerance of individual decision makers. The risk tolerance possessed by decision makers is not necessarily a fixed entity and may wax and wane between risk-averse, risk-neutral, to risk-seeking. If this is true, re-evaluation of tolerance and then re-selection of models from the tolerance trade-off curves would be prudent.

Changing the group of models (based on changing tolerance) applied to forecasting should not require a new 'search' for the best performing models. We do not recommend that the model search and performance testing be rerun on an annual basis. We believe the performance testing is sufficiently rigorous to rank only robust models having a capacity to forecast with acceptable precision across multiple years. However, one should not assume that any or all of the top models will remain robust to environmental changes over extended periods of time. It has been demonstrated that forecasts derived from multiple models are less susceptible to time series outliers and conditions of non-stationarity (Kuhn and Johnson, 2013). This evidence emphasizes the value in estimating an annual forecast from multiple, independently derived forecasts.

References

Bonjean, F. and Lagerloef, G.S.E. 2002. Diagnostic model and analysis of the surface currents in the tropical pacific ocean. Journal of Physical Oceanography 32(10): 2938–2954.

Folkes, M. J. P., Thomson, R. E., and Hourston, R. A. S. 2017 Evaluating Models To Forecast Fraser Sockeye Return Timing And Diversion Rate. Canadian Science Advisory Secretariat (in press).

Gilhousen, P. 1960. Migratory behaviour of adult Fraser River sockeye. Progress Report 7, International Pacific Salmon Fisheries Commission.

Haeseker, S.L., Peterman, R.M., Su, Z. and Wood, C.C. 2005. Retrospective evaluation of preseason forecasting models for pink salmon. North American Journal of Fisheries Management 25(3): 897–918.

Haeseker, S.L., Peterman, R.M., Su, Z. and Wood, C.C. 2008. Retrospective evaluation of preseason forecasting models for sockeye and chum salmon. North American Journal of Fisheries Management 28(1): 12–29.

Kuhn, M. and Johnson, K. 2013. Applied predictive modeling. Springer.

Pya, N. and Wood, S.N. 2015. Shape constrained additive models. Statistics and Computing 25(3): 543–559.

R Core Team. 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Reynolds, R.W., Rayner, N.A., Smith, T.M., Stokes, D.C. and Wang, W. 2002. An improved in situ and satellite SST analysis for climate. J. Climate 15(13): 1609–1625.

Wood, S. 2006. Generalized Additive Models: An Introduction with R. Taylor & Francis.