Guidelines for Application of State-Space Production Models to Japanese Fish Stocks (Fiscal Year 2023)

June 12, 2023[1]
Stock Assessment Working Group
(Authors: Momoko Ichinokawa, Mitsuyo Miyagawa, Kohei Hamabe, and Akira Hirao)

Introduction

The stock size estimation method used for stock assessment in Japan has mostly been VPA, which uses the catch in number at age as base information. However, when reliable catch in number at age data is not available, but catch size and the stock abundance index are available, then Harvest Control Rules (HCR) based on empirical evidence (bypassing the need for stock size estimation), such as the Group 2 Rules, have been used to calculate the acceptable biological catch (ABC). Meanwhile, expansions in the number of species covered by total allowable catch (TAC) policies under the revised Fishery Act have led to a spike in demand for the provision of decisive results of stock status according to Kobe plots, and for estimation of absolute stock size. In recent years, stock assessment methods have made it possible to perform stock size estimates according to assumptions and prior distribution despite a low volume of data, and it is recommended that these methods will also be applied more widely in performing stock size estimates for Japanese stocks.

Production models make it possible to theoretically estimate stock size and maximum sustainable yield (MSY) reference points based on only the stock abundance index and catch size. Because these models don't require high volumes of data, they have the longest track record of being used as stock assessment models. However, the parameter estimation method based on production model population dynamics equation(s) with assumptions at equilibrium (equilibrium method) is known to create an estimation bias in stock size estimates and MSY reference points (Mace and Mace 2001). Accordingly, there were years when the use of production models, or MSY reference points as estimated using production models with assumptions at equilibrium, were avoided intentionally (Larkin 1977) . However, these issues are being resolved thanks to the widespread use of appropriate methods for estimating parameters of production models without assumptions at equilibrium (Mace and Mace 2001), the development of Bayesian production models using prior distribution to compensate for low volumes of data (McAllister et al. 2011), and the development of state-space production models which estimate process errors using state-space models (Millar and Meyer 2000). Furthermore, these advancements are being distributed as R packages (SPiCT: Pedersen & Berg 2017; JABBA: Winker et al. 2018), which have allowed for production models to be used as stock assessment models

[1] The 2022 guidelines drafts have been adopted as the official guidelines. The main changes are: 1) feedback on the results of trial applications of production models to 11 stocks in 2022, 2) inclusion of information relating to regime shift revisions, 3) inclusion of information relating to the use of abundance estimated by VPA, 4) consideration of prior distribution in carrying capacity (K) and fishing gear efficiency (q), and 5) the addition of equations (6) through (10) as described in the appendix on documentation.

more easily (cf. Wang et al. 2020, see also Miyagawa et al. 2021).

There is sufficient data available to apply production models to Group 2 stocks in Japan (stocks for which numbers at age estimates cannot be performed, but for which stock abundance index and catch is available), so the applicability of these models must be investigated. The aforementioned software tools (JABBA, SPiCT) are designed to obtain estimates that are stable as possible (although these innovations are behind-the-scenes), which leads to different estimate results when the same data is processed by different software tools[2]. When introducing these production models to stock assessments in Japan, it is important to fully understand these behind-the-scenes settings, while agreeing on which assumptions should be made to achieve stability in estimates.

These guidelines are intended to present the procedures and points that should be kept in mind when performing stock assessments by fitting Japanese stock data in production models, and to explain utilization of stock assessment results according to the degree of uncertainty. There is a presumption that SPiCT is used as the software tool for state-space production models, but there are cases when it is also appropriate even if different software is used for assigning prior distribution or utilization of stock assessment models. If necessary, data which was pseudo-generated using simulations (Example 1) (https://github.com/ichimomo/frapmr/blob/dev/raw-data/example1.csv), or the results of trial applications of production models to 11 stocks in 2022 (Table 1) (Stock Assessment Working Group 2023a), can be included as examples of estimated results or model diagnostics results. We also included an appendix with an example of how to present the results of trial applications of production models in documentation. Figures were made using both SPiCT and frapmr (https://github.com/ichimomo/frapmr, not available to the public).

Production Model Fitting Procedures

The production model used here is a Pella-Tomlinson state-space production model (Equation 1).

$$B_{t+1} = \left[B_t + \tfrac{r}{n-1}B_t\left(1 - \left[\tfrac{B_t}{K}\right]^{n-1}\right) - F_tB_t\right]\exp(\varepsilon_t),\quad \varepsilon_t \sim N(-0.5\sigma_B^2,\, \sigma_B^2) \qquad \text{(Equation 1)}$$

$$I_{t,i} = q_iB_t\exp(e_{t,i}),\ \ e_{t,i} \sim N(0,\, \sigma_{I,i}^2)$$

In this equation, B_t and F_t are the biomass and fishing mortality in year t, r is the intrinsic natural growth rate, n is the shape parameter, K is the carrying capacity, ε_t is the process error in year t, σ_B is the magnitude of the process error, $I_{t,i}$ is the number i stock abundance index in year t, q_i is the fishing gear efficiency for the number i stock abundance index, $e_{t,i}$ is the observation error for the number i stock abundance index, and $\sigma_{I,i}$ is the magnitude of the observation error for the number i stock abundance index.

As discussed above, SPiCT is a differential production model, so Equation 1 is slightly different from the

[2] Because JABBA is based on a Bayesian production model, it is necessary to assume some form of prior distribution in estimated parameters, and estimates cannot be performed without applying assumptions. Meanwhile, SPiCT allows for estimates to be obtained without any data, which makes it possible to perform estimates in time increments shorter than 1 year. This is slightly different from the discrete models, which commonly rely on population dynamics equations. In addition, because the default settings are designed to stabilize estimates, parameter estimation is performed using a weak prior distribution for all estimated parameters.

population dynamics equations assumed within SPiCT. SPiCT allows users to process data in time increments shorter than 1 year (dteuler) to obtain numerical solution for differential population dynamics equations. However, it is also possible to imitate population dynamics by difference equation such as Equation 1 by setting dteuler to 1 (Pedersen and Berg 2017). And in practice, we found that data generation using a difference production model yielded the best true parameters when dteuler was set to 1 (Stock Assessment Working Group 2022b). Based on this, we established default settings that provided a structure equivalent to Equation 1 when applying SPiCT to Japanese stocks (dteuler = 1). With these settings, the meanings of the parameters will be the same as the discrete population dynamics equation as shown in Equation 1. However, in cases when convergence is not achieved with dteuler = 1, or when you want to utilize data with shorter time increments, dteuler can be adjusted appropriately. There are also many difference production models which do not estimate fishing mortality, and instead perform calculations using the exploitation rate C_t/B_t (Winker et al. 2018, and others). Meanwhile, there are other ways which SPiCT is different from other production models, such as assuming fishing mortality F as an unobserved process which follows a time series random walk model.

Preparation Step 1: Gather findings on biological characteristics and fishing of target species

First, it is necessary to gather information to assist in judging whether the parameters estimated using the production model are appropriate. If it is not possible to easily estimate these parameters in a model without any assumptions, then you should use the gathered findings as reference information for parameters. In addition to findings on biological characteristics and fishing as shown below, it is extremely important to gather information relating to regime shift and the presence or absence of other long-term trends in changes in productivity.

<u>Intrinsic natural growth rate (*r*)</u>

Parameters which describe the behavior of the entire population, such as the intrinsic natural growth rate r, are more difficult to perform estimation and judgement compared to parameters based on observation of individuals such as growth equations, age of maturity, and asymptotic length. However, the integration of a database of biological characteristics of each fish species (FishBase: Froese and Pauly 2022) and a database which accumulates population dynamics parameters obtained from stock assessment models (RAMLegacy database: Ricard et al. 2012) makes it possible to see the relationship between population dynamics parameters and biological characteristics at the individual level (Thorson 2020). The code by Thorson (2020) for obtaining projected values for *r* for each species is distributed as an R package called FishLife (https://github.com/James-Thorson-NOAA/FishLife), which enables us to gather fundamental findings on what kind of *r* can generally be expected for each target stock. It is strongly advised to obtain projected values for r_{fish} from this package as part of the information gathering stage. Estimates for r_{fish} from FishLife correspond to the slope near the origin of the surplus production curve, assuming a Schaefer production model with the shape parameter (*n*) set to 2 (this will be discussed in the next section). When $n \neq 2$ in Equation 1, the slope near the origin of the surplus production curve (in summary, r_{fish}) corresponds to *r*/(*n*-1). Therefore, to obtain the prior distribution of *r* from r_{fish}, it is necessary to theoretically use $\log(r_{fish}) + \log(n-1)$ as the mean value of prior distribution of $log(r)$. However, the performance when *r* values obtained this way are actually applied to stocks has not been fully investigated, so caution is advised (especially in cases when *n* is close to 1). Likewise, when $n < 1$, the slope at the origin theoretically becomes infinite, so it is not possible to directly assume estimated values from FishLife. When the likelihood profile plot (Fig. 2) is created using frapmr, the settings cause r to be shown as a Pella-Tomlinson model in the upper left horizontal axis, and r_{old} to be shown as a Schaefer model in the lower left horizontal axis, creating a relationship in which $r_{old} = r_{fish}$.

While FishBase and FishLife are both specialized for fish data, there is also SeaLifeBase (Palomares and Pauly 2022) which accumulates information on non-fish marine life. Both FishBase and SeaLifeBase rank species resilience (capacity of the population to recover after disturbances) in 4 levels: very low, low, medium, and high. The values set by Froese et al. (2017) can be used as references for setting the range of prior distribution of r, with very low = 0.015 to 0.1, low = 0.05 to 0.5, medium = 0.2 to 0.8, and high = 0.6 to 1.5.

SPiCT allows users to assume changes in maximum surplus production capacity due to regime shift (Mildenberger et al. 2020), and outputs 2 different intrinsic natural growth rates (r_1 and r_2) for before and after regime shifts. If regime shifts like these are assumed, then the change in production Δm will be estimated without changing K due to SPiCT's design, so the prior distribution of r cannot be set. See the Stock Assessment for Hokkaido Sea of Japan Stock of Pacific Cod (Sakai et al. 2022a, Sakai et al. 2022b) for a specific example of how this can be applied.

Shape parameters (n)

The shape parameter determines the shape of the surplus production curve. A larger parameter indicates a greater relative position of Bmsy compared to carrying capacity (K), for example, if $n \approx 1$ then Bmsy/K = 0.367 (Fox production model), if n = 2 then Bmsy/K = 0.5 (Schaefer production model), and if n = 4 then Bmsy/K = 0.630. In meta analysis by Thorson et al. (2012), it was demonstrated that the mean Bmsy/K for all fish species was 0.404, with some differences in mean within the same taxonomic groups, such as Pleuronectiformes at 0.395, Gadiformes at 0.439, Perciformes at 0.353, Clupeiformes at 0.261, and Scorpaeniformes at 0.463. Because the standard deviation of these estimated values ranges from 0.1 to 0.13, we believe that the realistic range for general Bmsy/K would be from 0.25 to 0.6.

Carrying capacity (K) and fishing gear efficiency (q)

Carrying capacity and fishing gear efficiency are parameters which determine the scale of absolute size of the entire population. Although it is not simple to determine an appropriate range for these parameters, it is possible to utilize other information such as estimated stock size from other data sources when available, for example, to investigate the possibility of introducing prior distribution to judge whether estimated parameters K and q are appropriate.

One example of utilizing estimated stock size values from other data sources is seen in the 2022 Stock Assessment for Northern Hokkaido Stock of Sohachi Flounder (and the 2022 Stock Assessment for Northern Hokkaido Stock of Littlemouth Flounder). In these papers, abundance estimated by VPA was used as a stock abundance index, and stock size estimates were performed by assuming prior distribution for q = 1 (Chiba et al. 2022a, Chiba et al. 2022b). It should be noted that abundance estimated by VPA follow a different definition than stock size estimated by production models, so before use, it is necessary to convert values to standing stock values which correspond to the production model. Stock size based on VPA represents the population size based on the population growth (maturity and recruitment) in a certain year, before the impact of fishing in that year. Meanwhile, stock size based on production models represents the population size in a certain year before the addition of surplus production, and before the impact of fishing in that year. In VPA, this is equivalent to the standing stock after decrease due to the impact of fishing and natural mortality in the previous year. Accordingly, the equation below will be used as the abundance index by extracting the standing stock *D* from abundance estimates based on VPA.

$$D_y = (B_{y-1} \cdot e^{\left(-\frac{M}{2}\right)} - C_{y-1})e^{\left(-\frac{M}{2}\right)} \quad \text{(Equation 2)}$$

In this equation, B_y is biomass (in weight) in year *y* as estimated based on VPA, C_y is the catch in year *y*, and *M* is natural mortality as assumed in VPA based analysis. It should be noted that even if *q* is available as reference information, it is always required to scrutinize findings when the data set is obtained.

If stock size is estimated based on VPA, then it's possible to theoretically calculate stock biomass $\widehat{B_0}$ when $F = 0$. Therefore, it should also be possible to use $\widehat{B_0}$ as an alternative index for *K*, but because the process of these calculations require assumptions in the stock-recruitment relationship, it is advised to use the biomass estimate results as the prior distribution of *q*, instead of *K*, if findings relating to the stock-recruitment relationship are uncertain.[3]

Catch (*C*) and stock abundance index (*I*)

If it is clear that catch is an estimate, then it is advised to gather information relating to the uncertainty of that estimate. Likewise, if the uncertainty of estimated stock size is expected due to standardization or other methods, then supporting information should be used. However, if the confidence interval as estimated based on standardization only evaluates data sampling errors, then the magnitude of observation error in the production model (σ_I) might not always be the same value (Francis 2011, Winker et al. 2018).

Other parameters ($B_1, \sigma_p,\ \sigma_{I,g}$)

When B_1 is the stock size at the time the stock assessment is started (the first year for which catch data is available), the relative ratio of B_1 to *K* ((B_1/K) can be inferred according to fishing conditions in the year(s) prior to the first year for which catch data is available. For example, it can be inferred that $B_1/K \approx 1$ in the initial fishing season, but $B_1/K \approx 0.5$ if fishing has been conducted for many years. Froese et al. (2017) proposed that if catch data is available for years before 1960, then B_1/K will range from 0.5 to 0.9, and if data is available after 1960 then this value will range from 0.2 to 0.6. Even if accurate catch statistics cannot be obtained for fishing conditions in the years before stock assessment was first performed, gathering other information about the prior intensity of fishing operations can help to presume the realistic range of priors to estimate B_1/K values.

If process error (σ_p) and observation error (($\sigma_{i,g}$) cannot easily be estimated at the same time, you might be able to make the assumption that $\sigma_p = \sigma_{i,g}$ (Thorson et al. 2013, Pedersen and Berg 2017).

Preparation Step 2: Re-confirming data used in the model(s)

Simple plots can be used to confirm understanding of the bigger picture of the data and how well the data aligns

[3] Up to 2022, the guidelines (draft) recommended the relationship proposed by Froese et al. (2017) as a benchmark for the validity of the scope of *K*. However, this assumed *n* = 2 in a Schaefer model, and the limits of validity were unclear in Pella-Tomlinson models, so this recommendation has been removed from the 2023 edition.

with assumptions in the production model. For example, SPiCT includes a function tool named plotspict.ci which outputs a plot similar to Fig. 1, which allows users to visualize exactly how much data deviates from assumptions in the production model (for example, that higher fishing effort (aka fishing pressure) directly corresponds to a lower stock abundance index). However, when this was applied to various species in 2022, plotspict.ci revealed some cases when parameters were not easily estimated even if the data indicated that higher fishing pressure directly corresponded to a lower stock abundance index. Therefore, it is advised that these plots should only be treated as a reference.

Fitting Data to State-Space Production Models

Developing models in stages

Ideally, parameters should be estimated based on available data, with as few assumptions as possible. But unless excellent data is available, it is generally agreed that it is not possible to estimate all parameters in a production model without some bias (Pedersen and Berg 2017). Accordingly, it is important to develop the model in a process which starts with zero assumptions for prior distribution of any parameters (model 0), then to add weak assumptions in stages, until the model can obtain realistic estimate values. For example, the models below were developed by gradually adding assumptions while confirming how each assumption contributes to the stabilization and results of the model. For 'model 0' and 'model 1', the results from settings shown below should be investigated regardless of fish species. For 'model 2+', the settings should be developed and adjusted according to findings and data for each fish species as necessary. Likewise, because it is known that reference points are highly dependent on the parameter n, which is difficult to estimate, we recommend to show results for $n = 2$ and $n = 1.19$ as 'sensitivity analysis 1' and 'sensitivity analysis 2', regardless of the assumption of n in 'model 2+'. Results for 'models 2+' should also be shown as 'sensitivity analysis 3+' to reflect assumption settings as necessary (include all relevant sensitivity analysis results).

Model 0	Zero assumptions for prior distribution of any parameters.
Model 1	Assume a weak prior distribution of r and n based on the information from FishLife and meta analysis. To find the mean of prior distribution, use convenient projected values from FishLife for log(r), and log(2) (Bmsy/K = 0.5) for log(n). Use sd = 1 for the standard deviation. [4]
Model 2+	This model should have stronger constraint than model 1. Here's an example of how to develop the model: - Take the prior distribution in 'model 1' and use sd < 1 for the standard deviation of prior distribution (for example, sd = 0.5) in order to increase the relative impact of prior information. - Assign prior distribution to q if the results of other stock assessments are usable. - Fix n at a value that fits the data better (higher likelihood) based on results from the likelihood profile (only if $n > 1$). If the likelihood is maximized when $n < 1$, then do not

[4] https://github.com/stan-dev/stan/wiki/Prior-Choice-Recommendations

	fix *n*, and instead set other parameters such as log(1.19) for the *n* prior distribution mean, and sd = 1. In these cases, the prior distribution mean of log(*r*) can be log(*r*) + log(*n*-1), which is adjusted according to *n* from FishLife. - Assume that the process error and observation error are equal. - Consider the uncertainty of catch. etc.
Sensitivity analysis 1	- If estimates for *r* are obtained from FishLife or FishBase, fix *n* = 2 as a reference case (Schaefer production model), and use log(*r*) and its standard deviation from FishLife for prior distribution of log(*r*).
Sensitivity analysis 2	- Fix *n* as estimated based on meta analysis results (Thorson et al. 2012) (for example, *n* = 1.19 for all fish species), and use log(*r*) + log(*n*-1) (values from FishLife) and the standard deviation of log(*r*) as the prior distribution of log(*r*).

To judge whether the results estimated in each model are appropriate, it is advised to investigate the following: (1) the numerical stability of estimated results, (2) model diagnostics results, and (3) whether estimates are within the appropriate parameter range(s) gathered during preparation. If you encounter problems regarding Step 1 (the stability of estimated parameters) as described below, then there is no reason to share model analysis results. At this point, the whole model needs to be reconsidered.

(1) The stability of estimated parameters

- **Presence/absence of convergence:** Convergence is present, and the parameter estimation errors are obtainable. If convergence is absent, then the required number of iterations might exceed the limit. Try raising the iteration limit settings.
- **Range of estimated parameters and the confidence interval:** Estimated parameters or the confidence interval are neither infinite nor diverge at extremely high values. Specifically, it is recommended that the upper 5% confidence interval of a parameter is approximately within 10 times the lower 5% confidence interval.
- **Robustness compared to initial values:** The same maximum likelihood estimates should be obtained even if initial values are adjusted.

(2) Model Diagnostics

- **Likelihood profile** [Set certain parameters to fixed values, estimate the remaining parameters, and repeat the process with different values of fixed parameters to observe how likelihood changes (Fig. 2)]: Construct a likelihood profile for *r* and *n*, and check the extent to which *r* and *n* values used for prior distribution in ‘model 1’ are consistent with the data.
- **Retrospective analysis** [Delete one year of data, counting back from the most recent year, and repeat estimates]: Even if stable parameters are obtained in Step 1, significant problems might be revealed in retrospective analysis, so retrospective analysis should always be performed to confirm the presence/absence of retrospective bias (See Fig. 3 for an analysis example). SPiCT added the hindcast cross-validation tool in Ver. 1.3.7. Hindcast cross-

validation is similar to retrospective analysis because it deletes one year of data, counting back from the most recent year, then repeats estimates. It also performs future projections for the abundance index and compares them to the observed abundance index, which allows for a true evaluation of projection accuracy.

- **Residual analysis** [Use residual patterns (observed value - projected value) to judge how well the model explains the abundance index which is used as data]: Do the residuals fit the assumption of a normal distribution? Are there any significant autocorrelation patterns in the residuals? SPiCT includes a model diagnostics program which automatically outputs test results regarding the significance of autocorrelation of residuals and the normality of residuals. If you encounter problems with these results, it indicates that there is room for improvement in the data or the construction of your model, but problems at this stage are not severe enough to reject the whole model.
- **Factor analysis** [Observe how, and how much, factors like surplus production, catch, and process errors influence fluctuations in population estimates (Fig. 4)]: If surplus production, catch, and process errors are estimated on the approximately the same scale, then they are considered to conform with the general assumptions in the production model. However, there are some cases when most fluctuations in population dynamics can be explained by process error(s), and in cases like these, factor analysis will reveal if information such as surplus production cannot be obtained from the available stock data.
- **Sensitivity analysis** [Change the value of some parameters (assumptions in the models) to observe how results change]: Specifically, ‘sensitivity analysis 1’ and ‘sensitivity analysis 2’ should always be performed. Then, if other assumptions are used in ‘model 2+’, the influence of these assumptions on results should be investigated using sensitivity analysis to identify how dependent the results are on these assumptions. It is taken for granted that adjustments in the model assumptions will produce different results, but it is worth exploring exactly which results change easily when the assumption(s) are adjusted, and conversely, it is also worth exploring which results remain in common despite adjustments of assumptions within the range of realistically possible values. This will allow you to write stock assessment reports which focus on the most robust stock assessment results. (For example, if absolute stock size changes dramatically when the assumption(s) are adjusted, then specific figures for absolute stock size should not be shared in the report. Meanwhile, if results for when B/Bmsy is higher/lower than 1 are robust, then those results should definitely be highlighted in the report.)

Utilization of Stock Assessment Results

As explained earlier, the construction of production models should start with ‘model 0’, which only uses data. Then, the production model can be developed by adding external information such as prior distribution and assumptions for parameters. This process helps guide the estimates into a realistic range. The confidence interval of estimated parameters will be narrower when the standard deviation of prior distribution is smaller, or when assumptions are assigned, so it will probably seem that better estimates are obtained as the model is developed. However, if there are errors in the prior distribution, this will lead to errors in judgement of stock status and ABC calculations, so caution is always advised. In addition, it is difficult to quantify the certainty of input such as prior information, so in some cases, it may be impossible to reduce candidates to a single model.

Furthermore, the results of a tentative study using management strategy evaluation (MSE) (Stock Assessment

Working Group 2022a) showed that if trustworthy prior information is unavailable for parameters (q and K) which determine absolute stock size, then a large error will occur in the estimate of absolute stock size. It has been shown that this leads to worse performance under Harvest Control Rules (HCR) for Group 1A (in which ABC is calculated by multiplying Fmsy by stock size as estimated in the production model), even more so than compared to management under HCR for Group 2. Meanwhile, if prior distribution of q and K is assigned without bias and with a certainty of approximately sd = 0.5, then the adjusted coefficient β used in HCR for Group 1A will be smaller (for example, approximately the standard deviation of estimated stock size), which will lead to equal or better performance than HCR for Group 2.

For these reasons, the utilization of production model results varies greatly according to the certainty of the prior information available for the target species, and the robustness of the estimated results. These classifications are described in Table 2. Within these classifications, it is acceptable to adopt one or both of 1C21 and 1C22. When compared against Group 2 stocks which don't utilize production model results, or which the results aren't included as an Appendix to stock assessment reports, there are some advantages to policies which publish the results (including partial results) of production models in stock assessment reports:

- Multiple stock abundance indices can be integrated into a single stock abundance index, and if this is used for HCR for Group 2 (1C21), then all available data for stock assessment and management can be reflected in ABC calculations. This also helps to avoid situations when reference points and ABC are different depending on which data is used.
- When decisive results for stock levels from production models based on population dynamics models are included (1C22), then it will also be possible to evaluate the plausibility of decisive results on stock status according to Group 2 rules. While it is very important to ensure a smooth start to management of stocks with high plausibility, it is also possible to assign a high priority to stocks with low plausibility that are in need of active introduction of species-specific MSE. If robust results are obtained for judgement of stock status or estimate process error(s), then those results can serve as evidence to support changing biomass targets (BT) in Group 2 rules from default values.
- If reasonably accurate population dynamic estimate results can be obtained, then MSE can be performed based on parameters as estimated in the production model(s), which will allow you to make species-specific improvements in the adjustment coefficient for Group 2 rules (e.g., Stock Assessment Working Group 2023b).

When production model results will also be included as an Appendix, it is important to share as much detailed information as possible in documentation, such as the estimates used in calculations, and model analysis results. We also included an example of including detailed information in documentation in "Appendix. Example of Documentation to Share the Results of Applying a State-Space Surplus Production Model."

References

Chiba, S., Sato, R., Morita, M., Sakai, O., and Hamatsu, T. (2022a) Stock Assessment for Northern Hokkaido Stock of Sohachi Flounder (Fiscal Year 2022). FRA-SA2022-AC-67, Marine fisheries stock assessment and evaluation for Japanese waters (Fiscal Year 2022), Fisheries Agency and Japan Fisheries Research and Education Agency, Tokyo, p. 37, https://abchan.fra.go.jp/wpt/wp-content/uploads/2022/details_2022_67-Souhachi-HN.pdf

Chiba, S., Sato, R., Morita, M., Sakai, O., and Hamatsu, T. (2022b) Stock Assessment for Northern Hokkaido Stock of Littlemouth Flounder (Fiscal Year 2022). FRA-SA2022-AC-70, Marine fisheries stock assessment and evaluation for Japanese waters (Fiscal Year 2022), Fisheries Agency and Japan Fisheries Research and Education Agency, Tokyo, p. 42, https://abchan.fra.go.jp/wpt/wp-content/uploads/2022/details_2022_70-Magarei-HN.pdf

Francis, R. I. C. (2011) Data weighting in statistical fisheries stock assessment models. Can. J. Fish Aqua. Sci., **68**, 1124–1138. https://doi.org/10.1139/f2011-025

Froese, R. and D. Pauly (2022) FishBase. World Wide Web Electronic Publication. www.fishbase.org, version (02/2022)

Larkin, P. A. (1977) An epitaph for the concept of maximum sustained yield. Transactions of the American Fisheries Society, **106(1)**, 1–10.

Mace, P. M. and P. Mace (2001) A new role for MSY in single-species and ecosystem approaches to fisheries stock assessment and management. Fish and Fisheries, ***2*(1)**, 2–32. https://doi.org/10.1046/j.1467-2979.2001.00033.x

McAllister, M. K., Pikitch, E. K. and E. A. Babcock (2011) Using demographic methods to construct Bayesian priors for the intrinsic rate of increase in the Schaefer model and implications for stock rebuilding. Can. J. Fish Aqua. Sci., **58**, 1871–1890. https://doi.org/10.1139/F01-114

Mildenberger, T. K., C. W. Berg, M. W. Pedersen, A. Kokkalis, J. Nielsen and R. Meyer (2020) Time-variant productivity in biomass dynamic models on seasonal and long-term scales. *ICES Journal of Marine Science*, **77(1)**, 174–187. https://doi.org/10.1093/icesjms/fsz154

Millar, R. B. and R. Meyer (2000) Non-linear state space modelling of fisheries biomass dynamics by using Metropolis-Hastings within-Gibbs sampling. Journal of the Royal Statistical Society: Series C (Applied Statistics), **49(3)**, 327–342. https://doi.org/10.1111/1467-9876.00195

Miyagawa, M., M. Ichinokawa, M. Yoda, H. Kurota, S. Ohshimo and H. Nishida (2021) Commentary: Stock Status Assessments for 12 Exploited Fishery Species in the Tsushima Warm Current Region, Southwest Japan and East China, Using the CMSY and BSM Methods. Frontiers in Marine Science, **8**. https://doi.org/10.3389/fmars.2021.703039

Ono, K., A. E. Punt and E. Rivot (2012) Model performance analysis for Bayesian biomass dynamics models using bias, precision and reliability metrics. Fisheries Research, **125–126**, 173–183. https://doi.org/10.1016/J.FISHRES.2012.02.022

Palomares, M. L. D. and D. Pauly (Editors) (2022). SeaLifeBase. World Wide Web electronic publication. www.sealifebase.org, version (04/2022).

Pedersen, M. W. and C. W. Berg (2017) A stochastic surplus production model in continuous time. Fish and Fisheries, **18 (2),** 226–243. https://doi.org/10.1111/faf.12174

Ricard, D., Minto, C., Jensen, O. P., and J. K. Baum (2011) Examining the knowledge base and status of commercially exploited marine species with the RAM Legacy Stock Assessment Database. Fish and Fisheries,

13(4), 380–398. https://doi.org/10.1111/j.1467-2979.2011.00435.x

Sakai, O., Chimura, M., Chiba, S., and Hamatsu, T. (2022) Stock Assessment for Hokkaido Sea of Japan Stock of Pacific Cod (Fiscal Year 2022). FRA-SA2022-AC-31, Marine fisheries stock assessment and evaluation for Japanese waters (Fiscal Year 2022), Fisheries Agency and Japan Fisheries Research and Education Agency, Tokyo, p. 23, https://abchan.fra.go.jp/wpt/wp-content/uploads/2022/details_2022_31-Madara-HJS.pdf

Sakai, O., Chimura, M., Chiba, S., and Hamatsu, T. (2022) Stock Analysis for Hokkaido Sea of Japan Stock of Pacific Cod by State-space Surplus Production Model (Fiscal Year 2022). FRA-SA2022-RC07-201, Marine fisheries stock assessment and evaluation for Japanese waters (Fiscal Year 2022), Japan Fisheries Research and Education Agency.

Stock Assessment Working Group (2022a) Exploring Robustness of Management Strategies Using Production Models. FRA-SA2022-ABCWG02-08, Fisheries Agency and Japan Fisheries Research and Education Agency, Yokohama, p. 21,

Stock Assessment Working Group (2022b) Time Step Settings in SPiCT Software for State-space Production Models. FRA2022-ABCWG02-12, Fisheries Agency and Japan Fisheries Research and Education Agency, Yokohama, p. 8,

Stock Assessment Working Group (2023a) Summary of the Results of Testing the Production Model (Fiscal Year 2022). FRA2022-ABCWG02-12, Fisheries Agency and Japan Fisheries Research and Education Agency, Yokohama, p. 9,

Stock Assessment Working Group (2023b) Management Strategy Evaluation and Future Projection Methods Using Stock Size Estimate Results Based on State-Space Production Models: Applied to the Northern Hokkaido Stock of Sohachi Flounder and Littlemouth Flounder. FRA-SA2023-BR03-101, Fisheries Agency and Japan Fisheries Research and Education Agency, Yokohama, p. 12, https://abchan.fra.go.jp/references_list/FRA-SA2023-BR03-101.pdf

Thorson, J. T. (2020) Predicting recruitment density dependence and intrinsic growth rate for all fishes worldwide using a data-integrated life-history model. Fish and Fisheries, **21(2)**, 237–251. https://doi.org/10.1111/FAF.12427

Thorson, J. T., J. M. Cope, T. A. Branch and O. P. Jensen (2012) Spawning biomass reference points for exploited marine fishes, incorporating taxonomic and body size information. Can. J. Fish Aqua. Sci., **69**, 1556–1568. https://doi.org/10.1139/F2012-077

Thorson, J. T., C. Minto, C. V. Minte-Vera, K. M. Kleisner, C. Longo and L. Jacobson (2013) A new role for effort dynamics in the theory of harvested populations and data-poor stock assessment. Can. J. Fish Aqua. Sci., **70**, 1829-1844. http://dx.doi.org/10.1139/cjfas-2013-0280

Wang, Y., C. Liang, W. Xian and M. L. Palomares (2020) Stock Status Assessments for 12 Exploited Fishery Species in the Tsushima Warm Current Region, Southwest Japan and East China, Using the CMSY and BSM Methods. Frontiers in Marine Science, **7**. https://doi.org/10.3389/fmars.2020.00640

Winker, H., F. Carvalho, and M. Kapur (2018) JABBA: Just Another Bayesian Biomass Assessment. Fisheries Research, **204**, 275–288. https://doi.org/10.1016/j.fishres.2018.03.010

Table 1. List of 11 stocks for trial applications of production models in 2022. The items listed here are reference materials for internal use during discussions at Research Institute Meetings. Each document is assigned an ID number beginning with FRA. Stock assessment reports and Research Institute Meeting materials for each stock in 2022 and 2023 can be viewed at https://abchan.fra.go.jp/.

Stock name	Stock assessment type	Time series length of catch	Number of series of abundance index	Report description
Western sand lance, Eastern Seto Inland Sea stock	Group 2	33	1	Details are presented in a separate document (FRA-SA2022-SC01-204).
Red-eyed round herring, Pacific stock	Group 2	43	1 or 2 (3 scenarios)	Details are presented in a separate document (FRA-SA2022-SC06-12).
Roughscale flounder, Northern Pacific stock	Group 2	51	1	Supplemental verbal explanation(s) at Research Institute Meeting(s)
Japanese Spanish mackerel, East China Sea stock	Group 2	38	1 or 2	Details are presented in a separate document (FRA-SA2022-BRP09-02).
Deep-sea smelt, Sea of Japan stock	Group 2	37	1	Unpublished reference materials
Amberstripe scads, East China Sea	Group 2	29	2	Details are presented in a separate document (FRA-SA2022-SC01-205).
Japanese scad	1C22	29	2	Kobe Plot (Appendix) is included in the Stock Assessment Report. Details are presented in a separate document (FRA-SA2022-SC01-204).
Pacific cod, Hokkaido Pacific stock	1C21 & 1C22	36	1	Included in the Stock Assessment Report along with the Kobe Plot (Appendix). Details are presented in a separate document (FRA-SA2022-RC-07-202).
Pacific cod, Hokkaido Sea of Japan stock	1C21	37	1	Details are presented in two separate documents (FRA-SA2022-RC07-201 and FRA-SA2022-BRP11-021). Summary is in the Appendix section of the Stock Assessment Report.

Sohachi Flounder, Northern Hokkaido stock	1C1	42	2	Stock Assessment Report, FRA-SA2022-SC08-01, and materials from Research Institute Meeting
Littlemouth flounder, Northern Hokkaido stock	1C1	42	2	Stock Assessment Report, FRA-SA2022-SC08-202, and materials from Research Institute Meeting

Table 2. Classification for production model results

Type	Conditions	Results which should be included	Harvest Control Rules
1C1	If reliable prior information on absolute stock size is available (stock size estimate results based on other data, or stock size estimate results based on surveys), and there are no noteworthy problems with model diagnostics results, etc.	- Model diagnostics results - Estimated stock size, Bmsy, Fmsy, B/Bmsy, F/Fmsy (main paper), etc. - Show estimated values as points and also include the confidence interval.	Catch calculated based on estimated stock size and Fmsy.
1C2	If the conditions for 1C1 do not apply, but there are some robust results, and there are no noteworthy problems with model diagnostics results, etc.	Explain which results were robust, and which were not, in an Appendix. Robust results can also be described in the main paper as supplementary information.	Group 2 rules apply due to high uncertainty in stock size estimates.
	1C21: If relative trends in stock size estimates are robust.	- Model diagnostics results - If stock size estimate results are similar from multiple models based on realistic assumptions, include figure(s) that explain the relative trend(s).	Relative trends in stock size can be applied to Group 2 rules as a stock abundance index.
	1C22: If there is robust evidence regarding the judgement of current stock status.	- Model diagnostics results - When using Kobe plots with confidence intervals, the plot for the most recent year shows the confidence interval and/or the results of multiple models in the same quadrant, or other results equivalent to this.	Group 2
Group 2	If there are problems with model	- Include documentation to	Group 2

	diagnostics results, or a lack of robust results.	describe the problem(s) encountered when applying production models, and describe the direction of future studies if necessary.	

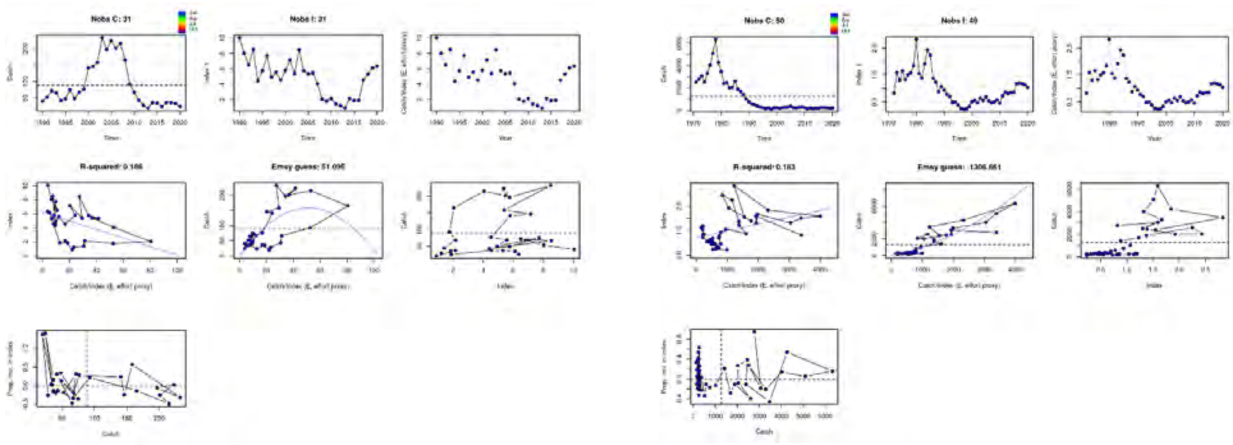
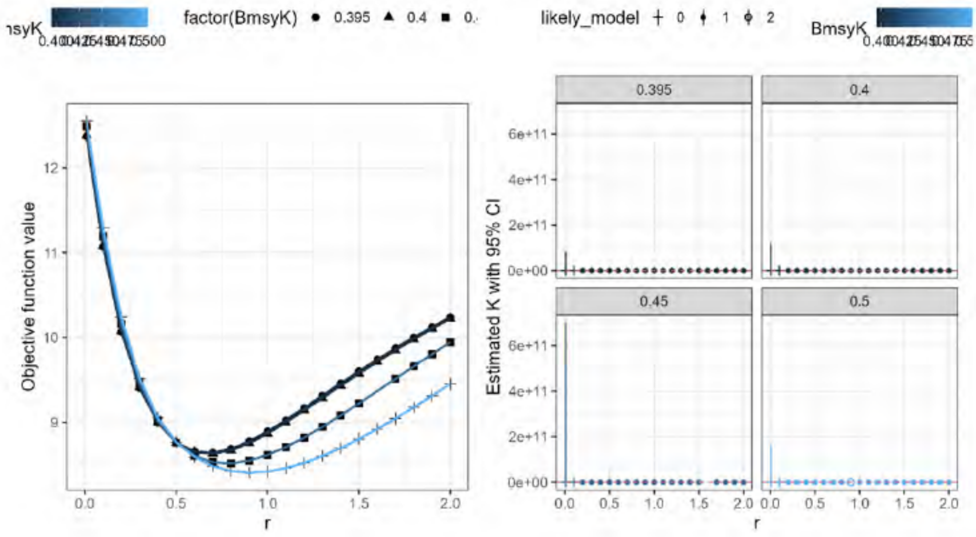


Fig. 1. Examples of plots generated using plotspict.ci. Left: Ideal example generated from simulation data. Right: Example of fitting actual data. If ideal data is available (left), the left figure in middle row will show a relationship that trends down and to the right, therefore, MSY and effort proxy (Emsy guess) can be estimated in the center figure in the middle row. If the data does not follow such a relationship (right), then estimated Emsy will be negative, which indicates the possibility of a problem with the data time series. The original function used for these plots is https://github.com/DTUAqua/spict/spict, but the plots shown here were output by a slightly modified program (https://github.com/ichimomo/spict/spict).

A) The r used in the x-axis is for the Pella-Tomlinson type model



B) The r used in the x-axis is for the Schaefer type model (denoted as "rold")

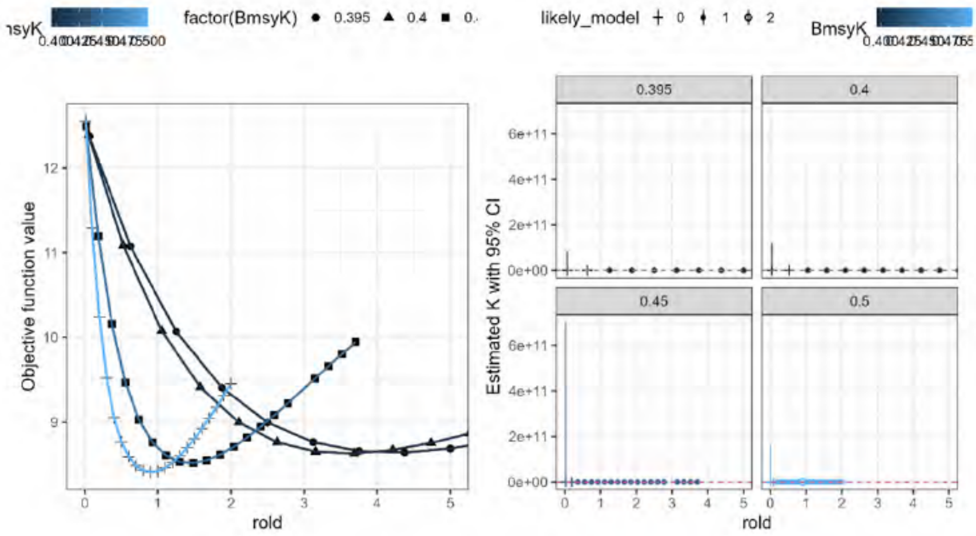
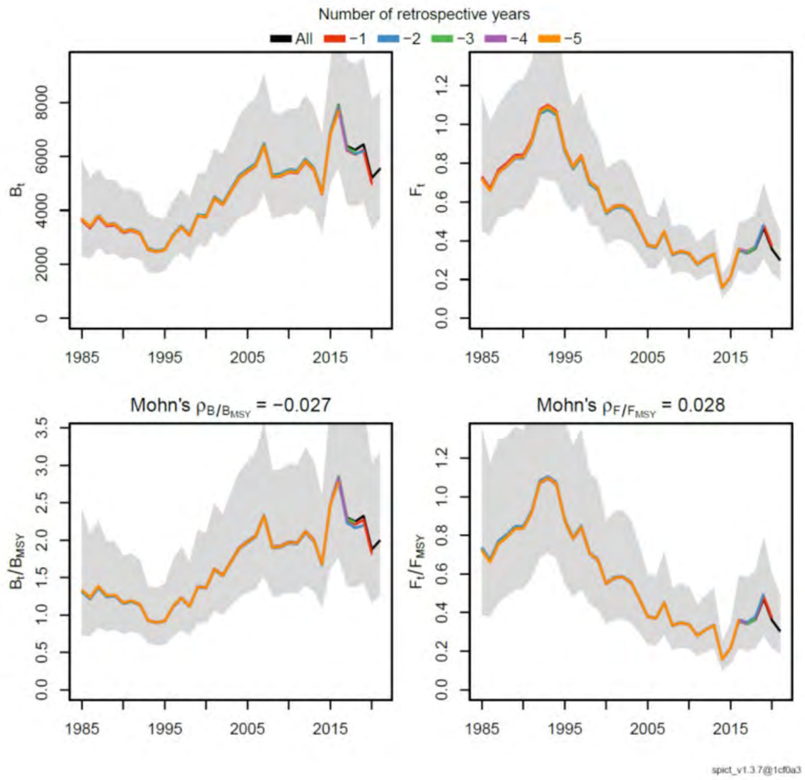


Fig. 2. Likelihood profile (Example for Northern Hokkaido Stock of Littlemouth Flounder). On the left, the vertical axis is the negative log-likelihood (smaller values make a better fit), and the horizontal axis is assumed *r* values. Different line symbols and colors indicate specific values for assumed Bmsy/K. On the right, the horizontal axis is *r*, the vertical axis is the carrying capacity *K* (and the 95% confidence interval), and the four panels are different assumptions for Bmsy/*K*. White circles represent models in which the minimum negative log-likelihood was obtained, black circles represent models in which the minimum likelihood difference was <2, and other models are not marked.

A)



B)

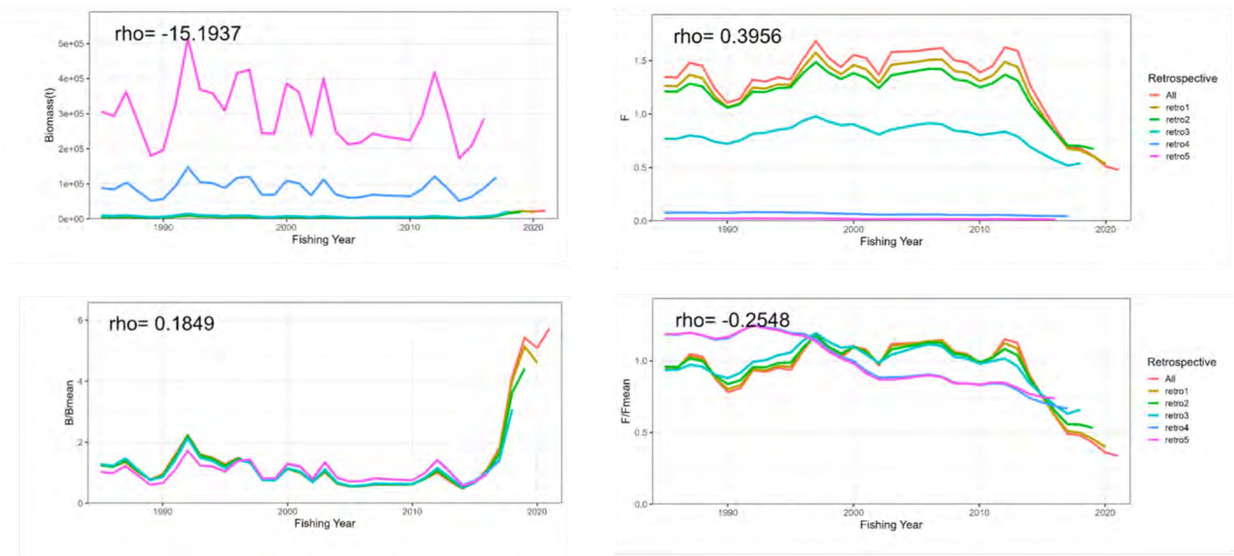


Fig. 3. Example of retrospective analysis. The top section A) shows examples without retrospective bias (Northern Hokkaido Stock of Sohachi Flounder). The bottom section B) shows examples with significant retrospective bias in three parameters (B, F, F/Fmsy), but without bias in B/Bmsy (bottom left) (Hokkaido Sea of Japan Stock of Pacific Cod).

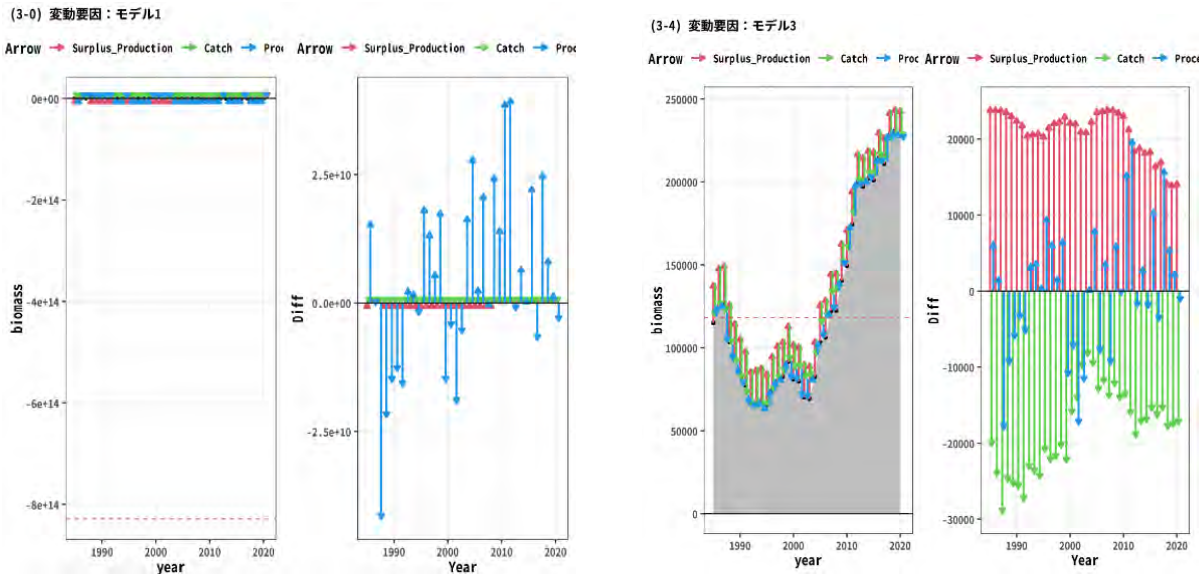


Fig. 4. Examples of factor stratification plots for stock fluctuations. On the left, the scale of stock size is so large compared to catch and other factors (colored arrows) that realistic estimates cannot be obtained. On the right, the scale of catch and other factors (colored arrows) are estimated in a realistic range compared to the scale of stock size, so there is a good balance with the process error(s). In these plots, red arrows represent surplus production, green arrows represent catch, and blue arrows represent the magnitude of the impact of process error(s).