Reliably Diagnosing Fraser River Sockeye Salmon Declines in a Variable World

by

Faye d'Eon-Eggertson

B.Sc., McGill University, 2008

Research Project Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Resource Management

Report No. 564

in the
School of Resource and Environmental Management
Faculty of Environment

SIMON FRASER UNIVERSITY

Spring 2013

Approval

Name: **Faye d'Eon-Eggertson**

Degree: **Master of Resource Management**

Title of Project: ***Reliably diagnosing Fraser River sockeye salmon declines in a variable world***

Report No.: **564**

Examining Committee:

Chair: **Aimée Brisebois**
Master of Resource Management Candidate

Dr. Randall M. Peterman
Senior Supervisor
Professor Emeritus
School of Resource and Environmental Management

Dr. Nicholas K. Dulvy
Supervisor
Professor
Department of Biological Sciences

Date Defended/Approved: January 17, 2013

Partial Copyright Licence

Abstract

There are numerous indicators that can be used to categorize populations as being a conservation concern, but these alternatives may vary in their reliability. To determine which of 18 quantitative decline indicators most reliably categorize the conservation status of a population, we used a stochastic model to simulate time series representing populations of sockeye salmon (*Oncorhynchus nerka*), allowing us to examine the effects of different levels of process variation and observation error. We examined whether each indicator's assessed status accurately predicted the subsequent trend of the population using a Receiver Operating Characteristic analysis as an integrated measure of reliability. Indicators that measured decline from a historical baseline were the most reliable. Indicators that measure decline over the most recent 3 generations, which are widely used, were sensitive to process variation and observation error. Our results suggest that, when available, longer time-series of abundance should be used to evaluate population decline.

Keywords: Decline indicators; sockeye salmon; IUCN; stochastic population dynamics model; Receiver Operating Characteristic; error trade-offs

Acknowledgements

I am grateful for funding to support this research, which was provided by Simon Fraser University and by grants to Randall M. Peterman from the Natural Sciences and Engineering Research Council of Canada and the Canada Research Chairs Program, Ottawa.

I would like to thank Randall Peterman for his support, guidance, thorough feedback, and enthusiasm. I would also like to thank Nick Dulvy for his expertise, and the Fisheries Science and Management Research Group, as well as the other students and staff in the School of Resource and Environmental Management for their support and camaraderie.

Table of Contents

List of Tables

List of Figures

1. Introduction

It is important to accurately rank which populations are most in need of protection, given the limited resources available to protect species and populations from going extinct. One of the most commonly-used worldwide classification systems for assigning species to categories of extinction risk was developed by the International Union for the Conservation of Nature (IUCN 2001). To date more than 60,000 animals, plants and fungi have been assessed using this approach. This classification system has been adapted by other agencies, for example, the Canadian Committee on the Status of Endangered Wildlife in Canada (COSEWIC 2011). These IUCN-like classification systems use various metrics, such as a high rate of decline in population size, a small geographic area, or small absolute numbers of organisms, to assess the level of extinction risk that a species is facing. In these cases, relative change in, or absolute sizes of, adult populations or geographic area is used to assign taxa to a threat category (IUCN 2001, IUCN 2010, COSEWIC 2011). These metrics can be viewed as measuring the symptoms of extinction (Mace et al. 2008), or at least serious problems, and can be used to provide information to managers about whether conservation action should be taken. Indicators that measure decreases in population abundance are especially widely used, in part because of available data, and are generally the preferred metric (Powles 2011, Wilson et al. 2011, Porszt et al. 2012).

However, declines in abundance can be challenging to measure, particularly in widely fluctuating populations such as Fraser River sockeye salmon (*Oncorhynchus nerka*) in British Columbia, Canada (Paulsen et al. 2007), where it can even be difficult to tell whether the underlying trend in abundance is a decline. The commonly used IUCN indicator of decline (criterion A) examines the reduction in population size over the longer of the most recent 10 years or three generations (where generation span is defined as the average age of adults) (IUCN 2001, COSEWIC 2011). If the decline is greater than a specified boundary, then the population is assigned to a conservation risk category. The reliability of this decline indicator (the likelihood of it correctly classifying a

population's status in various situations) remains largely uncertain and untested, although it is widely used to inform decisions regarding population conservation and management (Porszt et al. 2012). Decreases in population abundance could be measured using any of several metrics (e.g., decrease from some historical starting point, from some maximum abundance, or over some recent period), but these potential alternatives may vary substantially in their reliability. Unfortunately, a full analysis of reliability of such a wide range of indicators (or metrics) of abundance has not been conducted, although an initial analysis by Porszt et al. (2012) shows that certain indicators differ in their reliability. Obviously, in order to correctly assign a population to a conservation status, it is important to use the most reliable measures of decline (DeMaster et al. 2004, Keith et al. 2004, Regan et al. 2005, Wilson et al. 2011, Porszt et al. 2012). In this study, we sought to evaluate the reliability of a wide variety of indicators of decline in abundance (all of which could be viewed as potential alternatives to IUCN criterion A). Such evaluations of reliability are typically done to evaluate various proposed diagnostic tests in medicine.

Reliability of an indicator in the field of biological conservation reflects the probability of correctly and incorrectly categorizing a population as being at risk. If an indicator falsely suggests that a population is declining when it actually is not, then this is a false positive error. If an indicator falsely suggests that a population is not declining when it is, then this is a false negative error (Table 1). The costs and consequences of incorrectly classifying a population depend on the type of classification error. Scientific literature usually emphasizes the importance of avoiding false positive errors without examining the costs of making a false negative error, thus implicitly placing more weight on avoiding false positive errors than false negatives (Peterman 1990, Mapstone 1995, Field et al. 2004). However, in fisheries management, the costs associated with false negatives can often be at least as serious as false positives (Peterman 1990, Mapstone 1995, Field et al. 2004, Dulvy et al. 2006). For example, if a population is incorrectly classified as not being of conservation concern when it should be (false negative), then insufficient actions may be taken to protect this population from collapse or extinction. In the longer term, these false negative costs would also include the economic and social costs resulting from this reduction in population size (Peterman 1990). In contrast, if a population is incorrectly classified as being a conservation concern (false positive), this

may result in unnecessary action being taken to protect the population, such as fishing restrictions or remedial actions, which could result in a loss of revenue and employment, and a misdirection of funds away from populations that are in more dire need. Whether a false positive or a false negative is the more serious error depends on a manager's priorities and error tolerances. Different stakeholders will often have different error weighting preferences, because the costs of these errors might be borne by different groups (Peterman 1990). Classification errors may be impossible to eliminate, so there are often inherent trade-offs between the rates of false positive and false negative errors (Rice & Legacé 2007, Mace et al. 2008, Wilson et al. 2011), which we examine explicitly here.

Table 1. Possible outcomes when the indicator's assessed status of a population is compared to the estimated subsequent status of the same population.

		Indicator's assessed status	
		Non-declining	**Declining**
Subsequent status	**Non-declining**	True negative	False positive (*Type I error*)
	Declining	False negative (*Type II error*)	True positive

Previous analyses have evaluated decline criteria in terms of risk of extinction (e.g. Punt 2000). There is debate in the literature about the suitability of indicators of decline to correctly indicate extinction risk in marine fishes (e.g., Powles et al. 2000, Reynolds et al. 2005, Powles et al. 2011), and about the ability of fish populations to recover after a large reduction (Hutchings 2000, Hutchings 2001, Dulvy et al. 2003, Hutchings & Reynolds 2004). We chose not to focus on the extinction risk per se, but instead to examine whether status assessment diagnoses appropriately forewarn of subsequent population declines. In addition, major decreases in some sockeye salmon populations are economically, socially, and ecologically important (DFO 2005), even if the extinction risk is low.

To evaluate the performance of indicators of significant population decline, Porszt et al. (2012) conducted a retrospective analysis using historical data of Fraser

River sockeye salmon to determine which of a wide range of quantitative indicators of time trends in abundance were most reliable at indicating future trends in abundance that subsequently occurred. They found that indicators that measured the extent of decline from a historical baseline tended to better reflect a population's status in subsequent years than the commonly used IUCN indicator that measures the rate of decline over the previous 3 generations. A key limitation of the Porszt et al. (2012) retrospective analysis is that it only evaluates the criteria over one set of historical events or situations, that is, the actual environmental conditions that occurred, the state of other ecosystem components over that period, and management decisions that were made. One way around these caveats is to use simulation analysis for reliability testing. In a different and more generalizable approach, Wilson et al. (2011) used stochastic simulation modeling to assess three methods for detecting population decreases: measuring decline between two point estimates of abundance, use of linear regression on a time series of abundance, and state-space models. Their stochastic model estimated the error rates that would be generated by those methods. They found that linear regression and state-space models both had a low proportion of falsely detected declines, i.e., false positives (3-14%); however, a high percentage (33-75%) of small-magnitude declines in abundance were not detected (false negatives). They found that using two point estimates of abundance generated higher power (95%) to detect small declines than the other methods, but there was a high percentage (50%) of false detections. These studies illustrate the need to empirically and quantitatively assess the reliability of threat indicators before using them to inform managerial or policy decisions (Porszt et al. 2012).

To quantify the relative reliability of decline indicators, i.e., identify those with the highest probability of correctly classifying a population's status as decreasing or not, we used a simulation modelling approach, as has been done in similar situations by other researchers (Punt 2000, Holt et al. 2009, Regan et al. 2009, Wilson et al. 2011). Our approach extends the Wilson et al. (2011) method by examining a larger number of indicators and conducting a full analysis of false negative and positive rates across a large range of conditions (different levels of process variation and observation error) and thresholds for classifying a population as being a conservation concern. Evaluating the reliability of decline indicators is relevant for all taxa, but we based our model in part on

Fraser River sockeye salmon as a representative case study for which identifying and using the most reliable indicators is particularly important. Sockeye salmon are commercially-valuable, and many stakeholders are highly invested in ensuring the best decisions are made, which makes them a relevant case study not only for the purposes of species-at-risk listings, but also for maintaining their economic and social value (DFO 2005). Pacific salmon exhibit high variability in abundance over time due to inherent stochasticity in the system (process variation), and high variability can increase the probability of extinction of a population (Paulsen et al. 2007). Another problem with evaluating whether a population is decreasing is that estimates of spawner abundance are often inaccurate, due to imprecision and inadvertent bias in estimates of population size (observation error) (Rand 2011). Our simulation modelling approach allowed us to explore how a range of alternative indicators of population decrease perform across a wide range of process variation and observation error that might exist in the future, which is not possible to examine with a retrospective analysis of historical data.

In medicine, a Receiver Operating Characteristic (ROC) analysis is often used to evaluate the performance of diagnostic tests (Hibberd & Cooper 2008). ROC analyses combine the error rates produced by a test or indicator over different thresholds into a single measure. This type of analysis has recently been used in extinction-risk studies (e.g. Porszt et al. 2012), and can be used as an integrated measure of reliability to evaluate the indicators of symptoms of extinction risk. We used an ROC analysis to evaluate the reliability of decline indicators, but, because ROC analyses inherently weight false positives and false negatives as equally important, we also examined which indicators were the most reliable across different levels of weightings on those two types of errors.

Thus, the purpose of our study was to determine which quantitative indicators of time trends in abundance can be used to most reliably categorize the conservation status of a population. We also sought to examine how robust these indicators were to different magnitudes of process variation, observation error, and boundaries of conservation concern (the percentage decline in abundance beyond which a population is considered a conservation concern). We explored how different management error weightings would affect the ranking of alternative decline indicators, while explicitly examining the trade-offs between the error types. Our goal was to find the best way to

use the available data in order to correctly assign a conservation status to a population, and to illustrate the importance of evaluating indicators using simulation models for testing their relative reliability.

2. Methods

2.1. Stochastic population dynamics model

2.1.1. Model

We used a stochastic simulation model to simulate sockeye salmon time series. We used a spawner-to-spawner model rather than a spawner-to-recruit model because we were interested in how many fish made it back to spawn, representing a full life cycle, rather than how many were available for recruitment into the fishery. In this spawner-to-spawner model, the abundance in any given year was determined by the abundance of spawners four years prior, which was then input to a Ricker stock-recruitment function that included a stochastic productivity parameter, process variation term, and observation error term. This model is consistent with the reproductive life history of the semelparous sockeye salmon in the Fraser River, which has four distinct cycle lines arising from over 92% of the salmon reproducing and dying at four years of age with little gene exchange among the lines (Ricker 1997, Porszt et al. 2012). We set the equilibrium population size as 100,000 fish. The initial spawning population abundances were drawn randomly from a uniform distribution between 20,000 and 80,000 fish for each of the four cycle lines (as in Dorner et al. 2009). The model had an initialization period of 12 years, including the initial four years. We then analyzed each time series during a 52-year "evaluation period" after that initialization period. Following the evaluation period was a "subsequent period" of 12 years (Figure 1). The 52-year evaluation period is comparable to the length of historical time series available for Fraser River sockeye salmon (Porszt et al. 2012). We ran 500 stochastic simulations for each scenario (i.e., for each level of process variation, observation error, etc.); scenarios represented a range of different population productivities, environmental conditions, and harvest rates.

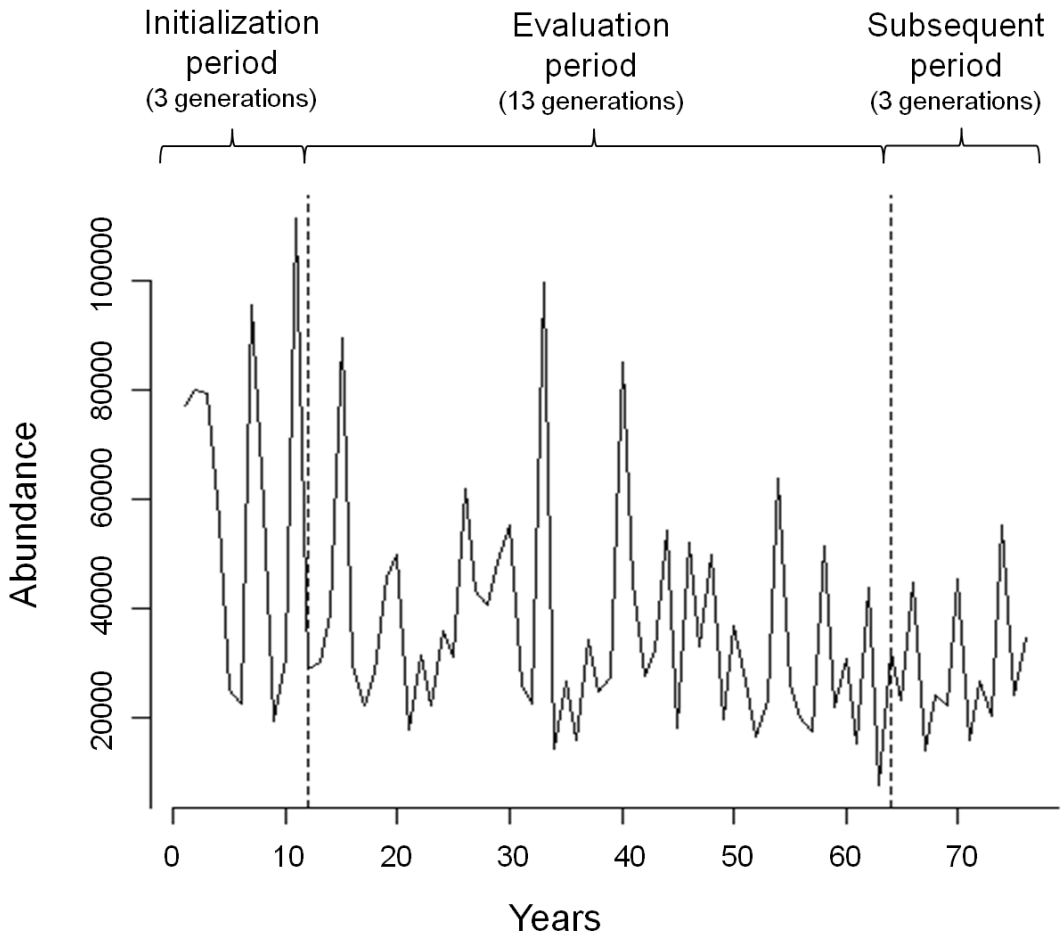


Figure 1. ***An example of periods in the time series output produced by the model.***

2.1.2. Productivity

Our model used a Ricker recruitment function of the form $S_t = aS_{t-1}e^{-bS_{t-1}}$, where *t* is the brood year. The productivity parameter (*a*) represents the slope at the origin, which is the maximum number of adults that return to spawn as a ratio to the spawners in the previous generation, in the absence of density dependence at low stock sizes. In this spawner-to-spawner model, this productivity parameter encompasses productivity from spawner-to-adult recruits prior to the onset of fishing, as well as the harvesting process and in-river pre-spawning mortality (Appendix 1). Each population (i.e., each Monte Carlo trial) was randomly assigned a productivity parameter value, representing differences in productivity among different populations due to local variables such as

harvest rate, habitat quality, and environmental conditions. Based on a realistic range observed for Fraser River sockeye salmon, we used a normally distributed productivity parameter, with a mean of 1 and a standard deviation of 0.3 (but constrained to not drop below zero) to ensure that some of the populations were declining and some were non-declining (Appendix 1). The density-dependent parameter (*b*) describes how quickly the number of returning spawners decreases as the number of spawners in the previous generation increases. We scaled *b* to 1 divided by equilibrium abundance, so in this case, 1/100,000 (Dorner et al. 2009).

2.1.3. Process variation

Process variation is the random interannual variability in productivity due to inherent stochasticity in salmon biology and environmental conditions (Hilborn & Walters 1992, Wilson et al. 2011). Process variation was added to the model with a multiplicative error term e^u (Walters & Ludwig 1981), where *u* was normally distributed with mean zero and a variance σ_u^2, and it varied for every year of each trial. Process variation can be quite high in Fraser River sockeye salmon (Paulsen et al. 2007), and we examined levels of σ_u^2 of 0.01, 0.05, 0.1, 0.3 and 0.5. With process variation included, the model becomes: $S_t = aS_{t-1}e^{-bS_{t-1}+u}$. Because the abundance with the inclusion of process variation still represents the "true" number of spawners (Wilson et al. 2011), this error is propagated through the model (i.e., the abundance of spawners in time *t* is based on the abundance of spawners in time *t-1* with process variation included), and is reflected in all periods of the time series.

2.1.4. Observation error

Observation error is the apparent interannual variability in abundance due to imprecision and inadvertent bias in estimates of population size, such as counting error and sampling error (Paulsen et al. 2007, Rand 2011, Wilson et al. 2011). Observation error was added to the model with a multiplicative error term e^v, where *v* was normally distributed with mean zero and a variance of σ_v^2 (Walters & Ludwig 1981). With observation error included, the full model was: $S_t = aS_{t-1}e^{-bS_{t-1}+u+v}$. The level of observation error varied stochastically for each year in each Monte Carlo trial, but was

added independently at each simulation step and thus did not propagate through the model, i.e., the abundance of spawners in time *t* was based on the "true" abundance of spawners in time *t-1* with process variation included but without observation error (Wilson et al. 2011). In actual stock assessments, levels of observation error can be quite high: "occasional errors of [a factor of 2 to 4] magnitude would not be considered unusual in most fisheries stock assessments" (Walters & Ludwig 1981). In sockeye salmon, there is "an unknown but high degree of random observer error" (Rand 2011). We thus examined variances of the observation error term (σ_v^2) of 0, 0.05, 0.1, 0.3 and 0.5 in different scenarios, which encompass the observed range of variability (e.g., Figure 2).

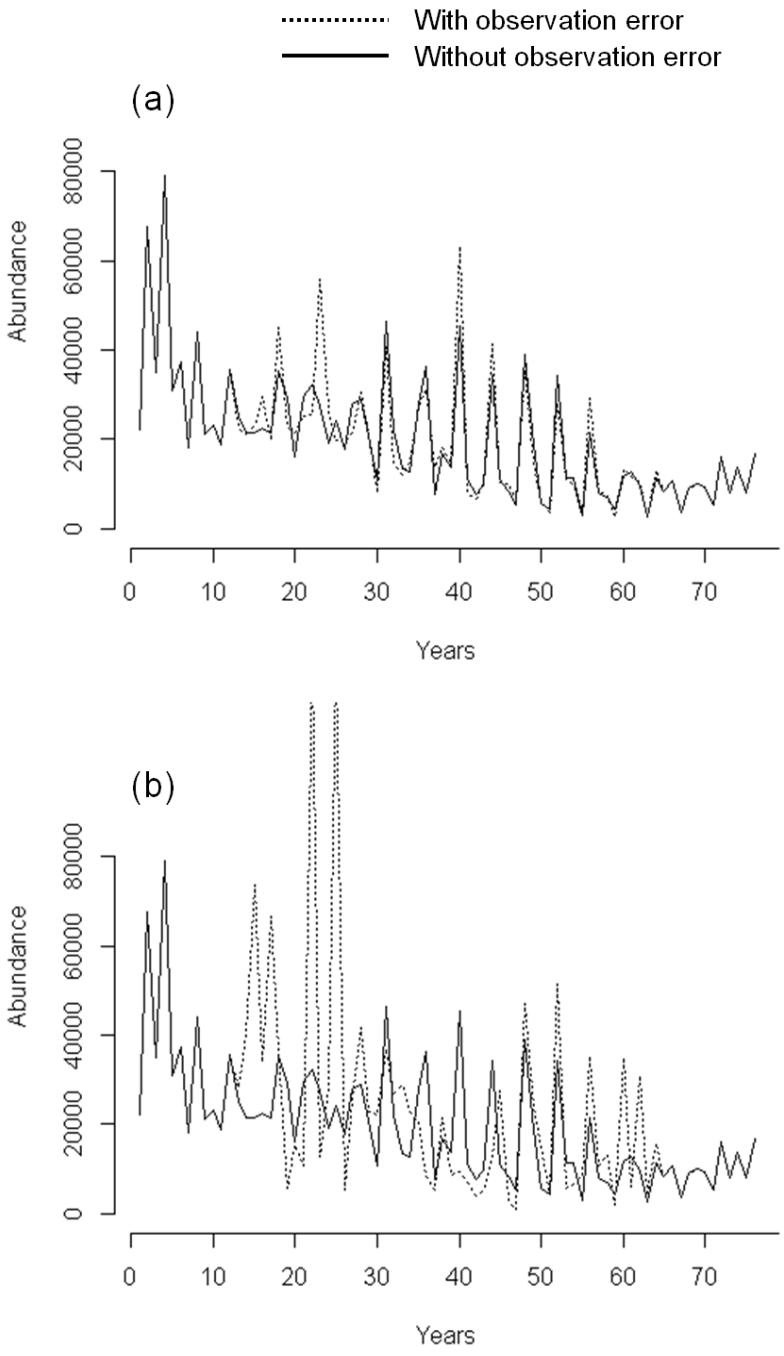


Figure 2: ***Examples of simulated population abundance produced by the model, with (dotted) and without (solid) observation error at levels of variance of the error term (σ_v^2) of 0.05 (a) and 0.5 (b).***

2.2. Threat indicators to estimate status during the "evaluation period"

We used 18 threat indicators of time trends in adult abundance to assess whether our populations were declining or non-declining. These indicators used different combinations of periods evaluated (either the most recent three generations, or from the beginning of the time series, or from the maximum abundance anywhere in the time series), smoothed vs. unsmoothed data, and transformed data (raw, $\log_e$-transformed, or generational means). Changes in abundance were measured using either regression or a percent decline from the baseline. Our indicators are similar to those used in Porszt et al.'s (2012) retrospective analysis; they are summarized in Table 2 and described in detail in Appendix 2.

Periods – Changes in abundance were measured either over the most recent three generations (12 years), as is done for criterion A by COSEWIC (2011) and IUCN (2001), or as a long-term change in abundance from some historical baseline level (as suggested by Mace et al. [2002] and Holt et al. [2009]). The historical baselines we used consisted of abundances occurring early in the time series (i.e., either the first year in the evaluation period, or the first year in the cycle line by examining each cycle line separately as opposed to all cycle lines together, or the geometric mean abundance of the first 4-year generation), or the maximum abundance anywhere in the evaluation period (i.e., maximum abundance in a single year or maximum geometric mean abundance of any 12-year or 3-generation period).

Smoothing – Abundance estimates were either smoothed with a 4-year (1-generation) running mean or were left unsmoothed.

Data transformations – Abundance estimates were taken from either raw values, $\log_e$-transformed values, or the geometric mean abundance of 4-year generations where the generations either moved one year at a time in sliding windows or in 4-year blocks with no overlapping of years.

Changes in abundance – Changes in abundance were measured using either robust linear regression over the designated period to minimize the influence of outliers

(Venables & Ripley 2002), or as a percent decline from a baseline to the current year or generation being evaluated.

Table 2. A summary of the 18 decline indicators used to assess population status.

Indicator	Periods[1]	Smoothed?	Transformation[2]	Change in abundance[3]
1	Recent 3 gen	No	Log_e	Regression
2	Recent 3 gen	Yes	Log_e	Regression
3	Hist: first year	No	Log_e	Regression
4	Hist: first year	Yes	Log_e	Regression
5	Hist: cycle year	No	Log_e	Regression
6	Hist: cycle year	Yes	Log_e	Regression
7	Max: single year	No	Log_e	Regression
8	Max: single year	Yes	Log_e	Regression
9	Hist: first gen	No	Mean: SW	% decline
10	Hist: first gen	Yes	Mean: SW	% decline
11	Hist: first gen	No	Mean: GB	% decline
12	Hist: first gen	Yes	Mean: GB	% decline
13	Max: 3 gen	No	Mean: SW	% decline
14	Max: 3 gen	Yes	Mean: SW	% decline
15	Max: 3 gen	No	Mean: GB	% decline
16	Max: 3 gen	Yes	Mean: GB	% decline
17	Hist: first gen	No	Raw	% decline
18	Max: 3 gen	No	Raw	% decline

[1]Recent 3 gen: Percent decline over the most recent 3 generations. Hist: Percent decline from the beginning of the time series (either the first year, mean of the first generation, or the first corresponding cycle year in the time series). Max: percent decline from maximum abundance in the time series (either single year or 3-generation period).

[2]SW: Generations moved one year at a time in sliding windows. GB: Generations moved in 4-year blocks with no overlap of years (i.e., status only assessed every four years).

[3]Regression: Changes in abundance were measured using robust linear regression over the designated period. % decline: Changes in abundance were measured as a percent decline from the baseline to the current year or generation being evaluated.

We evaluated these 18 indicators in each applicable year of the evaluation period of the time series. For each year, we determined whether the indicator categorized the population as “declining” or “non-declining” for each of 101 different threshold levels, ranging from 0-100% decline in increments of 1%. These threshold levels were used to classify status during the 13-generation (52-year) evaluation period (Figure 1). The assessed status during the evaluation period was then compared to the status in the subsequent period (Figure 1), as defined in the next section. Specifically, we compared whether the decline from the initialization to the subsequent period was greater than the boundary condition for subsequent status that indicated that the population was a conservation concern.

2.3. Subsequent status

For our analysis, we were concerned with how well our indicators predicted the longer-term trajectory of the population. We assumed that if populations declined in the past and continued to decline, they had a greater chance of extinction than populations that did not continue to decline (Mace et al. 2008, Porszt et al. 2012). To determine whether the population had declined and was going to continue to decline, we evaluated the “subsequent status” of the population, which we estimated as the decline from the initialization period through the subsequent period. We took a robust linear regression of the final 12 years in the time series (i.e., years 65-76, called the “subsequent period” – see Figure 1) and found the mean spawner abundance of the last generation (i.e., across the last 4 years, 73-76), as estimated from this regression. We found the mean spawner abundance at the end of the initialization period in the same way (i.e., the mean of years 9-12, calculated by robust linear regression of the first 12 years), and used the percent decline from the end of the initialization period to that last generation in the “subsequent period” as the estimate of subsequent decline in the population. If this reduction was greater than the boundary condition that indicated that the population was a conservation concern, the estimate of subsequent status was classified as “declining”, otherwise it was classified as “non-declining”. We examined the effects of different boundaries of estimated subsequent status which indicate that a population is a conservation concern. We used subsequent status boundaries of 90%, 70% and 50% decline. These are the levels of decrease that are used by COSEWIC and IUCN to

assign populations to a conservation status. A population that declines by ≥90% would be classified as “critically endangered”, by ≥70% would be “endangered”, and by ≥50% would be “threatened” or “vulnerable” (IUCN 2001, IUCN 2010, COSEWIC 2011).

We wanted to identify which indicators most reliably reflected the “true” subsequent outcome, therefore the “true” future time trend in abundance that we used to compare to an indicator’s assessed trend only reflected process variation and not observation error. Observation error was not included in the initialization and subsequent periods of the time series because for these periods, the “true” population abundance was considered as being known for the purposes of out simulations.

2.4. Measuring reliability

2.4.1. Receiver Operating Characteristic analysis

We used a Receiver Operating Characteristic (ROC) analysis to compare the reliability of the 18 different indicators. For our purposes, reliability was the ability of the indicator to correctly distinguish between a declining and non-declining population, as determined by the subsequent status. An ROC analysis provides an integrated measure of reliability, by combining into a single metric the true and false positive rates produced by a threat indicator across a wide range of thresholds for classifying status of a population (Hibberd & Cooper 2008, Porszt et al. 2012). For a given threat indicator under a given scenario, we compared the indicators’ assessed status in each applicable year in the evaluation period of the time series to the estimated subsequent status of the population, resulting in either a true negative (TN), true positive (TP), false negative (FN) or false positive (FP) outcome (see Table 1). For each indicator, these different categories of outcomes were tallied over all applicable years and over all 500 trials. For each threshold level (0-100%), the true positive rate (TP/(TP+FN)) was plotted against the false positive rate (FP/(FP+TN)). In this way, the true positive rate and the false positive rate for each threshold level generated a point on the ROC curve. The area under this resulting curve (AUC) can then be calculated, which reflects the ability of a threat indicator to correctly distinguish between two states (Hibberd & Cooper 2008), and can be used as a measure of overall reliability, allowing us to rank our threat

indicators. An AUC value of 1 indicates that the indicator has a perfect ability to distinguish a declining from a non-declining population, whereas an AUC value of 0.5 (falling near the 1:1 line) indicates a 50/50 chance level of distinguishing a declining from a non-declining population (Figure 3). The higher the AUC value is, the more reliable the indicator. ROC analyses are commonly used in medicine to determine the reliability with which test correctly diagnose a condition (e.g. Hibberd & Cooper 2008). Similarly, this analysis is used here to evaluate the reliability of threat indicators at identifying reduced population size, which can be viewed as a symptom of extinction for a population. Such analyses rank indicators relative to each other (Porszt et al. 2012).

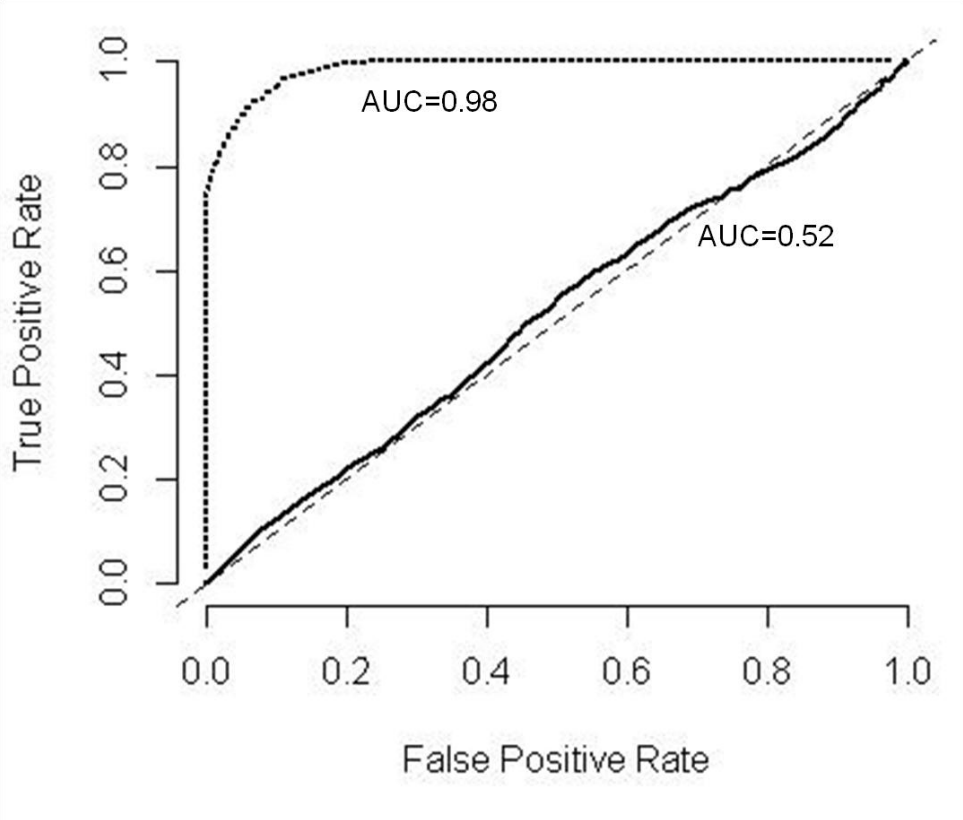


Figure 3. ***Two sample ROC curves for hypothetical indicators across threshold values of 0-100%. The ROC curves have an area under the curve (AUC) of 0.98 (dotted line) and 0.52 (solid line).***

2.4.2. Different error weightings and error tolerances

One limitation of an ROC analysis is that it inherently attributes equal weighting (i.e., relative importance to a decision) to the different types of errors (false positives and

false negatives), which is unrealistic in many situations in resource management (Peterman 1990, Mapstone 1995, Field et al. 2004, Dulvy et al. 2006, Porszt et al. 2012). In any given situation, managers could evaluate the costs of these two types of errors and then decide on an acceptable weighting for each (Peterman 1990, Mapstone 1995). Therefore, we also examined the reliability of the different threat indicators for each of several different error weightings. To do this, we examined, across different thresholds, the error rates for each of the two types of classification errors. The results showed which threat indicators gave the lowest error rates if one type of error was considered more important than the other.

Alternatively, rather than thinking in terms of a relative weighting of the error types, managers might have a maximum tolerable rate for a certain type of error. For example, a manager might say that any indicator with a greater-than-10% chance of a false positive is unacceptable due to the costs that would be incurred by that error. We therefore also examined the lowest error rate of a particular type that could be obtained if the other error rate was constrained to be below a certain value.

2.5. Sensitivity analyses

We performed sensitivity analyses in order to evaluate the degree to which our main findings were affected by changing levels of process variation (σ_u^2=0.01, 0.05, 0.1, 0.3, 0.5) and observation error (σ_v^2=0, 0.05, 0.1, 0.3, 0.5). This range of values encompasses low to high levels of each type of error, which are relevant because it can be difficult to know how much of the high apparent variability in Pacific salmon is due to either observation error or process variation (Paulsen et al. 2007, Rand 2011). Dorner et al. (2009) used a combined error variance (process variation and observation error) of 0.55 so that their data approximated the overall interannual variability observed in the 37 North American sockeye stocks that they used. Our range of values for process variation and observation error that we explored encompasses this level of variance. Our sensitivity analyses thus evaluated how robust the different threat indicators were to various situations that reflect potential real-world conditions.

We also examined different boundaries that were used to indicate a declining subsequent status for the population (90%, 70%, 50%), which are representative of thresholds of endangerment used to classify populations as at-risk. As well, we explored different values of the *a* (1, 1.4, 1.8) and *b* (1/100000, 1/200000, 1/400000) parameters to see whether our results were sensitive to these variables.

3. Results

At low levels of both process variation and observation error, all 18 indicators were almost equally reliable at discriminating between declining and non-declining populations (AUC values >0.9) (Figure 4a). Across all analyses of different levels of process variation and observation error, indicators that used a historical baseline based at the beginning of the time series (indicators #4, 6 and 9) consistently outperformed the other indicators, and ranked in the top 5 indicators across all scenarios (Figures 4, 5, and Appendix 3). In general, indicators calculated from a historical baseline were more reliable than either indicators with a baseline based on the maximum abundance anywhere in the time series or indicators based on decline over the most recent 3 generations, including the commonly used IUCN criterion A (which corresponds to indicator #2 in our study).

The indicators varied in how robust they were to increases in process variation and observation error. All of the indicators were somewhat sensitive to process variation. The AUC of the majority of the indicators decreased by ~0.1 when σ_u^2 (process variation) was increased from 0.01 to 0.5 (Figures 4a, b, 5a). The AUC of indicators #1, 2, 5, 17, and 18 were the most reduced by increases in process variation, with around twice the decrease in reliability exhibited by the other indicators (drop in AUC of ~0.2 when σ_u^2 was increased from 0 to 0.5) (Figures 4a, b, 5a, and Appendix 3). The latter more sensitive indicators included the indicators that evaluate the recent rate of decline (#1 and 2), the indicators that use raw abundance (#17 and 18), and one indicator based on decline from the first corresponding cycle year (#5). In contrast, most of the indicators were relatively insensitive to increases in *observation error*, exhibiting only slight decreases in reliability (drop in AUC of ~0.05 when σ_v^2 was increased from 0 to 0.5) (Figures 4a, c, 5b). The exceptions to this trend were indicators #1 and 2 (which are both based on a rate of decline over the most recent 3 generations), which were highly sensitive to observation error (drop in AUC of ~0.2 when σ_v^2 increased from 0 to 0.5) (Figures 4a, c, 5b, and Appendix 3).

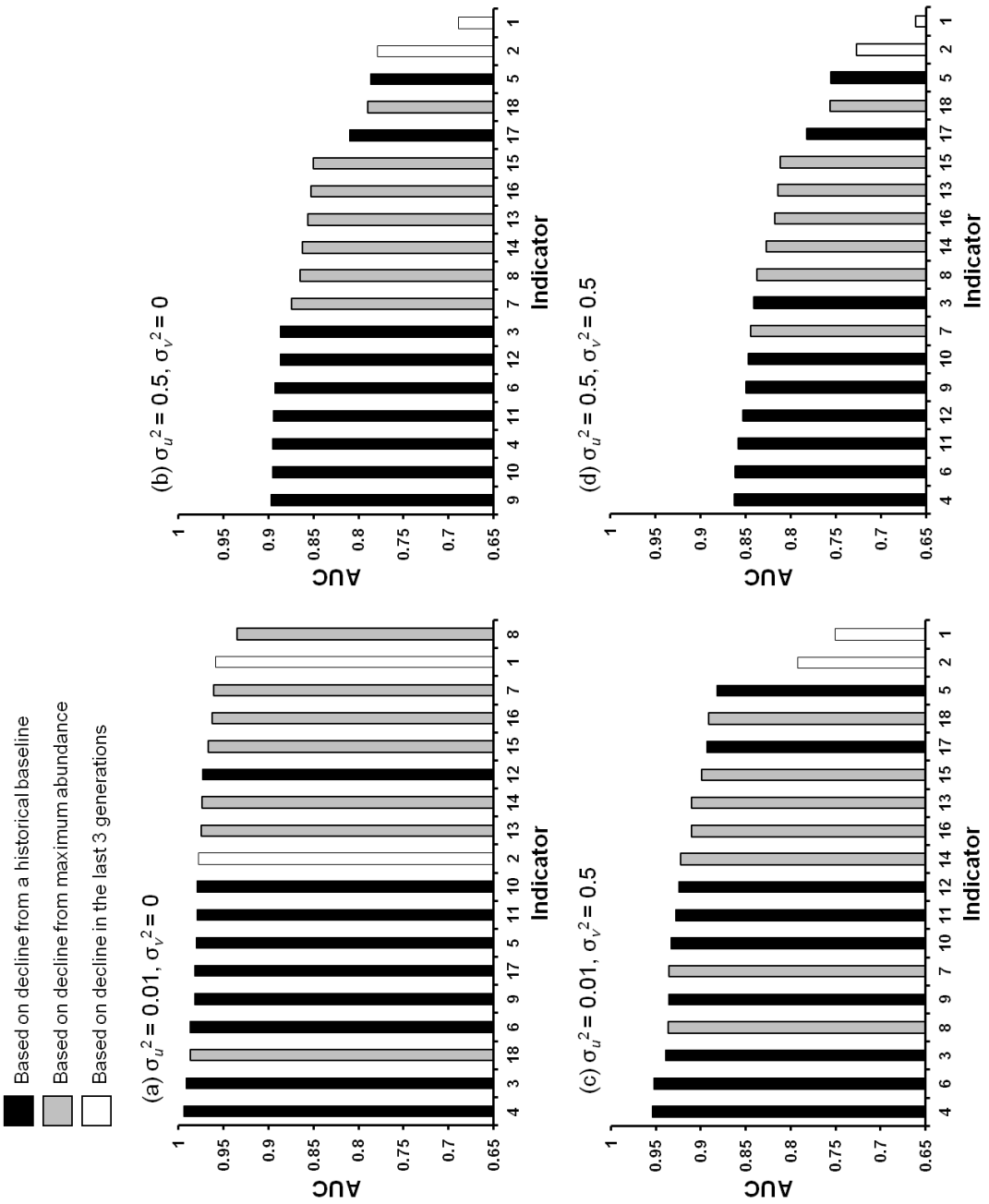


Figure 4. ***Ranking of indicators by area under the ROC curve (AUC), for (a) low process variation (σ_u^2=0.01) and no observation error (σ_v^2=0), (b) high process variation (σ_u^2=0.5) and no observation error (σ_v^2=0), (c) low process variation (σ_u^2=0.01) and high observation error (σ_v^2=0.5), and (d) high process variation (σ_u^2=0.5) and high observation error (σ_v^2=0.5). Indicators based on the decline in the last three generations are white, indicators based on decline from a historical baseline are black and indicators based on decline from maximum abundance are grey.***

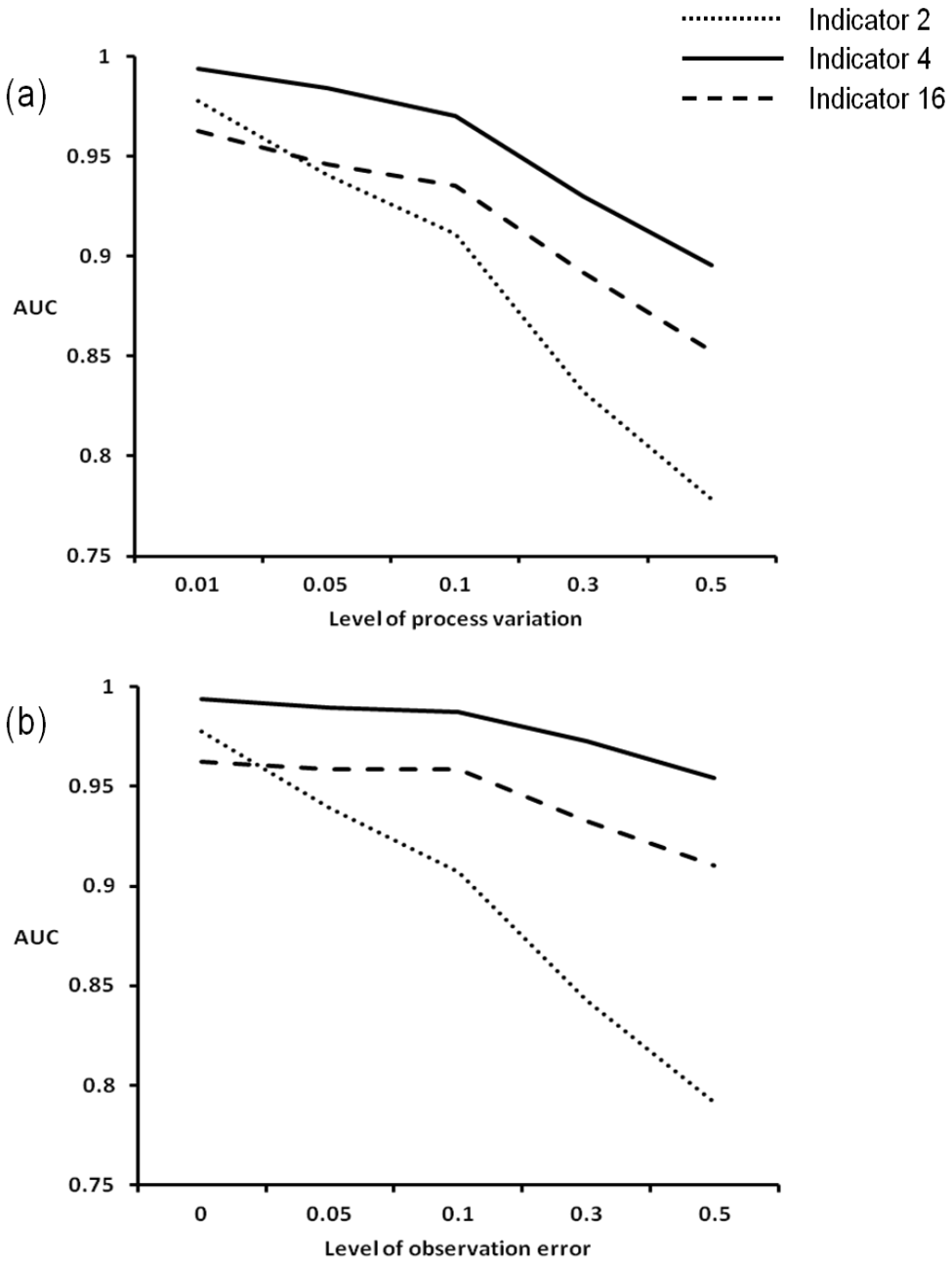


Figure 5. ***The area under the curve (AUC) values for selected indicators across increasing levels of (a) process variation and (b) observation error. The indicators are a high-reliability indicator based on a historical baseline at the beginning of the time series (indicator 4 [solid line]), a medium-reliability indicator based on a maximum abundance baseline (indicator 16 [dashed line]), and the commonly used IUCN indicator based on decline over the most recent three generations (indicator 2 [dotted line]).***

The AUCs of all indicators were relatively insensitive to changes in the boundary used to define a declining population in the "subsequent period" (AUCs changed by <0.04). However, the indicators that changed the most were #1 and 2 (Figure A4 of Appendix 3). The AUCs of all indicators declined with increases in the *a* parameter, however, the relative rankings of our indicators were relatively insensitive to changes in the *a* and *b* parameters (Figure A5 of Appendix 3).

3.1. Error weightings

There are inherent trade-offs between false positive and false negative error rates (as shown in Figure 6 for an example indicator #16), but there is a point (a threshold percentage decline for classifying a population as declining) at which the error rates are equal. As well, if one type of error is considered more important than the other, managers can specify a desired weighting or ratio of the error rates (e.g., that the false negative rate must be half the false positive rate) (Figure 6). We found that the rank order of an indicator's reliability was fairly robust to different error weightings, with the top-ranked indicators #4, 6 and 9 (indicators that used some historical baseline) being the best indicators across the majority of error weightings that we explored, across different levels of observation error (Table 3). The differences in the best-performing indicator across these error weightings are likely due to minor chance variations among closely ranked high-reliability indicators.

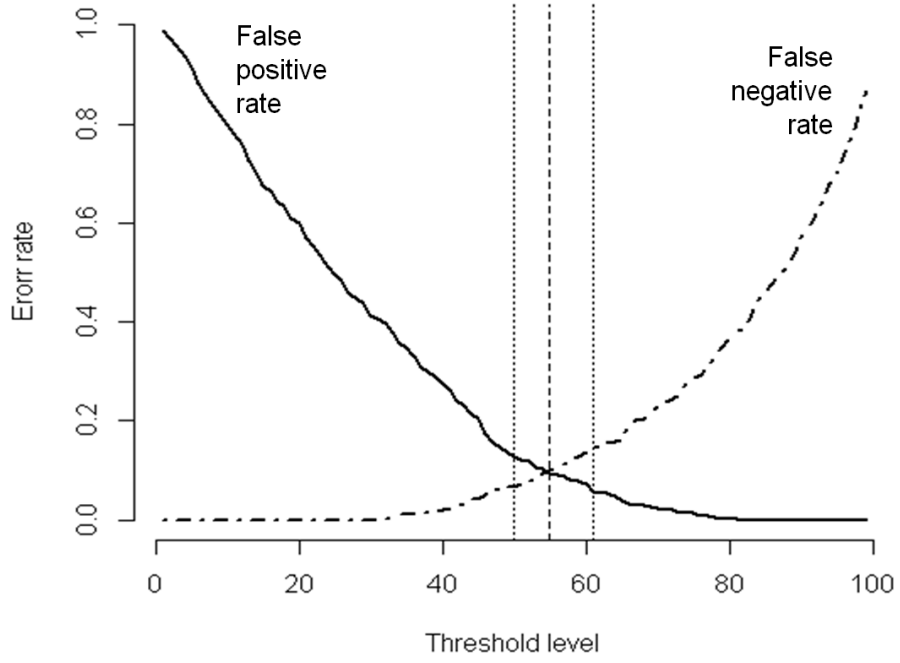


Figure 6. ***An example of the false positive rate (solid) and false negative rate (dot-dash) across threshold levels for classifying a population as declining ranging from 0-100% for indicator #16 for the base-case scenario with low process variation and no observation error (σ_u^2=0.01, σ_v^2=0). The black dashed vertical line indicates the threshold where both error rates are equal (i.e., the lowest rate of both error types if they are considered equally important). The dotted vertical lines indicate the lowest error rates if a false negative is considered twice as important as a false positive (left vertical dotted line, i.e., that the false negative rate must be half the false positive rate), and vice-versa (right vertical dotted line).***

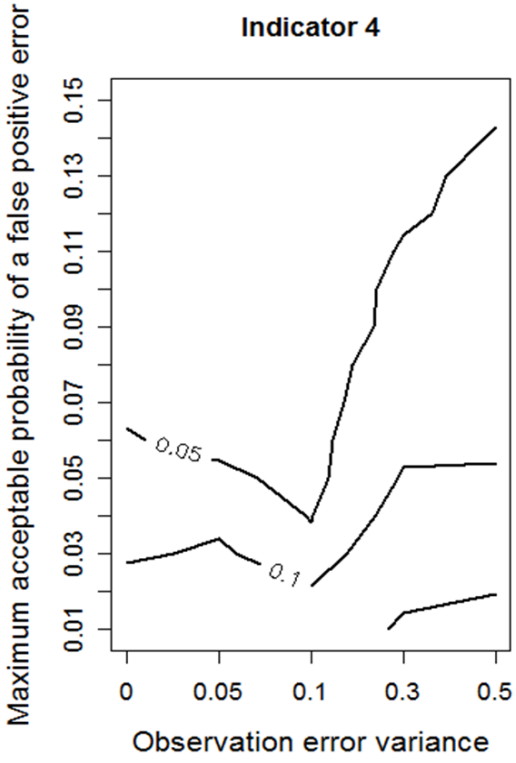
Table 3. The decline indicator number with the specified ratio of error rates of false positives (FP) to false negatives (FN) for different levels of (a) process variation and (b) observation error. Indicator numbers are defined in Table 2 and Appendix 2.

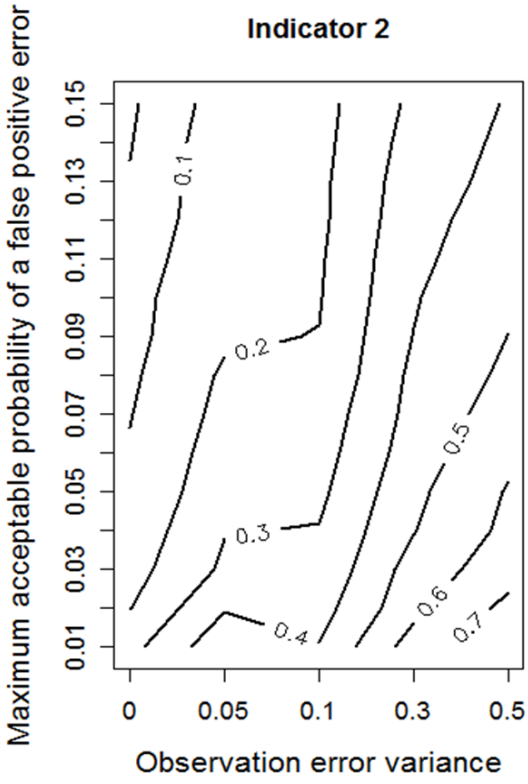
	"Best" indicator if FP is *x* times as important as FN.								
(a) Process variance (σ_u^2)	***x:*** **0.20**	**0.25**	**0.33**	**0.50**	**1.00**	**2.00**	**3.00**	**4.00**	**5.00**
0.01	4	4	4	4	4	4	4	4	4
0.05	4	4	6	6	6	9	9	6	10
0.1	4	6	4	4	6	11	11	6	6
0.3	4	6	6	11	9	10	6	7	7
0.5	11	10	10	9	4	6	10	8	8
(b) Observation error variance (σ_v^2)	"Best" indicator if FP is *x* times as important as FN.								
	x: **0.20**	**0.25**	**0.33**	**0.50**	**1.00**	**2.00**	**3.00**	**4.00**	**5.00**
0	4	4	4	4	4	4	4	4	4
0.05	4	4	4	4	4	4	4	4	7
0.1	4	4	4	4	4	3	3	4	4
0.3	4	4	4	4	4	4	4	7	7
0.5	4	4	4	4	4	4	4	4	4

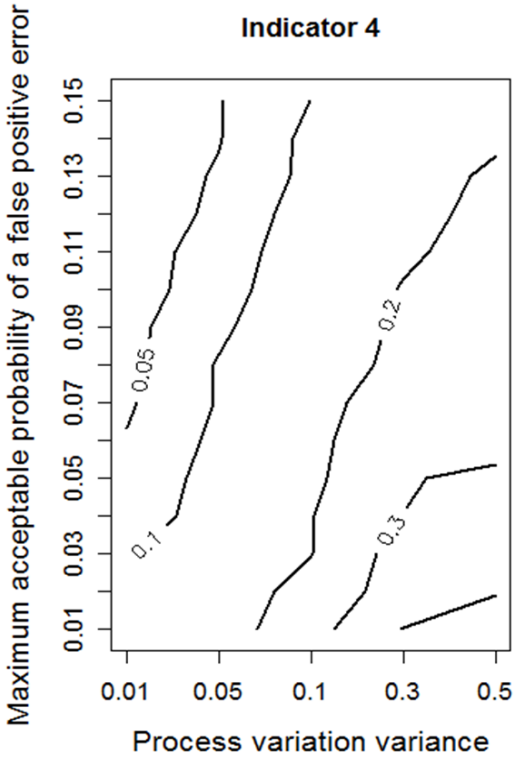
3.2. Error tolerance

Managers might have absolute values of rates for a certain type of error that they will not tolerate. For any given indicator, these tolerance levels can be mapped out to show the trade-offs between the error rate of the constrained type of error and the resulting error rate of the other type of error. For example, contours in Figure 7 show the resulting false negative rate if the false positive rate is constrained to values between 1 and 15% across several magnitudes of process variation and observation error for the high-ranking indicator #4 (Figures 7a, c) and the commonly used IUCN decline indicator #2 (Figures 7b, d). For example, if managers using that IUCN decline indicator #2 decided that it would be unacceptable to have greater than a 10% chance of a false positive error (i.e., incorrectly concluding that a population was declining), and if the

stock they were managing had a high level of observation error (say σ_v^2=0.3), then they could expect a false negative rate of about 39% (Figure 7c). This means that if the population is actually not declining, there will be less than a 10% chance of the indicator reporting that it is declining, but if the population is actually declining, there will be about a 39% chance of the indicator reporting that it is not declining. If managers want a lower false negative rate than that, they would have to increase their tolerance for false positives, decrease observation error, or pick a different indicator. For these same constraints, indicator #4 would give a much more desirable false negative rate of about 6% (Figure 7a). If someone wanted an even smaller chance of a false positive error (say 1%) under the same scenario, then they could expect around a 65% chance of a false negative error using indicator #2 or around a 21% chance of a false negative error using indicator #4. We did this analysis on other selected indicators to inform managers about the implications of their often-unstated error tolerances and to allow for visualization of the trade-offs (Figure A6 of Appendix 3). This type of analysis could be used to select the most desirable indicators, if managers had a maximum acceptable error rate for one type of error. We found that the trade-off characteristics of the 18 indicators were consistent with their overall performance. In other words, the top-ranked indicators (i.e., those based on a decrease from some historical baseline) based on AUC values (Figures 4 and 5) also showed the lowest range of probability of false *negative* errors (contours in Figure 7 and A6) for a given acceptable rate of false *positive* errors.


Indicator 4
Maximum acceptable probability of a false positive error
0.01
0.03
0.05
0.07
0.09
0.11
0.13
0.15
0.05
0.1
0
0.05
0.1
0.3
0.5
Observation error variance


Indicator 2
Maximum acceptable probability of a false positive error
0.01
0.03
0.05
0.07
0.09
0.11
0.13
0.15
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0
0.05
0.1
0.3
0.5
Observation error variance


Indicator 4
Maximum acceptable probability of a false positive error
0.01
0.03
0.05
0.07
0.09
0.11
0.13
0.15
0.05
0.1
0.2
0.3
0.01
0.05
0.1
0.3
0.5
Process variation variance

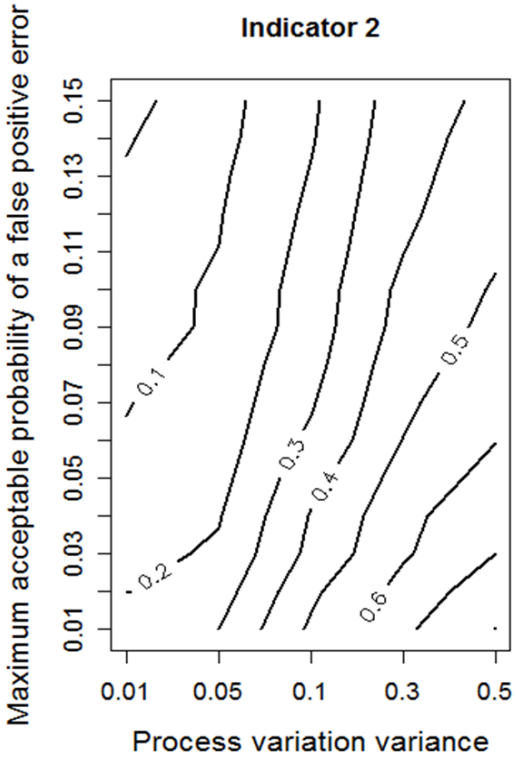

Indicator 2
Maximum acceptable probability of a false positive error
0.01
0.03
0.05
0.07
0.09
0.11
0.13
0.15
0.1
0.2
0.3
0.4
0.5
0.6
0.01
0.05
0.1
0.3
0.5
Process variation variance

Figure 7. ***The lowest rate of false negative (FN) errors (contour values) that can be obtained if the false positive (FP) rate is constrained to be below a certain value, as indicated on the y axis. False negative rates are shown for different levels of observation error (at σ_u^2=0.01) for the top-performing indicator in this study #4 (a) and commonly used IUCN decline indicator #2 (b). False negative rates are shown for different levels of process variation (at σ_v^2=0) for indicator #4 (c) and indicator #2 (d).***

4. Discussion

In order to make the best decisions regarding extinction risk classifications, it is important that the indicators being used to assess this risk be as reliable as possible and balance trade-offs between interest groups according to different levels of risk aversion. We found that the IUCN criterion A indicators of recent rate of decline were highly sensitive to process variation and observation error. With increased amounts of either process variation or observation error included in the analysis, the reliability of these indicators dropped substantially more than the reliability of the other indicators. This is because assessing status over a shorter time span (i.e., the most recent 12 years vs. as many as 50 years) would make it more difficult to identify longer-term trends in abundance, especially amid the noise of the process variation and observation error. Increasing the assessment window will increase the signal-to-noise (or decline-to-variance) ratio. This problem of the period of data used to assess status is a major concern for two compounding reasons: (1) indicators of recent rate of decline are the primary indicators currently being used to determine whether there has been a significant enough reduction in a population to consider it a conservation concern, and (2) there are often high levels of process variation and observation error inherent in data for wildlife and fisheries (Paulsen et al. 2007, Rand 2011, Wilson et al. 2011). The highest levels of error we examined (variance of the error term of 0.5) are representative of actual error rates in Fraser River sockeye salmon (Walters & Ludwig 1981, Dorner et al. 2009).

These results suggest that longer time-series of abundance, when they are available, should be used to evaluate whether a population is decreasing, rather than limiting assessments to the most recent 3 generations of data. When such lengthy series are not available, considerable effort should be made to secure and extend monitoring activities to minimize the chance of classification error and inadvertently incurring unexpected costs, such as the closure of a fishery due to an unpredicted

'extinction', e.g. collapse and non-recovery of Northern cod. This recommendation is consistent with results from other studies (Mace et al. 2002, Porszt et al. 2012).

Based on an empirical retrospective analysis, Porszt et al. (2012) found that indicators that were a measure of extent of decline from the maximum abundance in the data series were the least reliable indicators, and that the commonly used IUCN/COSEWIC criterion A indicators (decline over most recent 3 generations) ranked in the middle of the group of 20 indicators examined, most of which were similar to our indicators. Much like the empirical retrospective analysis of Porszt et al. (2012), we also found that indicators with a maximum abundance baseline tended to be less reliable than those based on decline in abundance from a historical baseline at the beginning of the time series. We also found that indicators based on recent rate of decline performed in the middle of the group, but only when there were low levels of process variation and observation error. When these sources of variance were larger, those indicators of recent rate of decline became much less reliable.

The area under the curve (AUC) values were often quite high (>0.8) across scenarios for the majority of the indicators. This was because our populations exhibited a high level of variability in order to reflect a range of productivities representative of the natural variability in sockeye salmon populations. Populations exhibiting an extremely large or extremely small amount of decline would be "easier" for indicators to classify than those that are close to the boundary rate of decline. It is important to note that it is not the absolute AUC vales that are the important measure in this study, but the relative AUC values of different indicators. As well, AUC reflects how well an indicator identifies populations that are more difficult to classify, rather than populations that are easy to classify.

Along with other recent studies (Wilson et al. 2011, Porszt et al. 2012), our work demonstrates the need to test indicators for reliability before using them to inform management and conservation decisions. A Receiving Operating Characteristic analysis provides a useful integrated measure of reliability that can be used to evaluate indicators over a full range of thresholds. However, when using ROC, it is important to be mindful that managers might not value both types of errors equally (which ROC does implicitly). Our analysis illustrates how to explicitly take these weightings into account.

A manager's weightings of the two types of errors might depend on the short- and long-term social, economic, and ecological costs of making a certain type of error, as well as political and administrative realities. While risk assessment is a scientific process, risk management and setting of priorities for conservation action is a societal process in fisheries and wildlife management (Irvine et al. 2005, Miller et al. 2006, Powles 2011). Managers could evaluate the costs of these errors and then decide what an acceptable level of each type of error is – either in absolute or relative terms (Mapstone 1995, Peterman 1990). There are inherent trade-offs between the rates of false positive and false negative errors such that both types of errors cannot be simultaneously minimized (Rice & Legacé 2007, Mace et al. 2008, Wilson et al. 2011). This paper demonstrates methods in which these error weightings and error tolerances could be incorporated explicitly into management decisions, allowing us to examine the trade-offs between the acceptable error rates and select for the best indicator(s) of population decline as a function of a manager's preferences for those weightings and tolerances. We also presented a method for managers to visualize the trade-offs between error types across different levels of process variation and observation error for their chosen indicator. In this study, we found that the best indicators to use (those with comparisons to some historical baseline) were surprisingly insensitive to different error weightings, but this might not always be the case under different conditions than examined here. Some indicators might be more predisposed to making certain types of errors (e.g., Wilson et al. 2011), and so might give more favourable results if a given error type is considered less important and weighted less heavily.

Our model was designed to simulate highly variable, semelparous fish populations with non-overlapping generations, such as sockeye salmon. Further studies would be required to examine if our general results about preferred indicators hold true for populations with different life-history traits or population dynamics (Dulvy et al. 2004). The methods used in this study can be applied to other taxa and to other environmental contexts to determine the reliability of indicators used to inform conservation decisions. Here, we assessed the reliability of indicators of decline in abundance, which are but one of the sets of metrics used to assign species to a threat category (IUCN 2001, COSEWIC 2011) – a measure of only one of the "symptoms" of extinction (Mace et al. 2008). Other indicators, such as those related to geographic area, absolute numbers of

individuals, or spatial distribution (Holt et al. 2009), should also be evaluated to ensure that the best possible indicators of conservation concern are being used.

A limitation of our model is that it did not include management responses to population declines. In reality, for a severely decreasing population, actions such as fisheries restrictions or habitat restoration would likely be taken (DFO 2005). In addition, although weighting false positive vs. false negative errors is mathematically feasible, it may be difficult in practice. Evaluating the costs of errors is complex, because they can be measured in different currencies (e.g. revenue vs. ecosystem health) and over various periods, and different groups may each incur the costs of different types of error (Peterman 1990).

Wherever possible, indicators should be evaluated both empirically and with simulation modelling (Punt 2000, DeMaster et al. 2004, Holt et al. 2009, Regan et al. 2009, Wilson et al. 2011, Porszt et al. 2012). That method allowed us to evaluate the relative reliability of a variety of indicators of decline, as well as to examine how robust they were to a range of relevant scenarios, such as changes in levels of process variation and observation error. We found that both of the latter have a significant impact on the relative ranking of reliability of certain indicators. Furthermore, evaluating indicators with simulation modelling is a relatively easy and inexpensive way to improve management decisions.

References

COSEWIC (Committee on the Status of Endangered Wildlife in Canada). 2011. COSEWIC's assessment process and criteria. Available from http://www.cosewic.gc.ca/pdf/Assessment_process_and_criteria_e.pdf (accessed October 2012).

DeMaster, D. (chair), R. Angliss, J. Cochrane, P. Mace, R. Merrick, M. Miller, S. Rumsey, B. Taylor, G. Thompson, and R. Waples. 2004. Recommendations to NOAA Fisheries: ESA Listing Criteria by the Quantitative Working Group, 10 June 2004. U.S. Dep. Commerce, NOAA Tech. Memo. NMFSF/SPO-67, 85 p.

Dorner, B., R. M. Peterman, and Z. Su. 2009. Evaluation of performance of alternative management models of Pacific salmon (Oncorhynchus spp.) in the presence of climatic change and outcome uncertainty using Monte Carlo simulations. Canadian Journal of Fisheries and Aquatic Sciences 66:2199-2221.

Dulvy, N. K., Y. Sadovy, and J. D. Reynolds. 2003. Extinction vulnerability in marine populations. Fish and Fisheries 4:25-64.

Dulvy, N. K., J.R. Ellis, N.B Goodwin, A. Grant, J.D. Reynolds, and S. Jennings. 2004. Methods of assessing extinction risk in marine fishes. Fish and Fisheries 5:255-276.

Dulvy, N. K., S. Jennings, N. B. Goodwin, A. Grant, and J. D. Reynolds. 2005. Comparison of threat and exploitation status in north-east Atlantic marine populations. Journal of Applied Ecology 42:883-891.

Dulvy, N. K., S. Jennings, S. I. Rogers, and D. L. Maxwell. 2006. Threat and decline in fishes: an indicator of marine biodiversity. Canadian Journal of Fisheries and Aquatic Sciences 63:1267-1275.

Field, S. A., A. J. Tyre, N. Jonzen, J. R. Rhodes, and H. P. Possingham. 2004. Minimizing the cost of environmental management decisions by optimizing statistical thresholds. Ecology Letters 7:669–675.

Fisheries and Oceans Canada (DFO). 2005. Canada's policy for conservation of wild Pacific salmon. DFO, Vancouver, B.C., Canada. Available from http://www.pac.dfo-mpo.gc.ca/publications/pdfs/wsp-eng.pdf (accessed January 2012).

Hibberd, P. L. and A. B. Cooper. 2008. Methodology: statistical analysis, test interpretation, basic principles of screening with application for clinical study. Pages 1517-1520 in Kleinman, R. E., O. Goulet, G. Mieli-Vergani, I. R. Sanderson, P. M. Sherman, and B. L. Shneider, editors. Walker's pediatric gastrointestinal disease: pathophysiology, diagnosis, management. 5th edition. B. C. Decker, Hamilton, Ontario.

Hilborn, R. and C. J. Walters. 1992. Chapter 7: Stock and Recruitment. In Quantitative Fisheries Stock Assessment: Choice, Dynamics, and Uncertainty. Chapman and Hall, New York.

Holt, C. A., A. Cass, B. Holtby, and B. Riddell. 2009. Indicators of status and benchmarks for conservation units in Canada's Wild Salmon Policy. DFO Canadian Science Advisory Secretariat Research Document 2009/058. viii + 74 p. Ottawa, Ontario, Canada. Available from http://www.dfo-mpo.gc.ca/CSAS/Csas/Publications/ResDocs-DocRech/2009/2009_058_e.pdf (accessed August 2012).

Hutchings, J. A. 2000. Collapse and recovery of marine fishes. Nature 406:882-885.

Hutchings, J. A. 2001. Conservation biology of marine fishes: perceptions and caveats regarding assignment of extinction risk. Canadian Journal of Fisheries and Aquatic Sciences 58:108-121.

Hutchings, J. A. and J. D. Reynolds. 2004. Marine fish population collapses: Consequences for recovery and extinction risk. BioScience 54:297-309.

Irvine, J. R., M. R. Gross, C. C. Wood, L. B. Holtby, N. D. Schubert, and P. G. Amiro. 2005. Canada's species at risk act: An opportunity to protect "endangered" salmon. Fisheries 30:11-19.

IUCN (International Union for Conservation of Nature). 2001. IUCN Red List categories and criteria. Version 3.1. Species Survival Commission, IUCN, Gland, Switzerland. Available from http://intranet.iucn.org/webfiles/doc/SSC/RedList/redlistcatsenglish.pdf (accessed October 2012).

IUCN (International Union for Conservation of Nature) Standards and Petitions Subcommittee. 2010. Guidelines for using the IUCN Red List categories and criteria. IUCN, Gland, Switzerland. Available from: http://intranet.iucn.org/webfiles/doc/SSC/RedList/RedListGuidelines.pdf (accessed October 2012).

Keith, D. A., et al. 2004. Protocols for listing threatened species can forecast extinction. Ecology Letters 7:1101-1108.

Mace, P. M., et al. 2002. National Marine Fisheries Service (NMFS) / Interagency Working Group evaluation of CITES criteria and guidelines. Technical memorandum NMFS-F/SPO-58. National Oceanic and Atmospheric Administration, Rockville, Maryland.

Mace, G. M., N. J. Collar, K. J. Gaston, C. Hilton-Taylor, H. R. Akçakaya, N. Leader-Williams, E. J. Milner-Gulland, and S. N. Stuart. 2008. Quantification of extinction risk: International Union for the Conservation of Nature's (IUCN) system for classifying threatened species. Conservation Biology 22:1424-1442.

Mapstone, B. D. 1995. Scalable decision rules for environmental impact studies: effect size, type I, and type II errors. Ecological Applications 5:401-410.

Matsuda, H., Y. Takenaka, T. Yaharat, and Y. Uozumi. 1998. Extinction risk assessment of declining wild populations: The case of the southern bluefin tuna. Researches on Population Ecology 40:271-278.

Miller, R. M., et al. 2006. Extinction risk and conservation priorities. Science 313:441.

Paulsen, C. M., R. A. Hinrichsen, and T. R. Fisher. 2007. Measure twice, estimate once: Pacific salmon population viability analysis for highly variable populations. Transactions of the American Fisheries Society 136:346-364.

Peterman, R. M. 1990. Statistical power analysis can improve fisheries research and management. Canadian Journal of Fisheries and Aquatic Sciences 47:2-15.

Porszt, E. J., R. M. Peterman, N. K. Dulvy, A. B. Cooper, and J. R. Irvine. 2012. Reliability of indicators of decline in abundance. Conservation Biology 26:894-904.

Powles, H., M. J. Bradford, R.G. Bradford, W.G. Doubleday, S. Innes, and C. D. Levings. 2000. Assessing and protecting endangered marine species. ICES Journal of Marine Science 57:669-676.

Powles, H. 2011. Assessing risk of extinction of marine fishes in Canada - the COSEWIC experience. Fisheries 36:231-246.

Punt, A. E. 2000. Extinction of marine renewable resources: a demographic analysis. Population Ecology 42:19-27.

Rand, P. S. 2011. Oncorhynchus nerka. IUCN Red List of Threatened Species. Version 2011.2. International Union for Conservation of Nature, Gland, Switzerland. Available from http://www.iucnredlist.org/apps/redlist/details/135301/0 (accessed January 2012).

Regan, T. J., M. A. Burgman, M. A. McCarthy, L. L. Master, D. A. Keith, G. M. Mace, and S. J. Andelman. 2005. The consistency of extinction risk classification protocols. Conservation Biology 19:1969–1977.

Regan, T., B. Taylor, G. Thompson, J. Cochrane, R. Merrick, M. Nammack, S. Rumsey, K. Ralls, and M. Runge. 2009. Developing a structure for quantitative listing criteria for the U.S. Endangered Species Act using performance testing, phase 1 report. Technical memorandum NOAA-TM-NMFS-SWFSC-437. National Oceanic and Atmospheric Administration, Rockville, Maryland.

Reynolds, J. D., N. K. Dulvy, N. B. Goodwin, and J. A. Hutchings. 2005. Biology of extinction risk in marine fishes. Proceedings of the Royal Society of London B 272:2337–2344.

Rice, J. C., and E. Legacé. 2007. When control rules collide: a comparison of fisheries management reference points and International Union for Conservation of Nature (IUCN) criteria for assessing risk of extinction. ICES Journal of Marine Science 64:718-722.

Ricker, W. E. 1997. Cycles of abundance among Fraser River sockeye salmon (Oncorhynchus nerka). Canadian Journal of Fisheries and Aquatic Sciences 54:950-968.

Venables, W. N., and B. D. Ripley. 2002. Modern applied statistics with S. 4th edition. Springer Science, New York.

Walters, C. J., and D. Ludwig. 1981. Effects of measurement errors on the assessment of stock-recruitment relationships. Canadian Journal of Fisheries and Aquatic Sciences 38:704-710.

Wilson, H. B., B. E. Kendall, and H. P. Possingham. 2011. Variability in population abundance and the classification of extinction risk. Conservation Biology 25:747-757.

Appendices

Appendix A.

Data on the *a* parameter

To estimate values of the Ricker *a* parameter in nature, we used data from Dorner et al. (2009), encompassing 37 stocks of wild North American sockeye salmon with spawning areas ranging from northern Washington to western Alaska and covering brood years (spawning years) from the 1950s to the late 1990s. The estimated parameter values from Dorner et al. (2009) were in the form $R = Se^{\alpha - bS}$ (for a spawner-to-recruit model). When substituting $a=e^{\alpha}$, we get the $R = aSe^{-bS}$ form of the equation used here, and the spread of the *a* values ranged from 0.5 to 14.6, with a mean value of 6.2 (Figure A1).

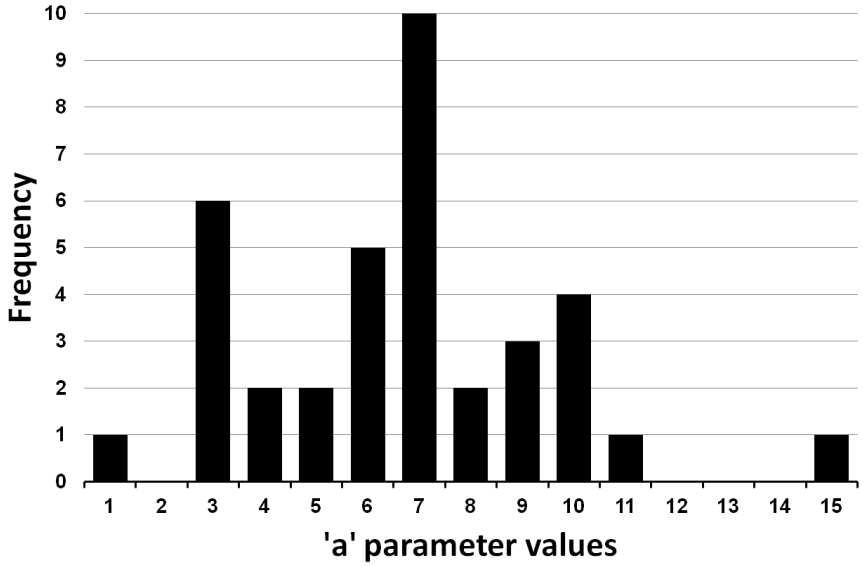


Figure A1. Frequency of 'a' parameter values derived from the α parameter values from Dorner et al. (2009) data.

Because our model was a spawner-to-spawner model rather than a spawner-to-recruit model, we also took into account annual harvest rates. The average annual harvest rate of Fraser River sockeye salmon from 1952-2006 was 69% (Porszt et al. 2012). When this value was taken into account and incorporated into the *a* parameter, the resulting spread of the distribution shifted downward and ranged from 0.16 to 4.5 with a mean *a* value of 1.9 and a standard deviation of approximately 0.9.

In addition to fishing mortality, pre-spawner in-river mortality rates can be substantial (e.g., Porszt et al. 2012), which would further decrease the *a* parameter value in a spawner-to-spawner model. Because we are concerned with how reliable decline indicators are at assessing populations that are truly declining, we further reduced the mean *a* value to 1.0 for our model, and reduced the standard deviation to 0.3. Our *a* parameter values thus capture realistic *a* values for Fraser River sockeye salmon, albeit with more emphasis on the lower end of the productivity values that have been observed, which is the set of situations in which we are interested (i.e., when populations are declining).

Appendix B.

Decline indicators

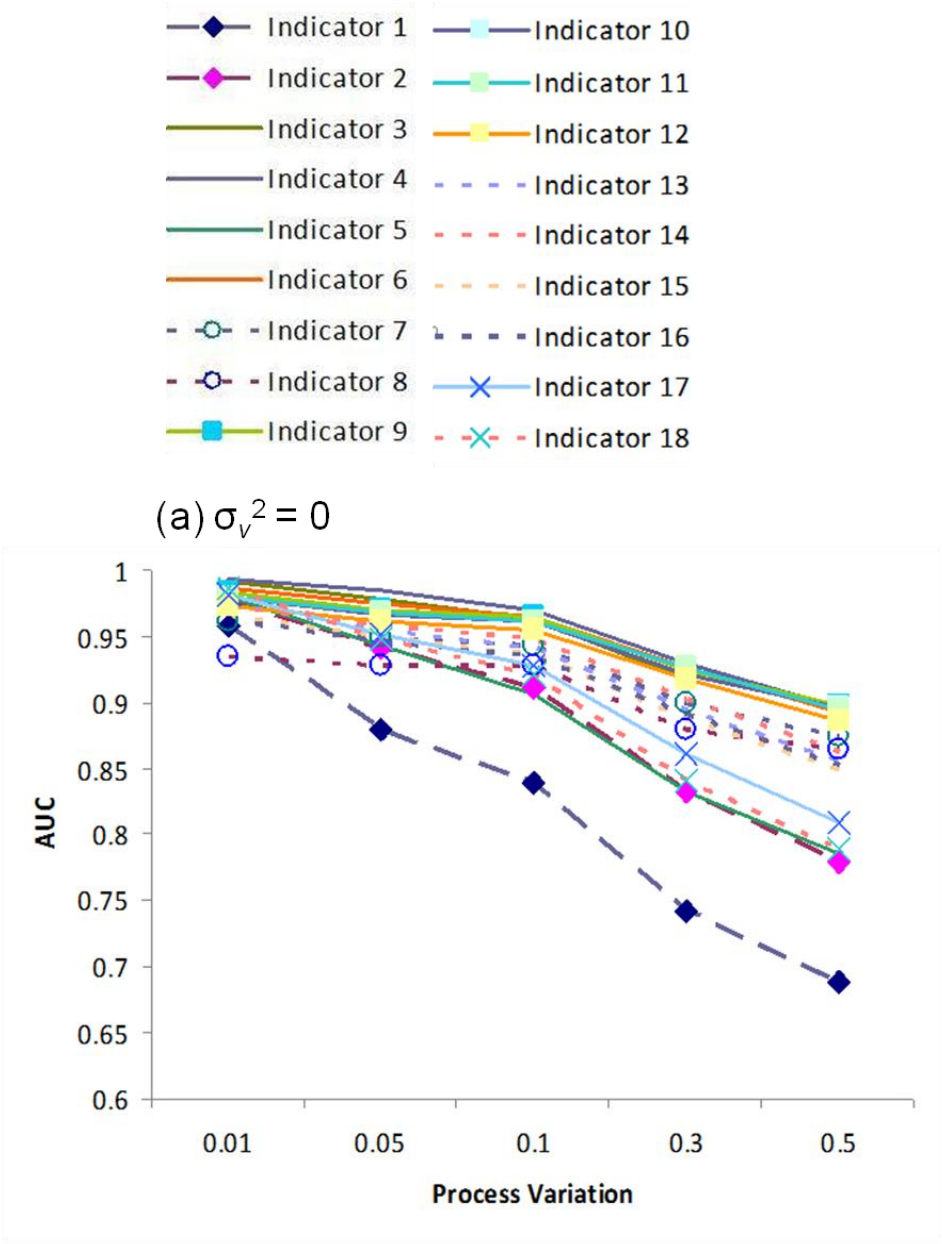
1. Percent decline in spawner abundance over the most recent three generations (i.e., rate of decline over 12 years in the case of Fraser River Sockeye salmon), estimated by exponentiation of best-fit values from the robust regression of loge (unsmoothed abundance) on years.
2. Same as indicator #1 except we used smoothed (4-year running mean) abundances rather than unsmoothed abundances.
3. Percent decline between abundance in first year of the data series and abundance in a subsequent assessment year (at least 12 years later), using values estimated by exponentiation of best-fit values from the robust regression of $\log_e$(unsmoothed abundance) on years.
4. Same as indicator #3 except we used smoothed (4-year running mean) abundances rather than unsmoothed abundances.
5. Same as indicator #3 except we used data from the first corresponding cycle year up to the year of analysis (e.g., dominant compared with another dominant cycle year).
6. Same as indicator #4 except we used data from the first corresponding cycle year up to the year of analysis (e.g., dominant compared with another dominant cycle year).
7. Same as indicator #3 except we used the maximum unsmoothed annual abundance anywhere in the time series as the historical baseline, instead of abundance in the first year.
8. Same as indicator #7 except we used smoothed (4-year running mean) abundances rather than unsmoothed abundances.
9. Percent decline between geometric mean abundance (equivalent to the exponentiated arithmetic mean of $\log_e$(unsmoothed abundance)) of the first 4-year generation and the mean of a subsequent generation being assessed, where generations move one year at a time in sliding windows.
10. Same as indicator #9 except we used smoothed (4-year running mean) rather than unsmoothed abundances.
11. Same as indicator #9 except generations moved in 4-year blocks with no overlap of years (i.e., status only assessed every four years).
12. Same as indicator #10 except generations moved in 4-year blocks with no overlap of years.
13. Same as indicator #9 except for the historical baseline, we used the maximum geometric mean abundance of any three-generation (12-year) period in the time series.
14. Same as indicator #13 except we used smoothed (4-year running mean) rather than unsmoothed abundances.
15. Same as indicator #13 except generations moved in 4-year blocks with no overlap of years.
16. Same as indicator #15 except we used smoothed (4-year running mean) rather than unsmoothed abundances.
17. Percent decline between the geometric mean abundance (which is by definition described as raw number of fish) of the first 4-year generation and raw abundance in a subsequent assessment year (starting at least 12 years later).
18. Same as indicator #17 except for the historical baseline, we used the maximum geometric mean abundance of a three-generation (12-year) period that occurred anywhere in the time series.

Appendix C.

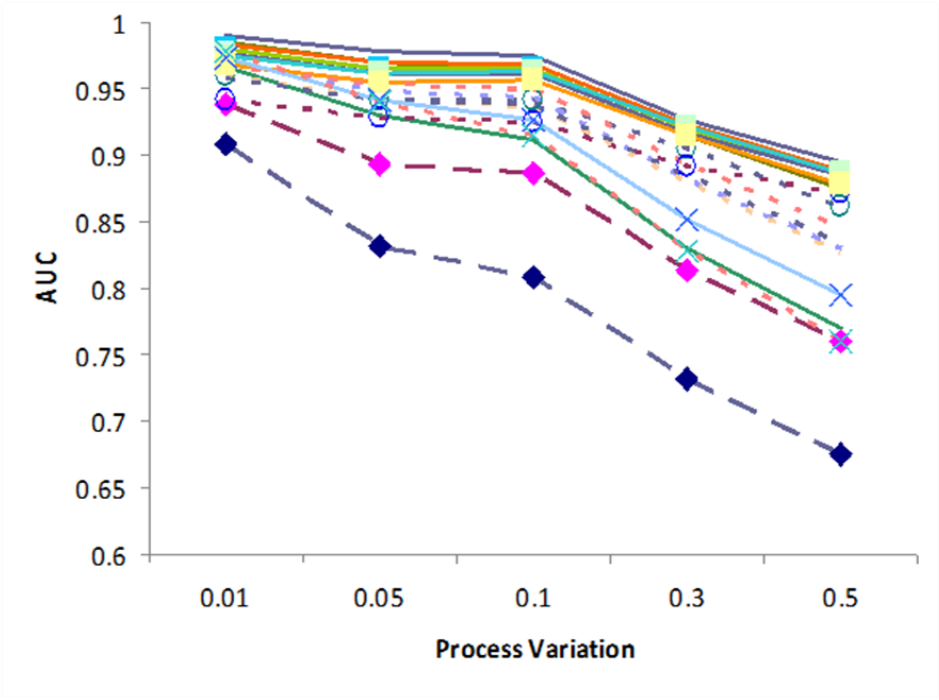
Additional results

Here we present the full results for the 18 indicators in our study, as well as the results of our sensitivity analyses.

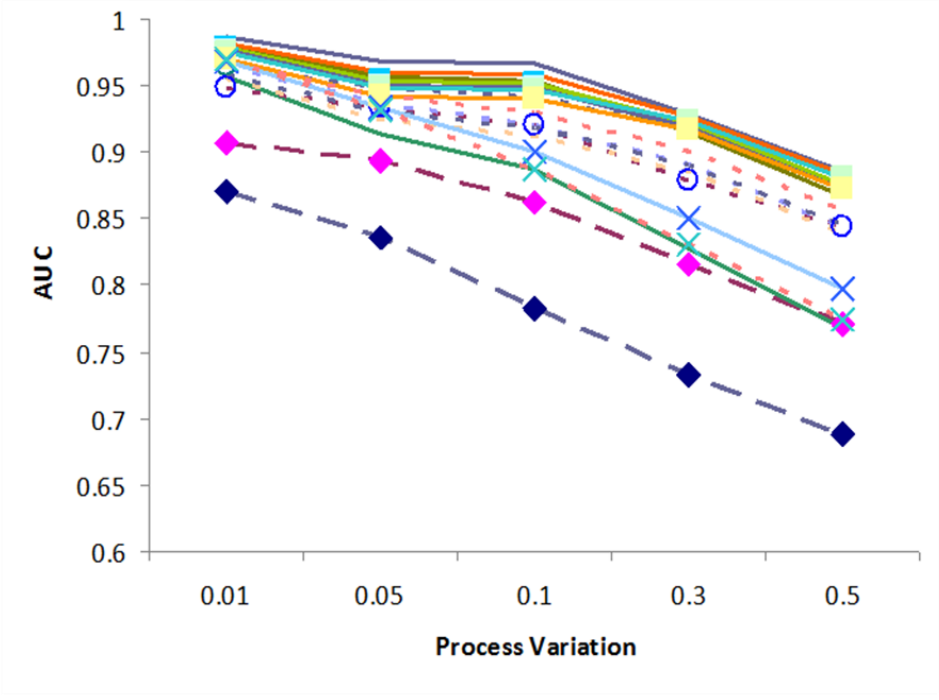
Figure A2 shows the effects on AUC (probability that an indicator is able to differentiate between a declining and non-declining population) of increasing process variation at different levels of observation error (σ_v^2=0, 0.05, 0.1, 0.3, 0.5) for all indicators.



(b) $\sigma_v^2 = 0.05$



(c) $\sigma_v^2 = 0.1$



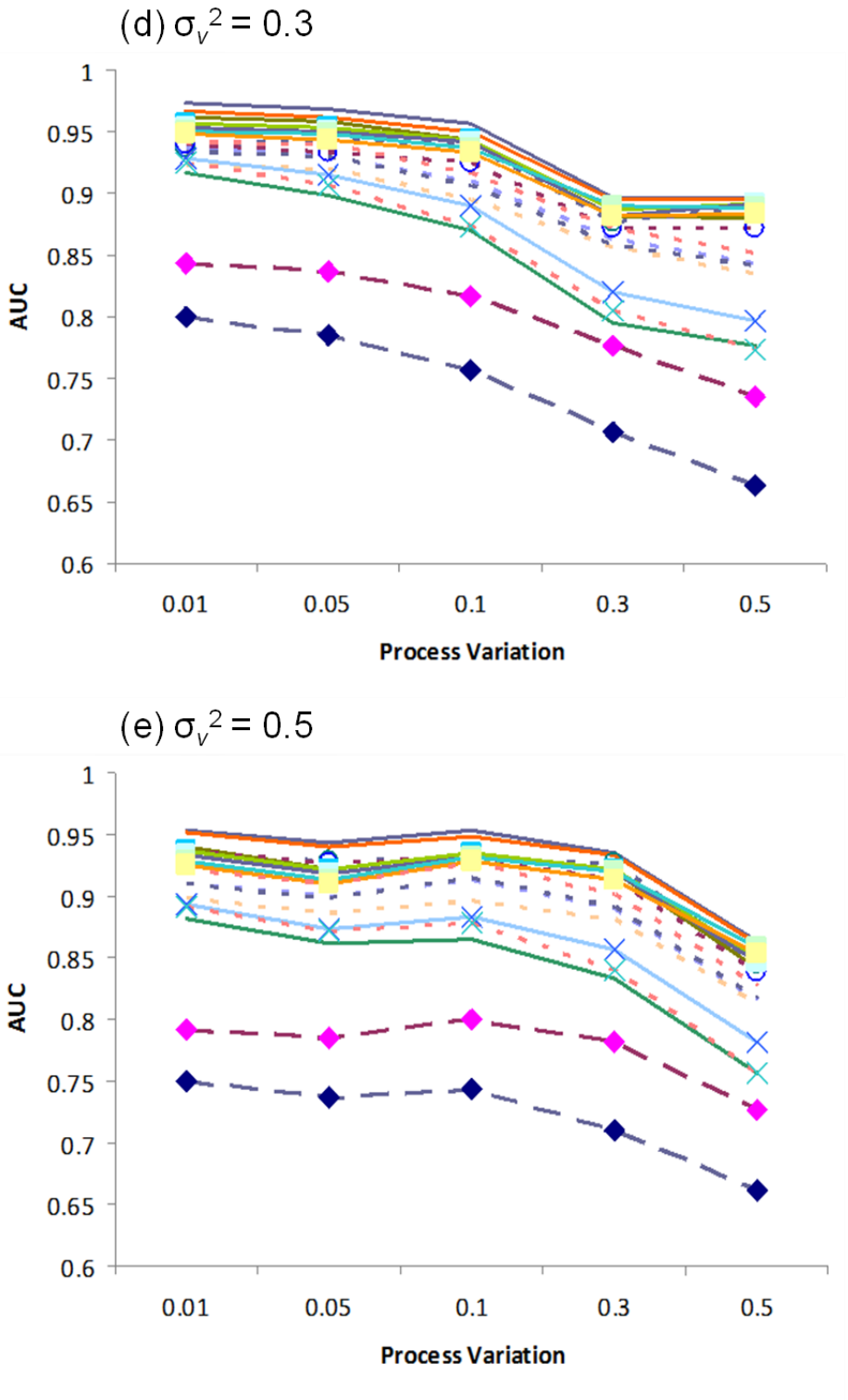
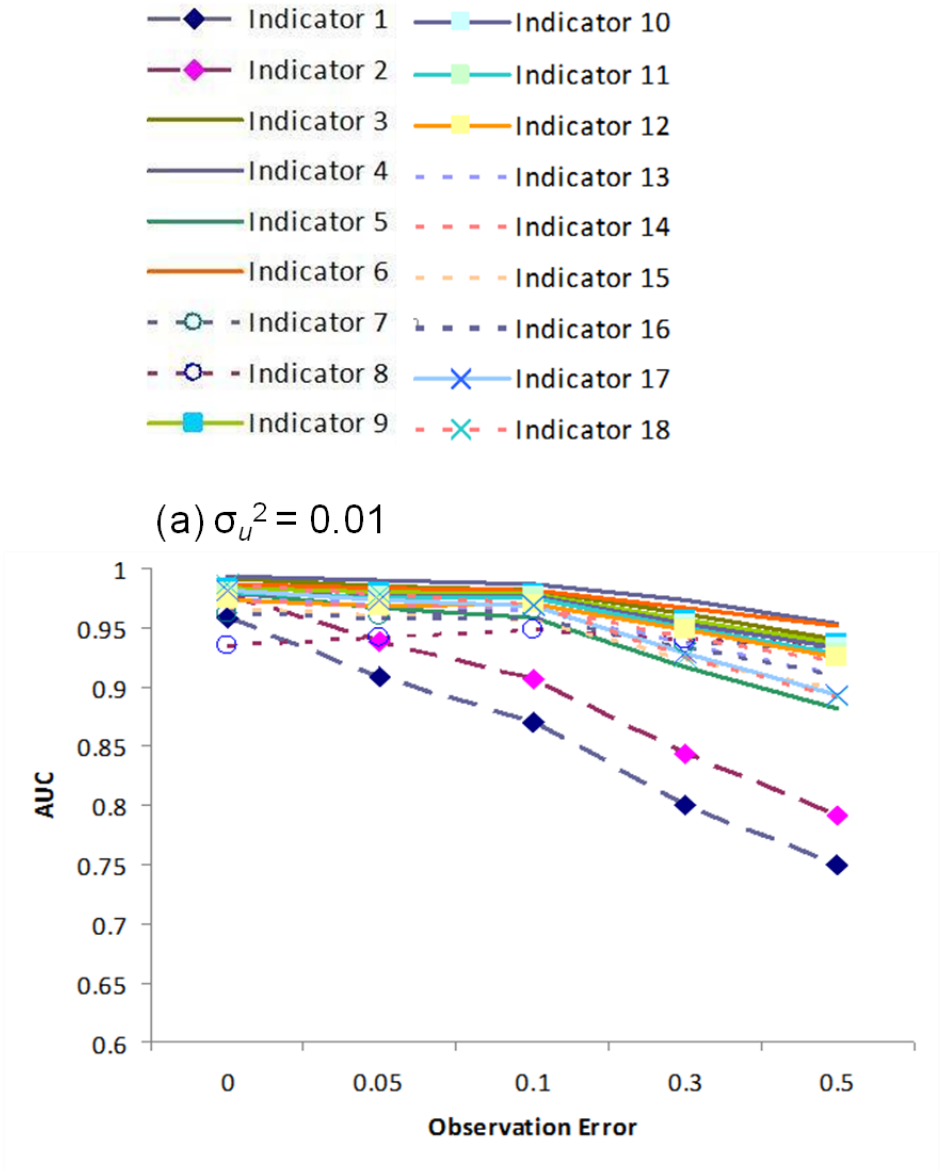
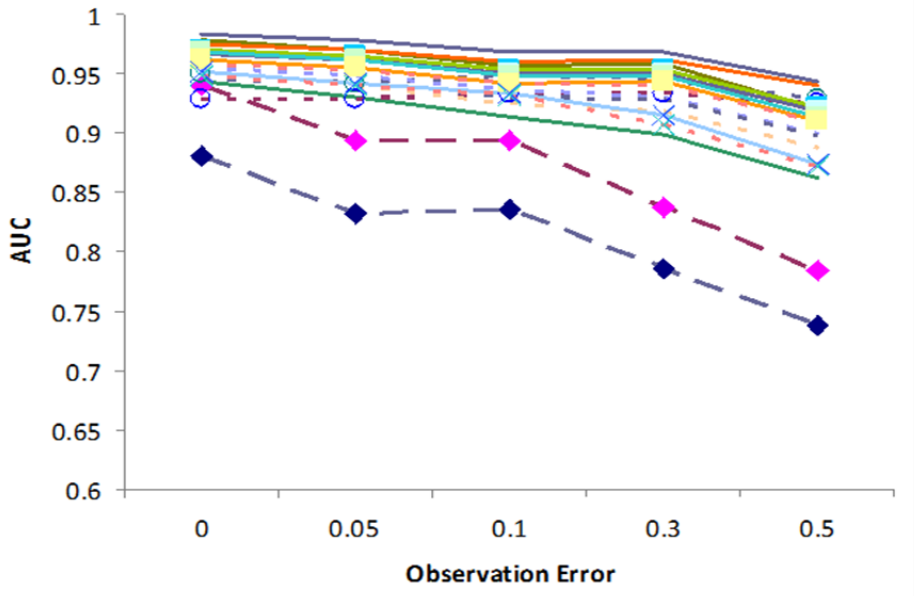

(d) $\sigma_v^2 = 0.3$
AUC
1
0.95
0.9
0.85
0.8
0.75
0.7
0.65
0.6
0.01
0.05
0.1
0.3
0.5
Process Variation
(e) $\sigma_v^2 = 0.5$
AUC
1
0.95
0.9
0.85
0.8
0.75
0.7
0.65
0.6
0.01
0.05
0.1
0.3
0.5
Process Variation

Figure A2.

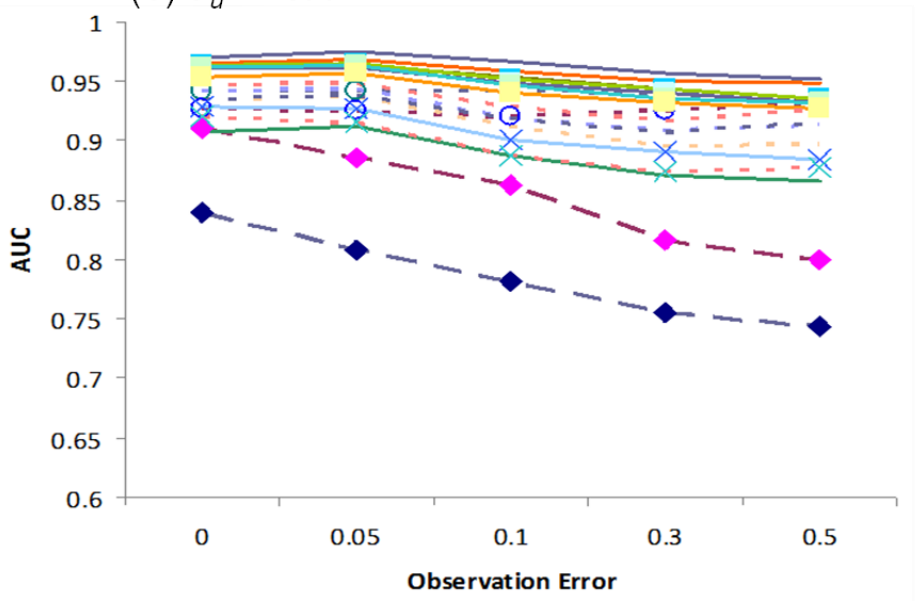
Figure A3 shows the effects on AUC of increasing observation error at different levels of process variation (σ_u^2=0.01, 0.05, 0.1, 0.3, 0.5) for all indicators.



(b) $\sigma_u^2 = 0.05$



(c) $\sigma_u^2 = 0.1$



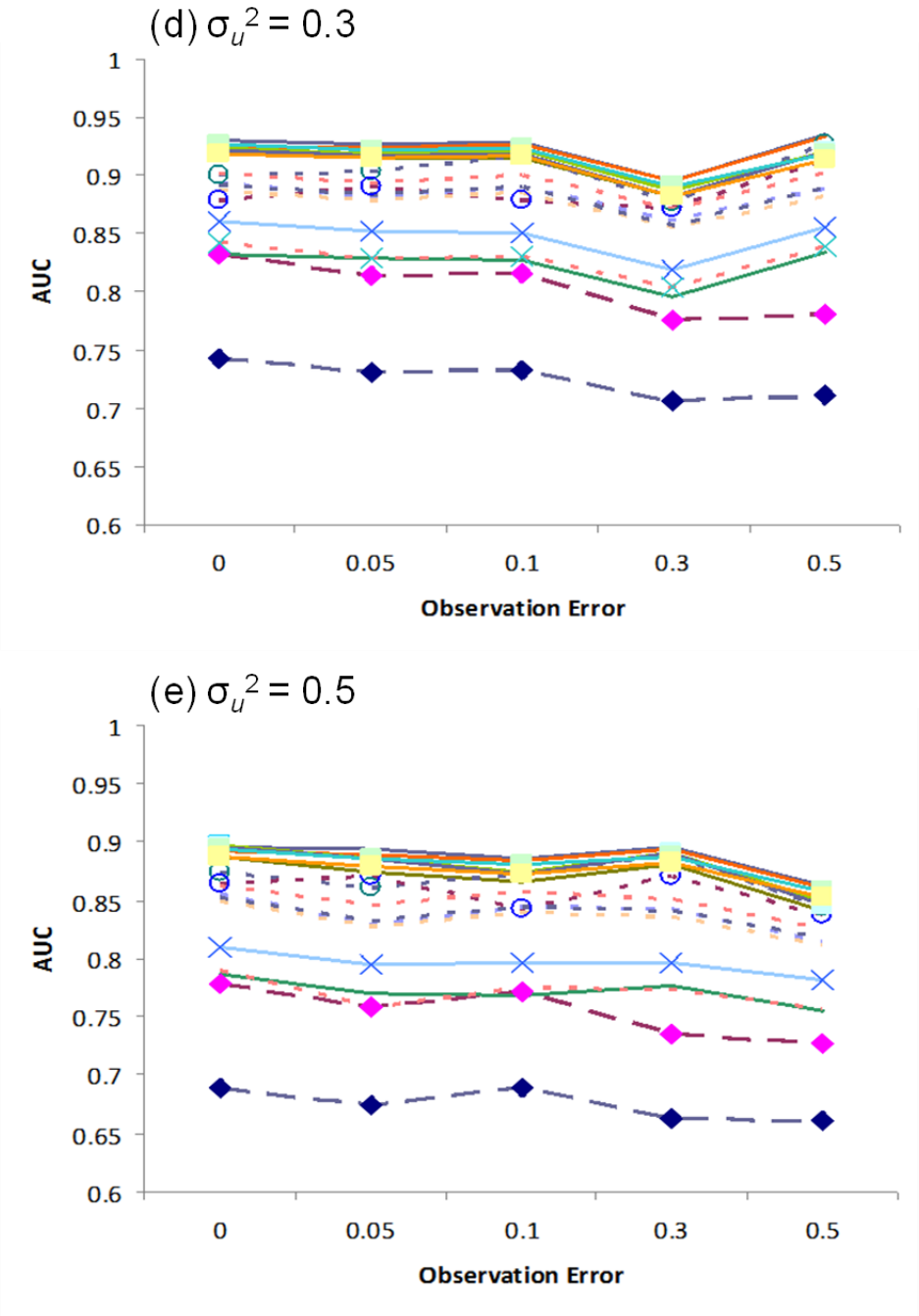

(d) $\sigma_u^2 = 0.3$
AUC
1
0.95
0.9
0.85
0.8
0.75
0.7
0.65
0.6
0
0.05
0.1
0.3
0.5
Observation Error
(e) $\sigma_u^2 = 0.5$
AUC
1
0.95
0.9
0.85
0.8
0.75
0.7
0.65
0.6
0
0.05
0.1
0.3
0.5
Observation Error

Figure A3.

Figure A4 shows the effect on AUC of the boundary used to define a subsequent status of a declining population for all indicators (σ_v^2=0, σ_u^2=0.01).

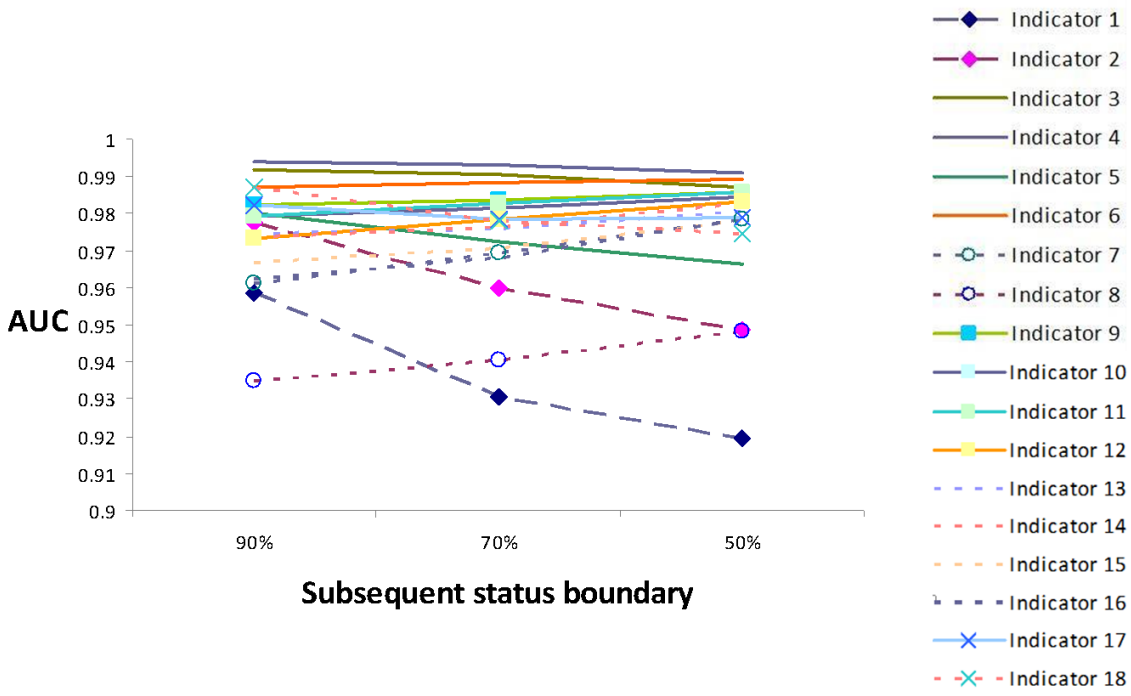


Figure A4.

Figure A5 shows the effects on AUC of the *a* (a) and *b* (b) parameters for all indicators. Increasing *a* or decreasing *b* results in more productive populations, resulting in fewer populations being evaluated as declining using our initial boundaries. Without both declining and non-declining populations, it is impossible to calculate an ROC curve. To get around this problem, we used the mean percentage decline generated by each set of populations as the boundary which defines a declining population.

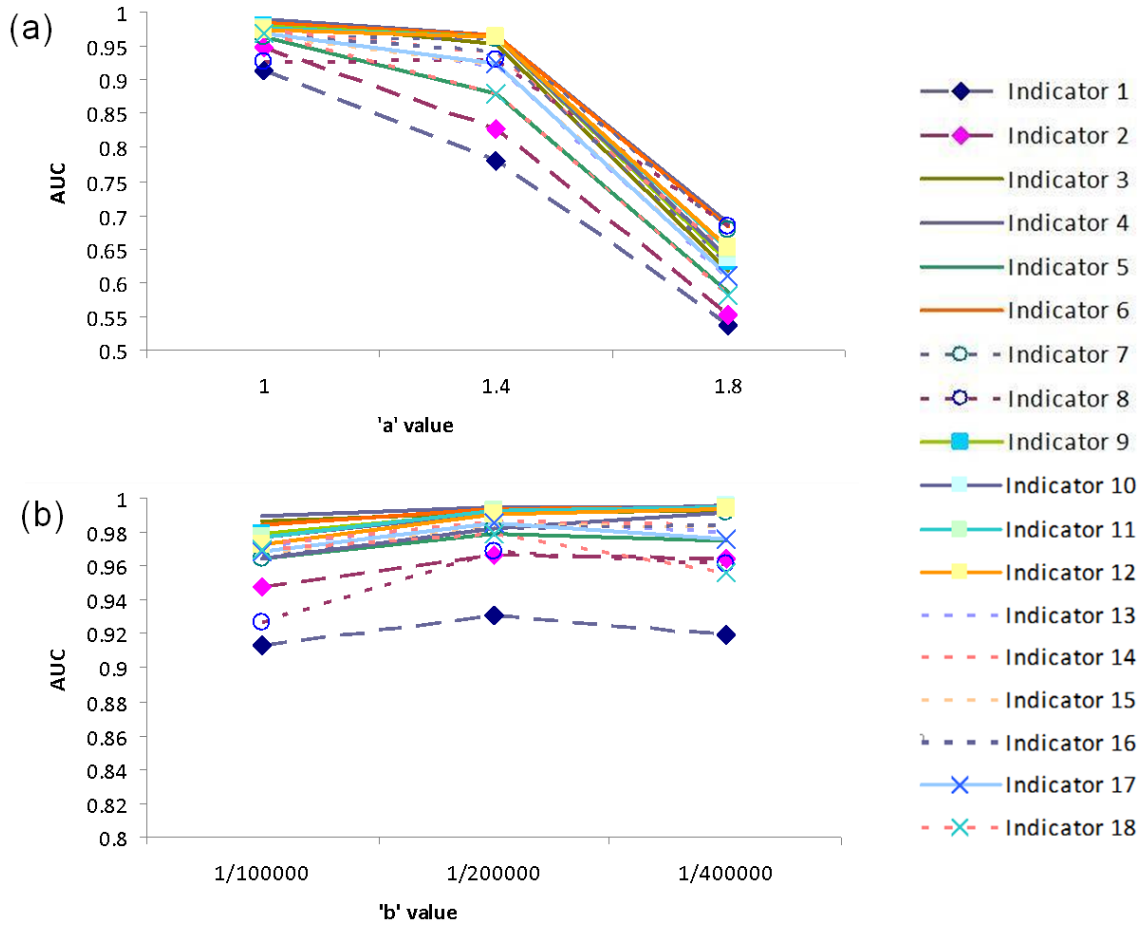
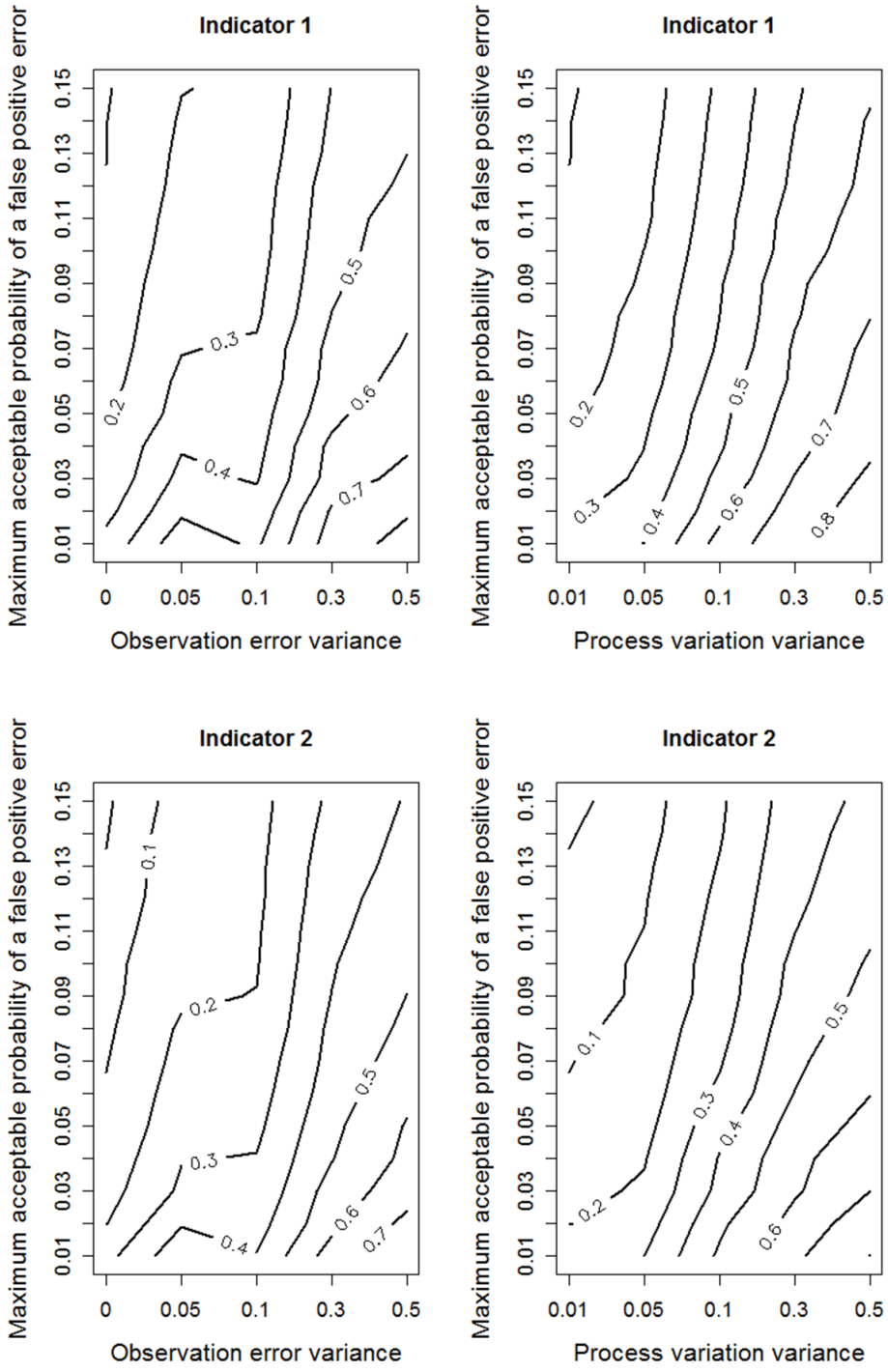
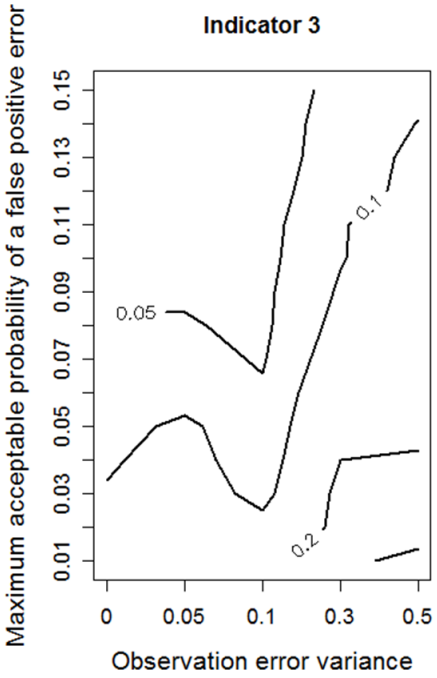
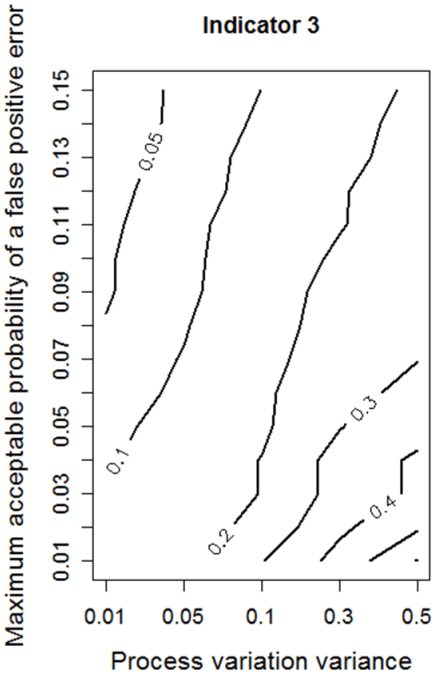


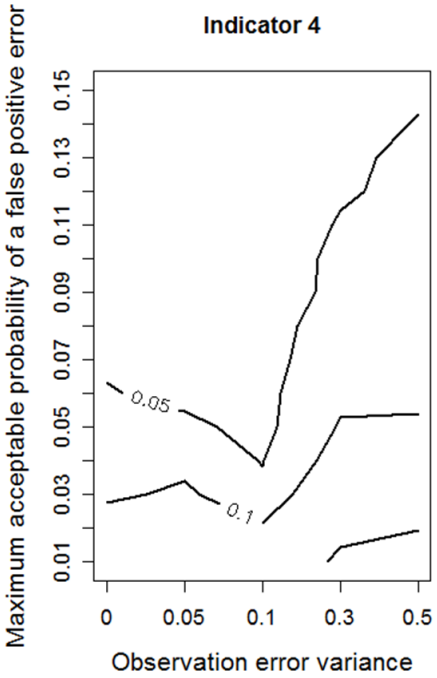
Figure A5.

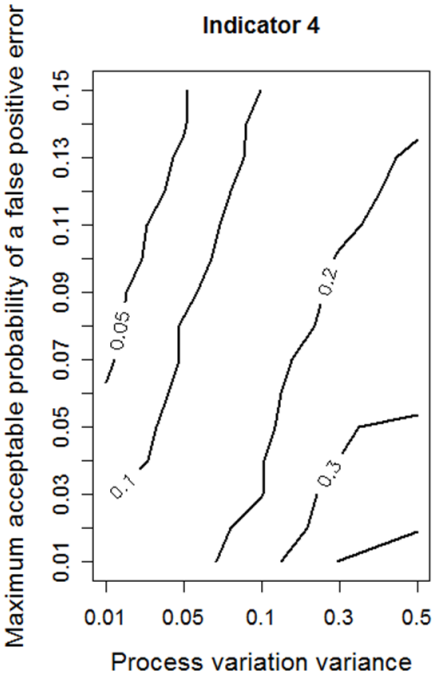
Figure A6 shows the contour plots of the lowest rate of false negative errors that can be obtained if the false positive rate is constrained to be below a certain value for all of the indicators (as in Figure 7).

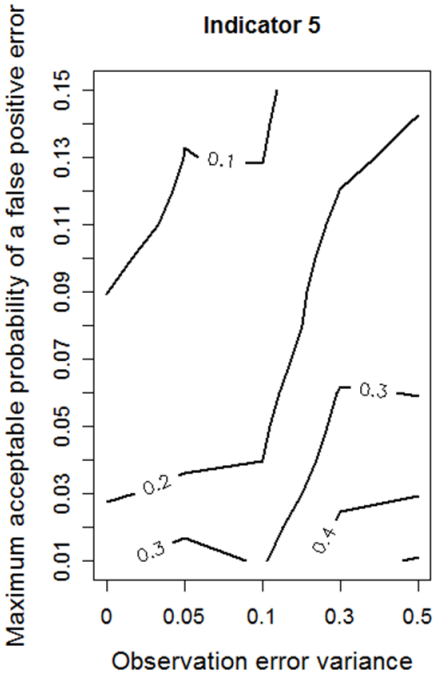


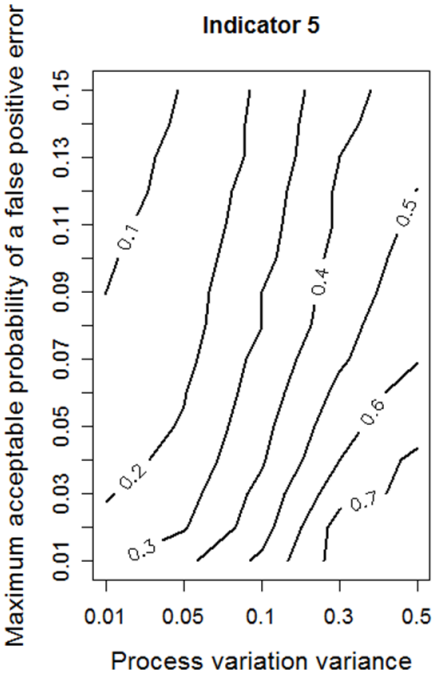

Indicator 3
Maximum acceptable probability of a false positive error
0.01
0.03
0.05
0.07
0.09
0.11
0.13
0.15
0.05
0.1
0.2
0
0.05
0.1
0.3
0.5
Observation error variance

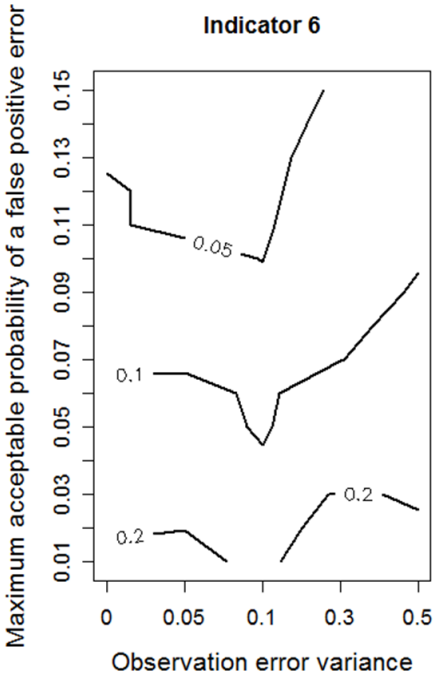

Indicator 3
Maximum acceptable probability of a false positive error
0.01
0.03
0.05
0.07
0.09
0.11
0.13
0.15
0.05
0.1
0.2
0.3
0.4
0.01
0.05
0.1
0.3
0.5
Process variation variance

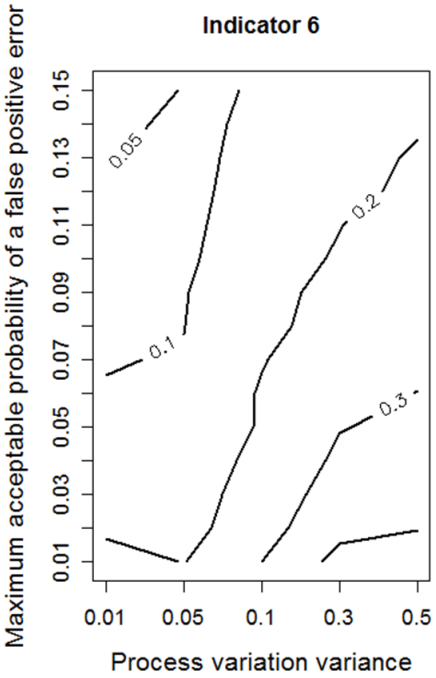

Indicator 4
Maximum acceptable probability of a false positive error
0.01
0.03
0.05
0.07
0.09
0.11
0.13
0.15
0.05
0.1
0
0.05
0.1
0.3
0.5
Observation error variance

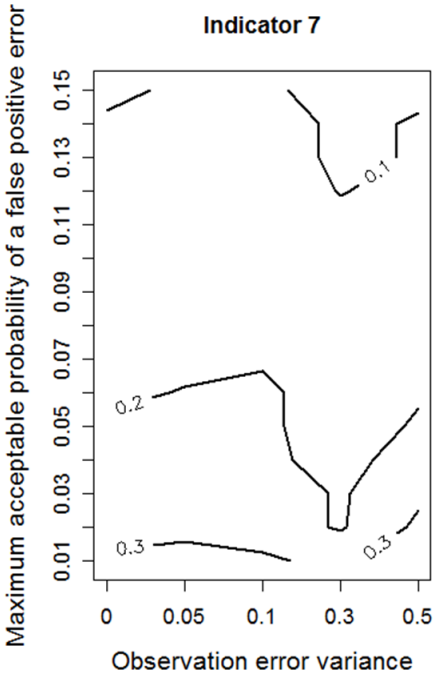

Indicator 4
Maximum acceptable probability of a false positive error
0.01
0.03
0.05
0.07
0.09
0.11
0.13
0.15
0.05
0.1
0.2
0.3
0.01
0.05
0.1
0.3
0.5
Process variation variance

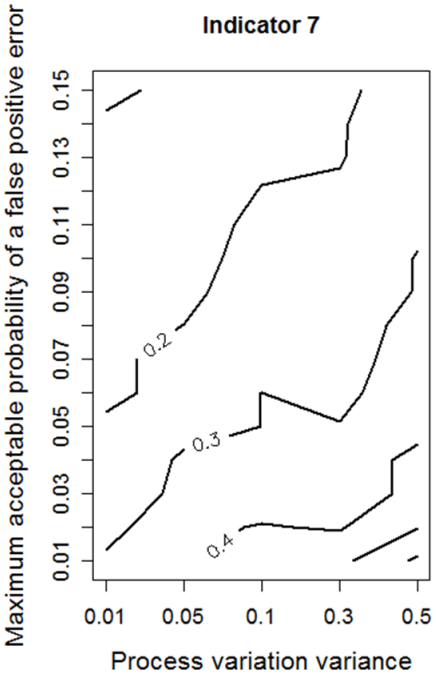

Indicator 5
Maximum acceptable probability of a false positive error
0.01
0.03
0.05
0.07
0.09
0.11
0.13
0.15
0.1
0.2
0.3
0.4
0
0.05
0.1
0.3
0.5
Observation error variance

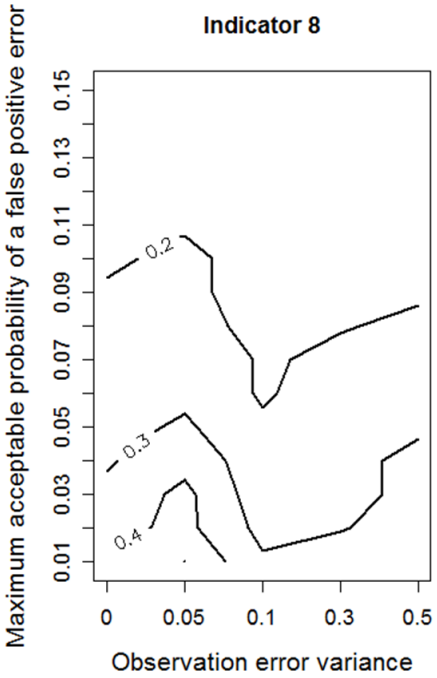

Indicator 5
Maximum acceptable probability of a false positive error
0.01
0.03
0.05
0.07
0.09
0.11
0.13
0.15
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.01
0.05
0.1
0.3
0.5
Process variation variance

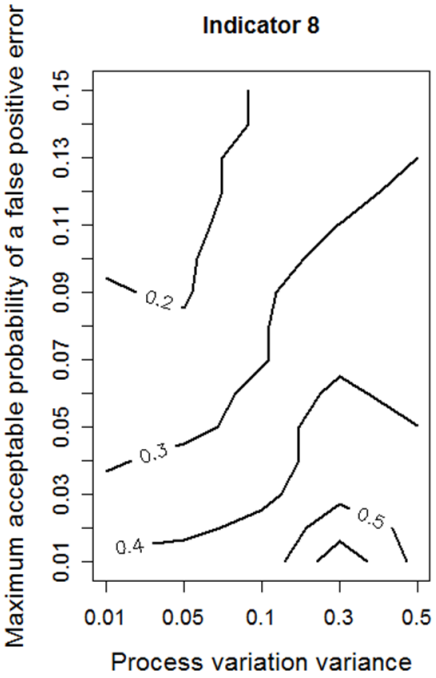

Indicator 6
Maximum acceptable probability of a false positive error
0.01
0.03
0.05
0.07
0.09
0.11
0.13
0.15
0.05
0.1
0.2
0
0.05
0.1
0.3
0.5
Observation error variance

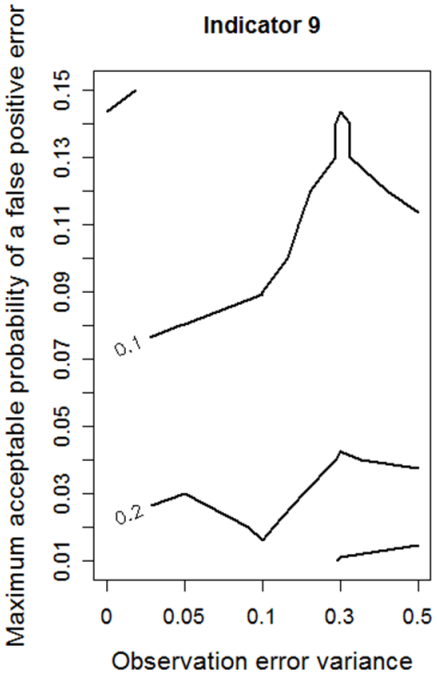

Indicator 6
Maximum acceptable probability of a false positive error
0.01
0.03
0.05
0.07
0.09
0.11
0.13
0.15
0.05
0.1
0.2
0.3
0.01
0.05
0.1
0.3
0.5
Process variation variance

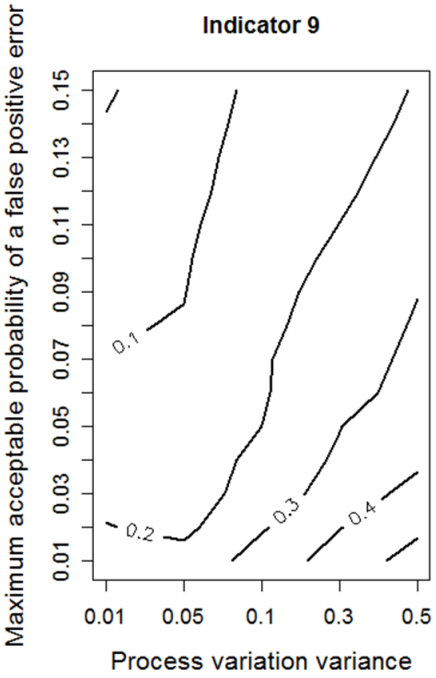

Indicator 7
Maximum acceptable probability of a false positive error
0.01
0.03
0.05
0.07
0.09
0.11
0.13
0.15
0.1
0.2
0.3
0.3
0
0.05
0.1
0.3
0.5
Observation error variance

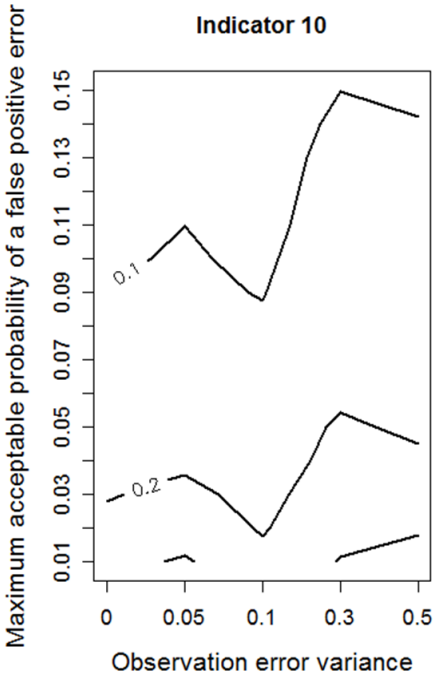

Indicator 7
Maximum acceptable probability of a false positive error
0.01
0.03
0.05
0.07
0.09
0.11
0.13
0.15
0.2
0.3
0.4
0.01
0.05
0.1
0.3
0.5
Process variation variance

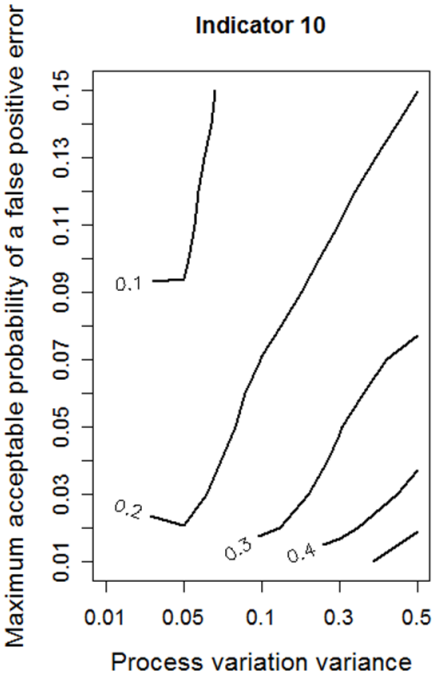

Indicator 8
Maximum acceptable probability of a false positive error
0.01
0.03
0.05
0.07
0.09
0.11
0.13
0.15
0.2
0.3
0.4
0
0.05
0.1
0.3
0.5
Observation error variance

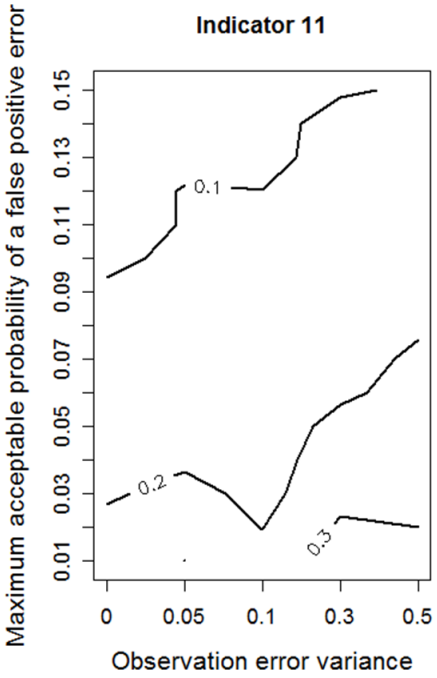

Indicator 8
Maximum acceptable probability of a false positive error
0.01
0.03
0.05
0.07
0.09
0.11
0.13
0.15
0.2
0.3
0.4
0.5
0.01
0.05
0.1
0.3
0.5
Process variation variance

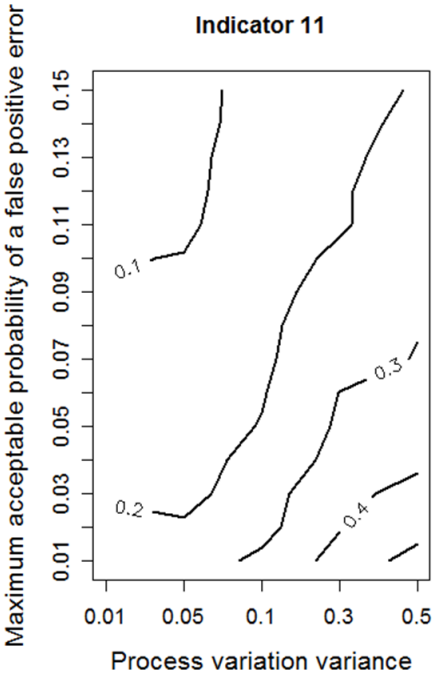

Indicator 9
Maximum acceptable probability of a false positive error
0.01
0.03
0.05
0.07
0.09
0.11
0.13
0.15
0.1
0.2
0
0.05
0.1
0.3
0.5
Observation error variance

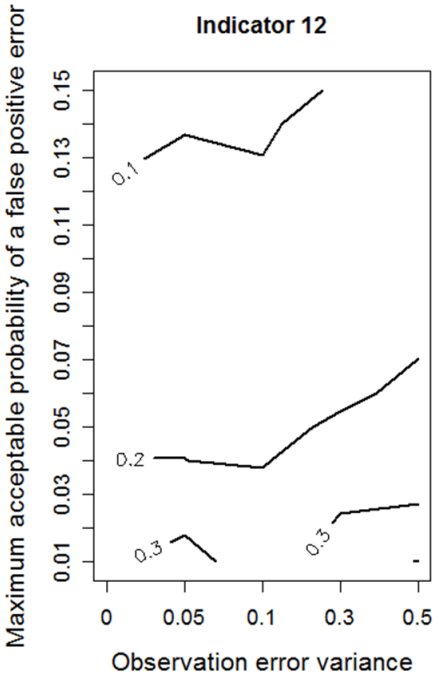

Indicator 9
Maximum acceptable probability of a false positive error
0.01
0.03
0.05
0.07
0.09
0.11
0.13
0.15
0.1
0.2
0.3
0.4
0.01
0.05
0.1
0.3
0.5
Process variation variance

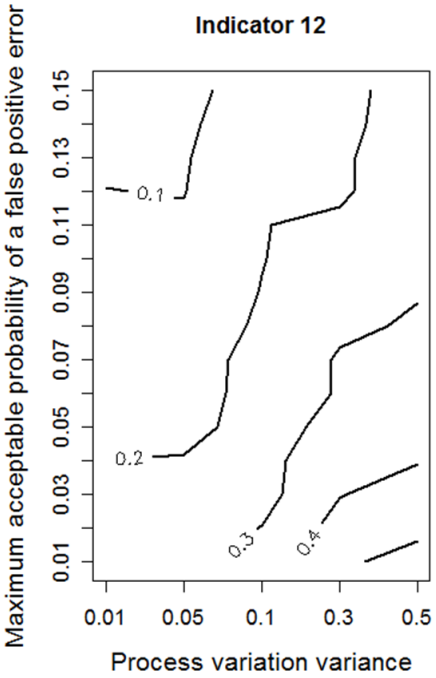

Indicator 10
Maximum acceptable probability of a false positive error
0.01
0.03
0.05
0.07
0.09
0.11
0.13
0.15
0.1
0.2
0
0.05
0.1
0.3
0.5
Observation error variance

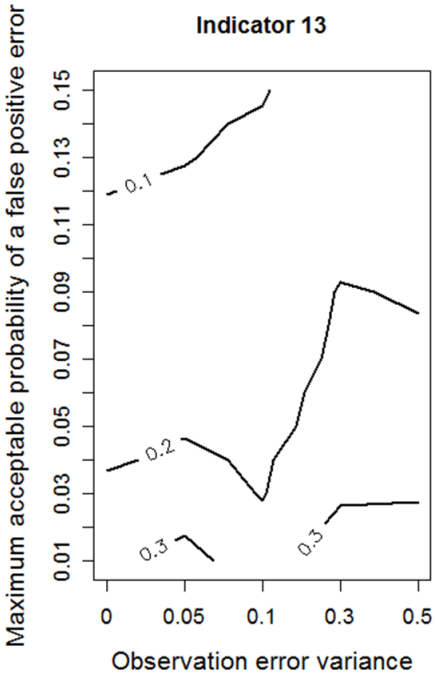

Indicator 10
Maximum acceptable probability of a false positive error
0.01
0.03
0.05
0.07
0.09
0.11
0.13
0.15
0.1
0.2
0.3
0.4
0.01
0.05
0.1
0.3
0.5
Process variation variance

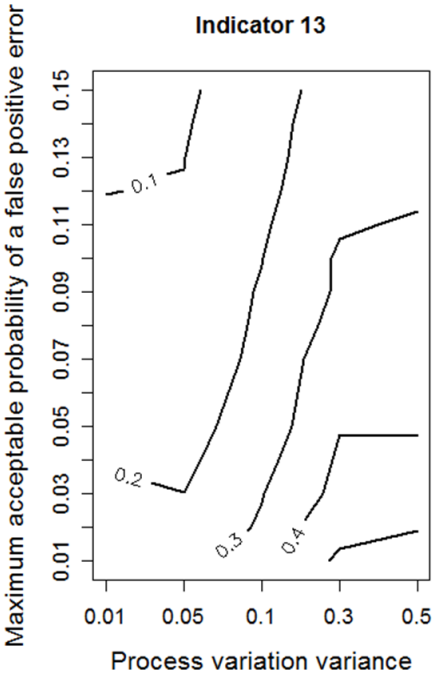

Indicator 11
Maximum acceptable probability of a false positive error
0.01
0.03
0.05
0.07
0.09
0.11
0.13
0.15
0.1
0.2
0.3
0
0.05
0.1
0.3
0.5
Observation error variance

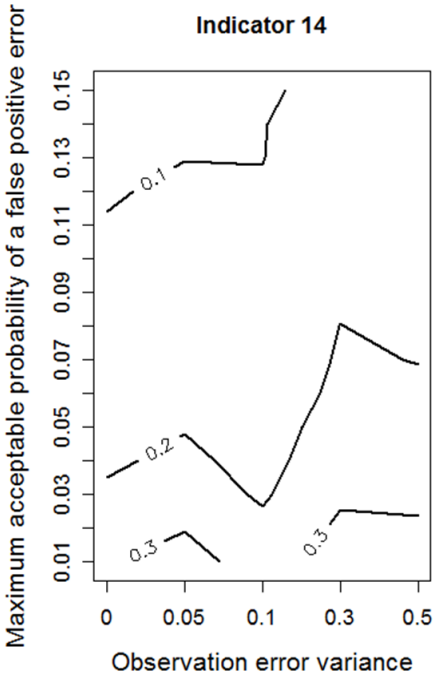

Indicator 11
Maximum acceptable probability of a false positive error
0.01
0.03
0.05
0.07
0.09
0.11
0.13
0.15
0.1
0.2
0.3
0.4
0.01
0.05
0.1
0.3
0.5
Process variation variance

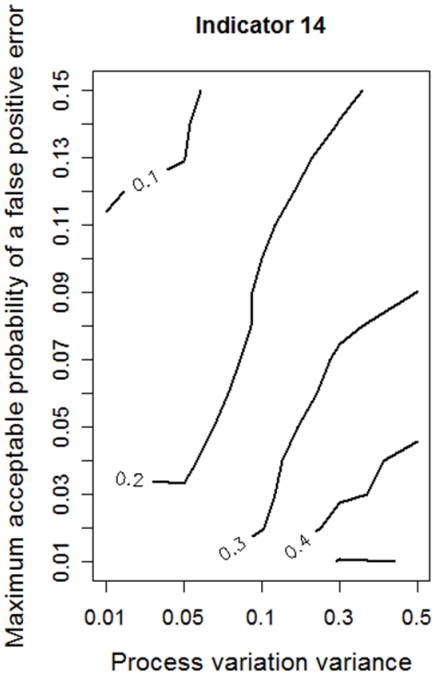

Indicator 12
Maximum acceptable probability of a false positive error
0.01
0.03
0.05
0.07
0.09
0.11
0.13
0.15
0.1
0.2
0.3
0.3
0
0.05
0.1
0.3
0.5
Observation error variance

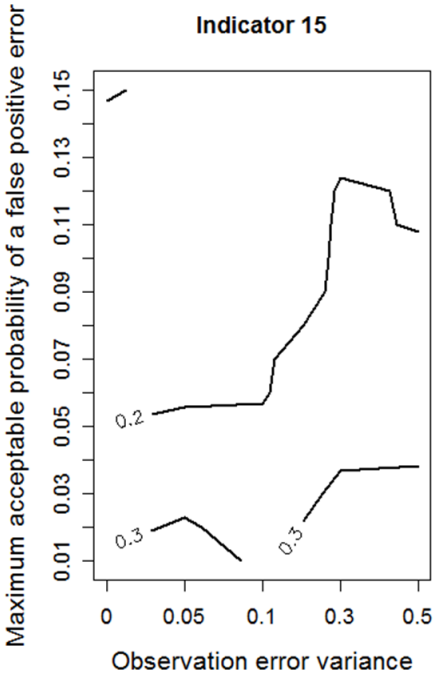

Indicator 12
Maximum acceptable probability of a false positive error
0.01
0.03
0.05
0.07
0.09
0.11
0.13
0.15
0.1
0.2
0.3
0.4
0.01
0.05
0.1
0.3
0.5
Process variation variance

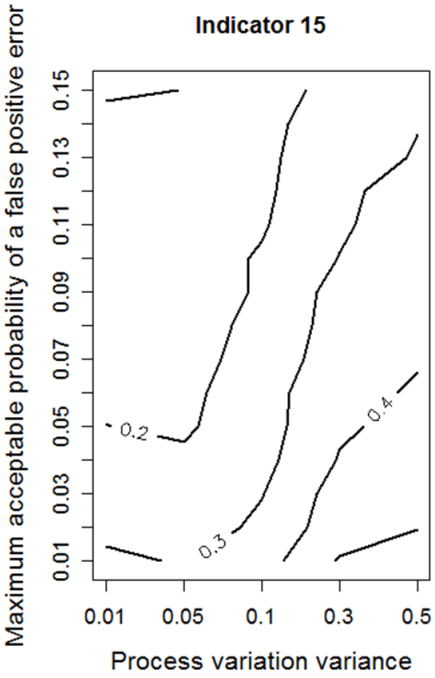

Indicator 13
Maximum acceptable probability of a false positive error
0.01
0.03
0.05
0.07
0.09
0.11
0.13
0.15
0.1
0.2
0.3
0.3
0
0.05
0.1
0.3
0.5
Observation error variance

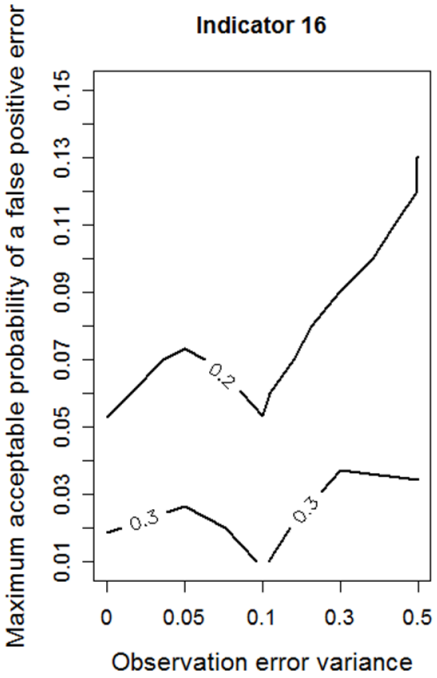

Indicator 13
Maximum acceptable probability of a false positive error
0.01
0.03
0.05
0.07
0.09
0.11
0.13
0.15
0.1
0.2
0.3
0.4
0.01
0.05
0.1
0.3
0.5
Process variation variance

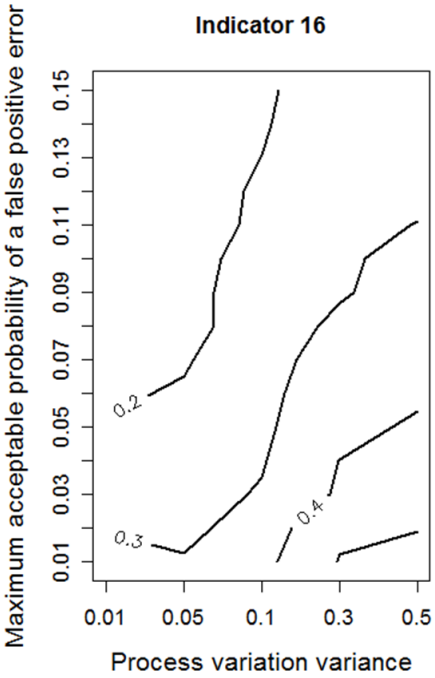

Indicator 14
Maximum acceptable probability of a false positive error
0.01
0.03
0.05
0.07
0.09
0.11
0.13
0.15
0.1
0.2
0.3
0.3
0
0.05
0.1
0.3
0.5
Observation error variance


Indicator 14
Maximum acceptable probability of a false positive error
0.01
0.03
0.05
0.07
0.09
0.11
0.13
0.15
0.1
0.2
0.3
0.4
0.01
0.05
0.1
0.3
0.5
Process variation variance


Indicator 15
Maximum acceptable probability of a false positive error
0.01
0.03
0.05
0.07
0.09
0.11
0.13
0.15
0.2
0.3
0.3
0
0.05
0.1
0.3
0.5
Observation error variance


Indicator 15
Maximum acceptable probability of a false positive error
0.01
0.03
0.05
0.07
0.09
0.11
0.13
0.15
0.2
0.3
0.4
0.01
0.05
0.1
0.3
0.5
Process variation variance


Indicator 16
Maximum acceptable probability of a false positive error
0.01
0.03
0.05
0.07
0.09
0.11
0.13
0.15
0.2
0.3
0.3
0
0.05
0.1
0.3
0.5
Observation error variance


Indicator 16
Maximum acceptable probability of a false positive error
0.01
0.03
0.05
0.07
0.09
0.11
0.13
0.15
0.2
0.3
0.4
0.01
0.05
0.1
0.3
0.5
Process variation variance

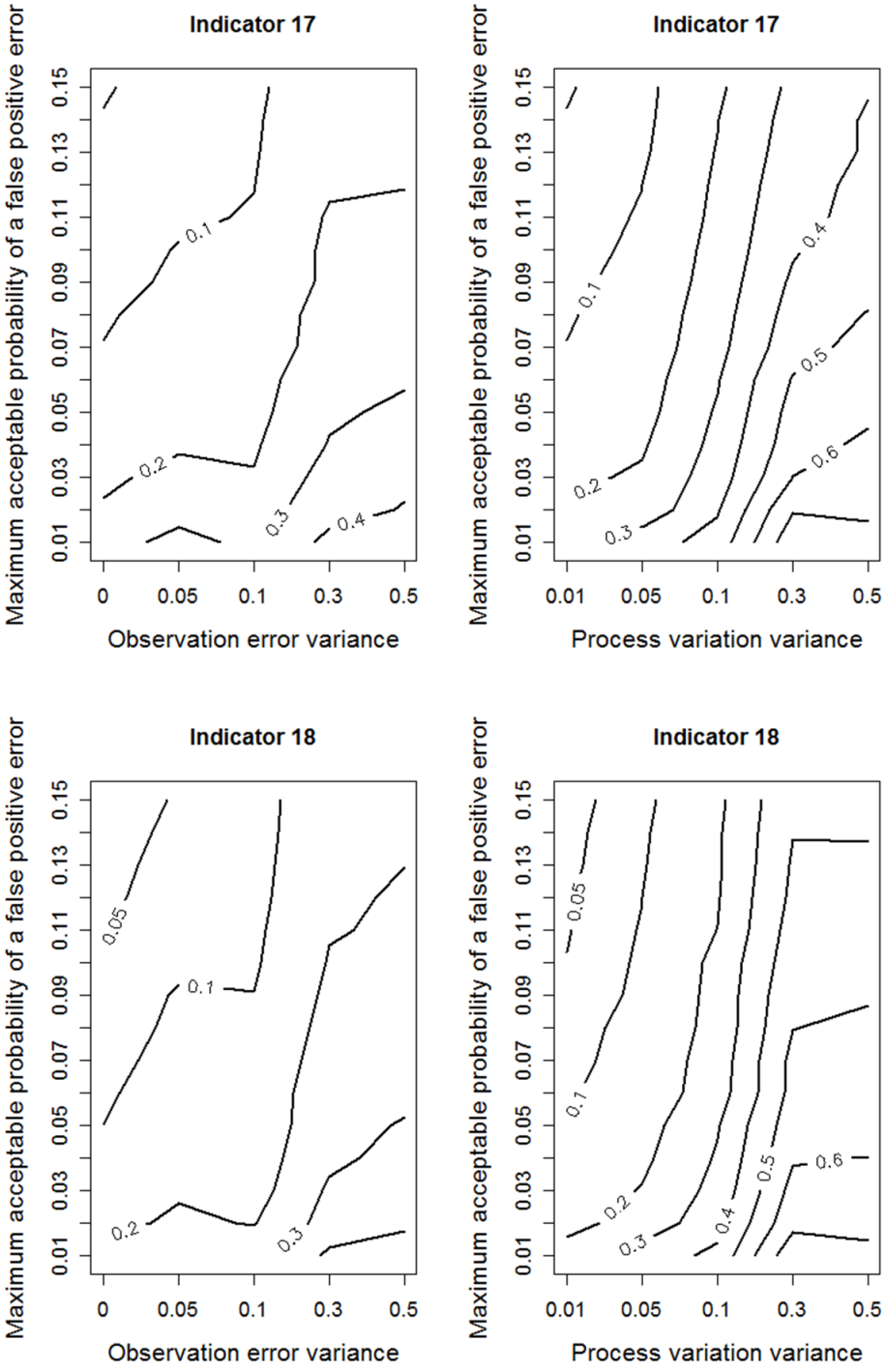

Indicator 17
Maximum acceptable probability of a false positive error
0.01
0.03
0.05
0.07
0.09
0.11
0.13
0.15
0.1
0.2
0.3
0.4
0
0.05
0.1
0.3
0.5
Observation error variance
Indicator 17
Maximum acceptable probability of a false positive error
0.1
0.2
0.3
0.4
0.5
0.6
0.01
0.05
0.1
0.3
0.5
Process variation variance
Indicator 18
Maximum acceptable probability of a false positive error
0.05
0.1
0.2
0.3
0
0.05
0.1
0.3
0.5
Observation error variance
Indicator 18
Maximum acceptable probability of a false positive error
0.05
0.1
0.2
0.3
0.4
0.5
0.6
0.01
0.05
0.1
0.3
0.5
Process variation variance

Figure A6.