Revised October 21 2014

Supporting Information for d'Eon-Eggertson et al.:
"Reliable identification of declining populations in an uncertain world"

Parts of this SI section:

A. Sockeye salmon life history and parameterization of the model

About 94% of adult sockeye salmon (*Oncorhynchus nerka*) from the Fraser River, British Columbia, Canada, reproduce and die at four years of age. Over six decades of abundance records from the Pacific Salmon Commission (Vancouver, B.C.; www.psc.org) document the high temporal variation in abundance of these populations. In some populations, the predominant 4-year life history has created four distinct "cycle lines", with little gene flow among the lines (Ricker 1997). In each 4-year period, one line ("dominant" cycle line) tends to have much higher abundance than the other three lines. The 4-year life span and cycle-line features led us to calculate some indicators of cycle-line-specific values and mean abundances across every 4 years (Table 2 of the main text and Supporting Information (SI) part B below).

We based the simulation model's parameter values on realistic ranges observed for sockeye salmon. To create simulations so that some populations were declining and some were not, we used a normally distributed a parameter in equation 1, with a mean of 1 and a standard deviation of 0.3 ($\sim N(1,0.3)$, but constrained it to not drop below zero). Our a parameter values capture realistic a values but 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). As well, we explored different parameter values for the mean of a (1, 1.4, 1.8) and b (1/100000, 1/200000, 1/400000) to see whether the rank order of indicators (based on their reliability) was sensitive to these changes.

To start each simulation, 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 then simulated a 3-generation (12-year) initialization period, a 52-year evaluation period, which is comparable to the duration of historical time series available for Fraser River sockeye salmon (Porszt et al. 2012), and a subsequent period of 12 years.

Process variation changed randomly for every year, t, of each Monte Carlo trial through a multiplicative error term, e^{ε_t} (Hilborn & Walters 1992), where ε_t is a lag-1-year autocorrelated process (equation 2 in the main text) with an error term u_t that is normally distributed with a mean of zero and a variance σ_u^2, i.e., $u_t \sim N(0, \sigma_u^2)$. Because the abundance with the inclusion of process variation still represents the "true" number of spawners, this error is propagated throughout the model and is reflected in all periods of the time series (i.e., the abundance of spawners in time t is based on the abundance of spawners in time t-4 with process variation included).

Observation error reflects 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 (equation 3 of the main text) with a multiplicative error term e^v, where v was normally distributed with a mean of zero and a variance of σ_v^2 (Walters & Ludwig 1981; Hilborn & Walters 1992). 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. That is, the abundance of spawners in time *t* was based on the "true" abundance of spawners in time *t*-4 with process variation included but without observation error (Wilson et al. 2011).

The values 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) that we explored encompassed most observed ranges of these parameters. Dorner et al. (2009) used a combined error variance (process variation plus observation error) of 0.55 so that their simulated dynamics approximated the total empirically observed interannual variability in the 37 North American sockeye stocks that they used. Peterman and Dorner (2012) used a state-space Ricker model to separately estimate process variance and observation-error variance for the 64 sockeye populations that they analyzed and found medians of 0.26 for process variation, σ_u^2, (0.13 to 0.6 for the 25th and 75th percentiles) and 0.04 for observation error, σ_v^2, (0.01 to 0.12 for the 25th to 75th percentiles) (Brigitte Dorner, personal communication, 10 Dec. 2010, bdorner@driftwoodcove.ca). The full range of estimates extended beyond those percentile values. A recent study by Connors et al. (2014) also used a state-space model to separately estimate process variance and observation-error variance for 627 abundance time series of a wide variety of animal species in the Global Population Dynamics Database (GPDD). However, they used a Gompertz model instead of our Ricker model. Nevertheless, broadly speaking, their estimates of process variance and observation-error variance are likely to be roughly comparable to Ricker-derived values. They found medians of σ_u^2=0.03 for process variation (from 0.01 to 0.16 for the 25th to 75th percentiles) and σ_v^2=0.16 for observation error (from 0.02 to 0.5 for the 25th to 75th percentiles). In their analysis of 1386 time series of abundances for plants as well as animal species in the GPDD, Wilson et al. (2011) also used a Gompertz-based state-space model and estimated the 25 to 75th percentiles of process variance to be 0.004 to 0.23 (median 0.03) and observation-error variance to be 0.002 to 0.1 (median 0.03).

Thus, these empirically estimated parameter values were within the ranges examined in our simulations.

Estimates of lag-1 temporal autocorrelation, Φ, for North American sockeye salmon populations range between -0.47 and +0.79 (mean of 0.22) (Korman et al. 1995 for 30 populations) and -0.21 to +0.77 (mean of 0.31) for the 64 sockeye populations analyzed in Peterman and Dorner (2012) (Brigitte Dorner, personal communication, 11 Feb. 2014, bdorner@driftwoodcove.ca). The range for Φ that we explored (-0.5 to +0.75) essentially covers those observed ranges. For populations other than sockeye salmon, Connors et al. (2014) found that 57% of the time series of abundances of animals in the GPDD had zero lag-1 autocorrelation in process variation; the other species ranged from -0.4 to +0.95.

Thus, our sensitivity analyses across a wide range of values of process variation, autocorrelation in that variation, and observation error evaluated how robust the different indicators of population decline were to various situations that potentially reflect real-world conditions.

B. Details of the 20 indicators of population decline

Our analysis expands upon the Porszt et al. (2012) retrospective analysis of past observed data, which examined many of the same indicators as we did. However, their empirically based analysis only evaluated indicators over the single set of observed historical events. In contrast, our simulation analysis is much broader and more comprehensive than Porszt et al. (2012) in that we explore a wide range of plausible future sequences of events as delineated by the process variation and observation error that we included.

Many organizations, such as the Canadian Committee on the Status of Endangered Wildlife in Canada (COSEWIC 2011), CITES, U.S. EPA, and others use indicators similar to

some of those listed below. The following indicators were calculated during the "evaluation period" (Figure 1a).

1. Percent decline in spawner abundance over the most recent three generations (i.e., 12 years in the case of Fraser River Sockeye salmon) prior to each assessment year, estimated by exponentiation of best-fit values from the robust regression of the annual values for $\log_e$(unsmoothed abundance) on year. Robust regression reduces the influence of outliers (Venables & Ripley 2002). In other words, the rate of decline indicators (#1 and #2) examine all of the possible 3-generation periods over the evaluation period of the time series; they calculate the rate of decline in 12-year blocks of years 13-24, 14-25, 15-26, … up through years 53-64. In comparison, the indicators of historical extent of decline look at years 13-24, 13-25, 13-26, … up through 13-64.
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 later 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). Many populations of Fraser River sockeye salmon show four distinct successive "cycle lines" arising from the 4-year life span, with little gene flow among the lines (Ricker 1997; SI part A). Therefore, for "cycle-line" indicators, we calculated trends in abundance by "cycle

line", that is by only using data from every fourth year, e.g., years 1, 5, 9, etc. for cycle-line 1, and years 2, 6, 10, etc. for cycle-line 2, etc.

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 the geometric mean of raw abundances of the first 4-year generation and the geometric mean of another 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 raw abundances of the first 4-year generation and the raw abundance in an 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.
19. Same as indicator #3 (decline since the first year in the data set), except we calculated a three-generation (12-year) rate of decline from the regression.
20. Same as indicator #19 except we used smoothed (4-year running mean) abundances rather than unsmoothed abundances.

For each relevant simulated year of the 13-generation evaluation period, we determined for each of the 20 indicators whether the indicator classified the current status of the population as "declining" or "non-declining". To do this, in a set of separate calculations, we compared the estimated decrease in abundance against various thresholds that delineated "declining"; those thresholds ranged from 0 to 100% in increments of 1%. A population was classified as "declining" if the estimated decline was greater than a given threshold shown on the X-axis, but was classified as "non-declining" if the estimated decline was less than that threshold. For example, a given population at year 50 would be assessed using indicators of recent rate of decline, which would calculate a regression over the previous 12 years (years 39-50). If this regression corresponded to a decline of, say, 40%, then the assessed status at a threshold of 1% would be classified as “declining” but the assessed status at a threshold of 99% would be classified as “non-declining”, and so on for thresholds between 0 and 100%. This example population at year 50 would also be assessed using historical baseline indicators, which would

calculate a regression on all years since the beginning of the evaluation period (years 13-50). The decline estimated from this regression would also be compared to thresholds from 0 to 100% in order to classify the population as "declining" or "non-declining" in each case. This analysis would be completed for all indicators, for all applicable years, and for each population.

C. Overall long-term population trend

To classify this overall long-term status, we first took a robust linear regression of the final 12 years in the time series (i.e., years 65-76; the "subsequent period" in Figure 1b) and found the mean spawner abundance of the last 4-year generation (i.e., across years 73-76), as estimated from this regression. We then 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 in those mean abundances from the end of the initialization period to that last generation of the subsequent period as the estimate of the overall long-term trend in the population.

We wanted to identify which indicators that were estimated during the "evaluation period" most reliably reflected the actual simulated long-term trend. Therefore, for the steps in our estimation of the overall long-term trend, observation error was not included in the initialization and subsequent periods of the time series; for these periods, the "true" population abundance, which reflected process variation, was considered as being known for the purposes of our simulations.

D. Receiver Operating Characteristic (ROC) analysis

In medicine, a Receiver Operating Characteristic (ROC) analysis is often used to evaluate the reliability of diagnostic tests (Hibberd & Cooper 2008). ROC analyses combine into a single

196 measure of reliability the true and false positive error rates (which are the complements of the
197 true and false negative rates, see Table 1 of the main text) produced by a test or indicator across
198 different classification thresholds. This type of analysis has recently been used in extinction-risk
199 studies (e.g. Porszt et al. 2012), biogeography to compare statistical models of habitat (Pearce
200 and Ferrier 2000), and management of invasive species (Baxter and Possingham 2011).
201 In our analysis, we combined into a single metric for each decline indicator the true and
202 false positive rates that were produced across a wide range of thresholds of decline in abundance
203 for classifying population status. Specifically, for each threshold (0 to 100% in 1% increments),
204 the true positive rate (TP/(TP+FN)) was plotted against the false positive rate (FP/(FP+TN)),
205 which created 101 points and generated an ROC curve (examples in Figure 1c). The resulting
206 area under the ROC curve (AUC) reflects the ability of a given indicator to correctly distinguish
207 whether a population is declining; higher AUC values mean a more reliable indicator (Hibberd &
208 Cooper 2008; Porszt et al. 2012).
209
210 **E. Additional results and sensitivity analyses**
211 This section contains figures of detailed main results for the 20 indicators in our study, as
212 well as results of sensitivity analyses.
213

Table S1. Tables of the most reliable indicator (numbers defined in Table 2 and SI part B), that is, the one with the largest area under the ROC curve (AUC), which means the greatest ability to correctly classify a population's abundance as decreasing or not. Results are shown for different ratios of weightings (importance of avoidance) placed on rates of false positive (FP) errors to false negative (FN) errors for different levels of (a) process variation, i.e., environmentally-driven variability in productivity (at observation error σ_v^2=0) and (b) observation error (at process variation σ_u^2=0.01), with no temporal autocorrelation in process variation. Analogous cases with an autocorrelation coefficient of 0.5 are also shown (c and d).

(a)

Process variance (σ_u^2)	Best indicator if a false positive is *x* times as important to avoid as a false negative *x:*								
	0.20	0.25	0.33	0.50	1.0	2.0	3.0	4.0	5.0
0.01	20	20	20	20	20	20	20	20	4
0.05	20	20	20	20	20	9	9	6	10
0.1	20	19	20	20	20	11	11	6	6
0.3	4	20	20	20	9	7	6	7	7
0.5	11	10	10	9	4	6	10	8	8

226 Table S1
227

(b)

Observation error variance (σ_v^2)	Best indicator if a false positive is *x* times as important to avoid as a false negative *x:* 0.20	0.25	0.33	0.50	1.0	2.0	3.0	4.0	5.0
0	20	20	20	20	20	20	20	20	4
0.05	4	4	4	4	4	4	4	4	7
0.1	20	4	4	20	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

(c)

Process variance (σ_u^2)	Best indicator if a false positive is *x* times as important to avoid as a false negative *x:* 0.20	0.25	0.33	0.50	1.0	2.0	3.0	4.0	5.0
0.01	20	19	20	20	19	19	18	4	18
0.05	20	20	20	20	20	4	9	9	9
0.1	20	9	9	9	11	9	11	6	9
0.3	9	9	9	11	11	11	11	11	9
0.5	7	7	7	7	7	7	9	7	18

228

229 Table S1
230

(d)

Observation error variance (σ_v^2)	Best indicator if a false positive is *x* times as important to avoid as a false negative *x:*								
	0.20	0.25	0.33	0.50	1.0	2.0	3.0	4.0	5.0
0	20	19	20	20	19	19	18	4	18
0.05	20	20	19	20	19	6	20	7	18
0.1	20	20	20	20	20	4	4	4	6
0.3	20	20	20	6	4	6	6	4	6
0.5	20	20	20	20	20	6	6	6	6

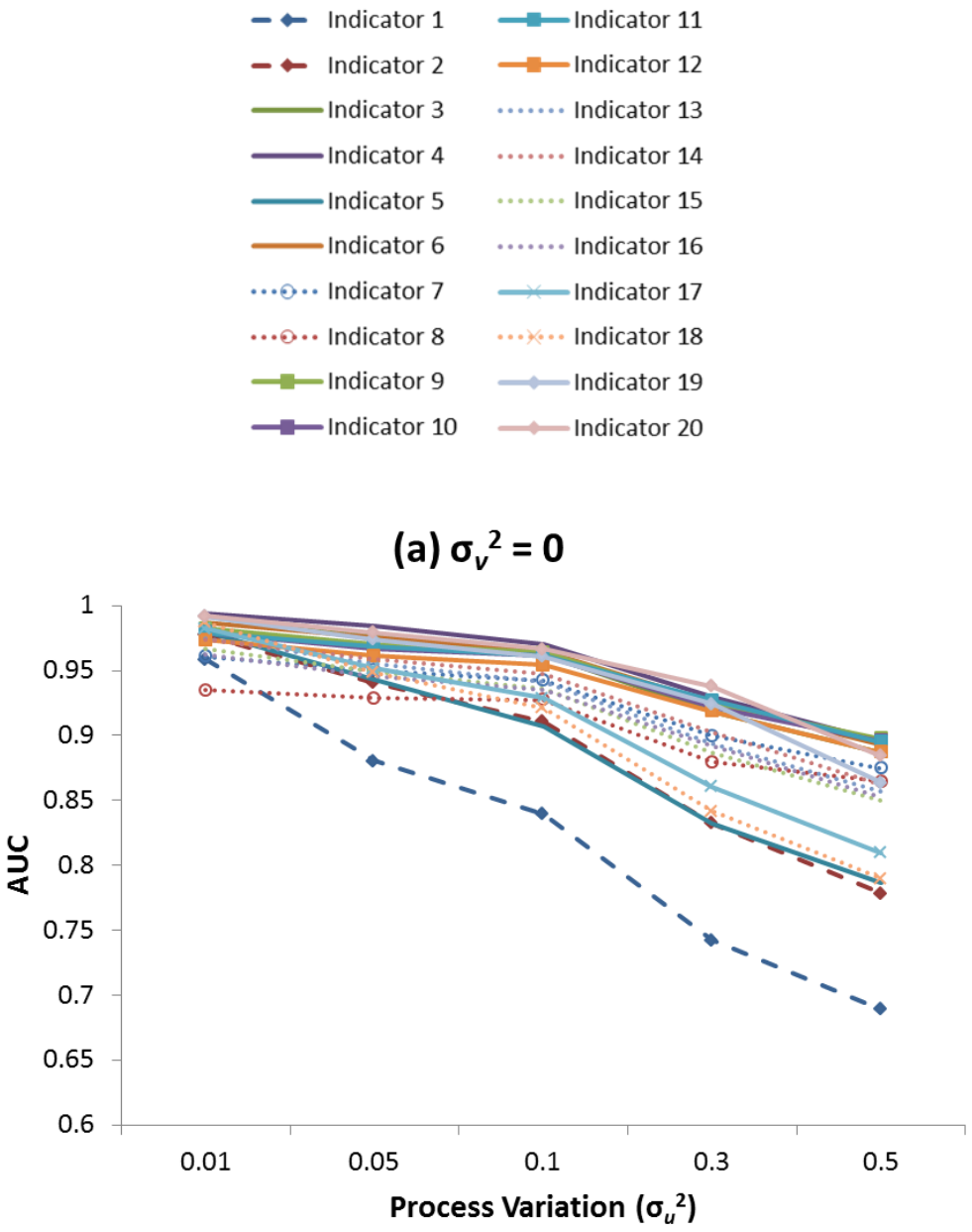
231

232

233

234 **Figure S1.** Changes in AUC as a function of increasing process variation (σ_u^2) at different levels
235 of observation error (σ_v^2 = 0, 0.05, 0.1, 0.3, 0.5 in panels a-e, respectively) for all indicators.
236 Higher AUC values reflect a better ability of an indicator to differentiate between a declining and
237 non-declining population. Table 2 and Supporting Information (SI) part B define the indicators.
238 Thick broken dashed lines with solid symbols are for indicators that are based on decline in the
239 last 3 generations (indicators 1 and 2). Solid lines are for indicators that are based on decline
240 from some historical baseline (indicators 3-6, 9-12, 17, and 19-20). Thin dotted lines are for
241 indicators based on decline from a maximum abundance (indicators 7-8, 13-16, and 18).
242

243 Figure S1

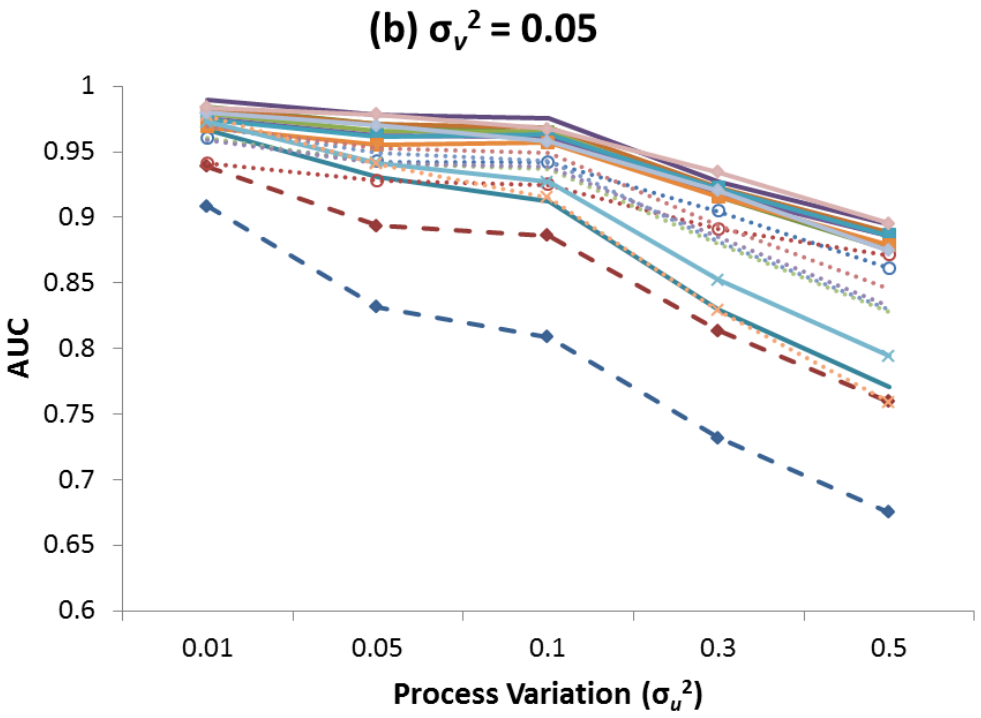


244

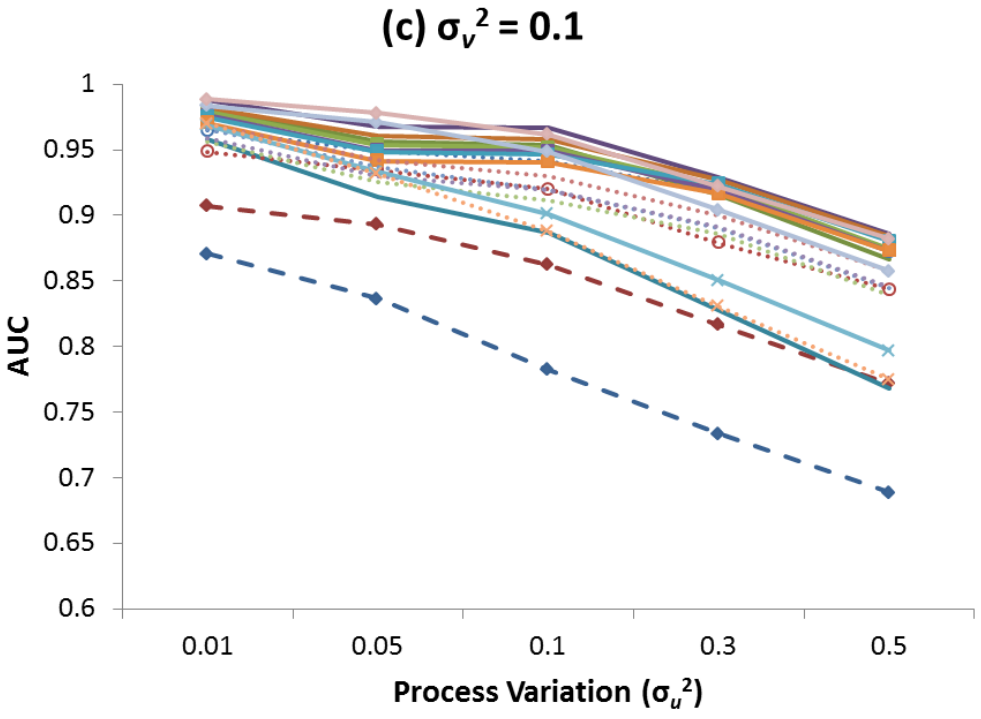
245

246

247 Figure S1



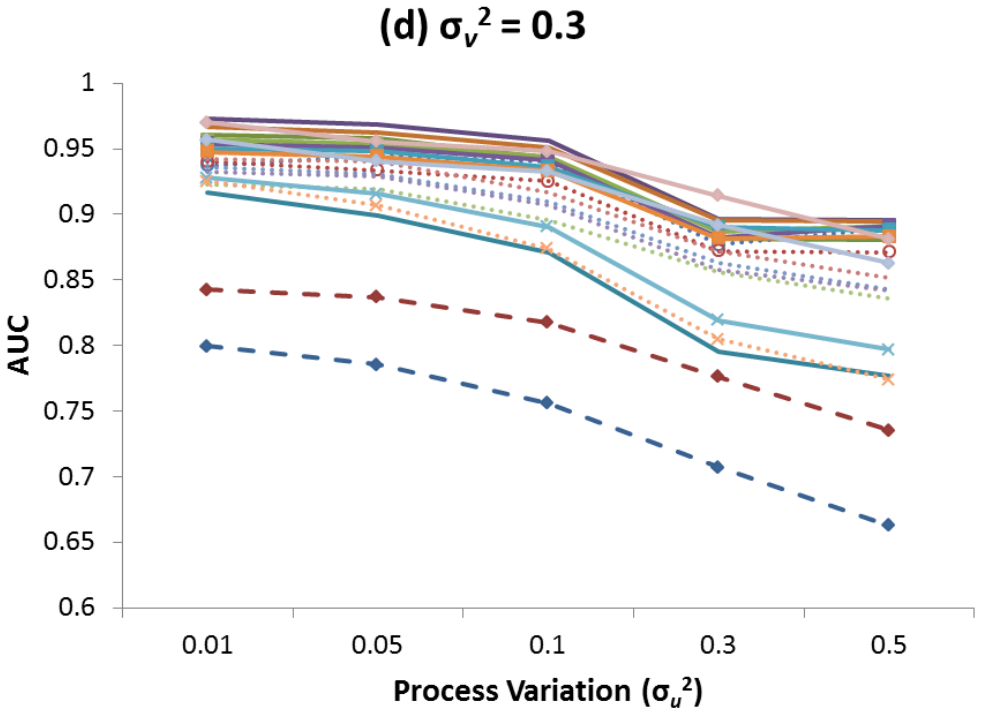
248



249

250

Figure S1


(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 (σ_u^2)

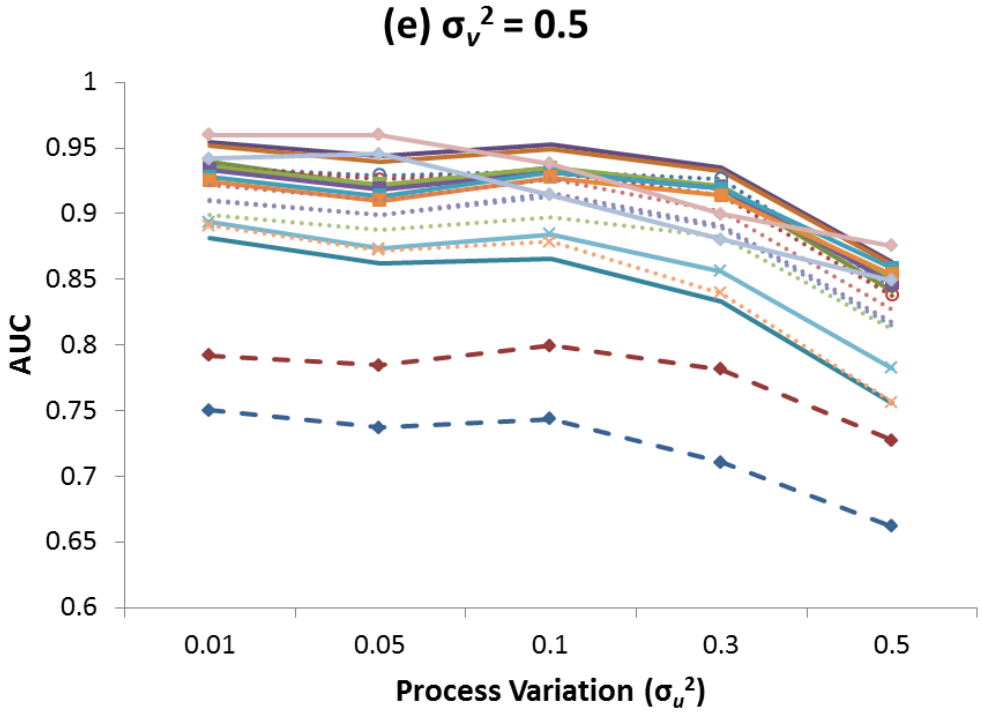
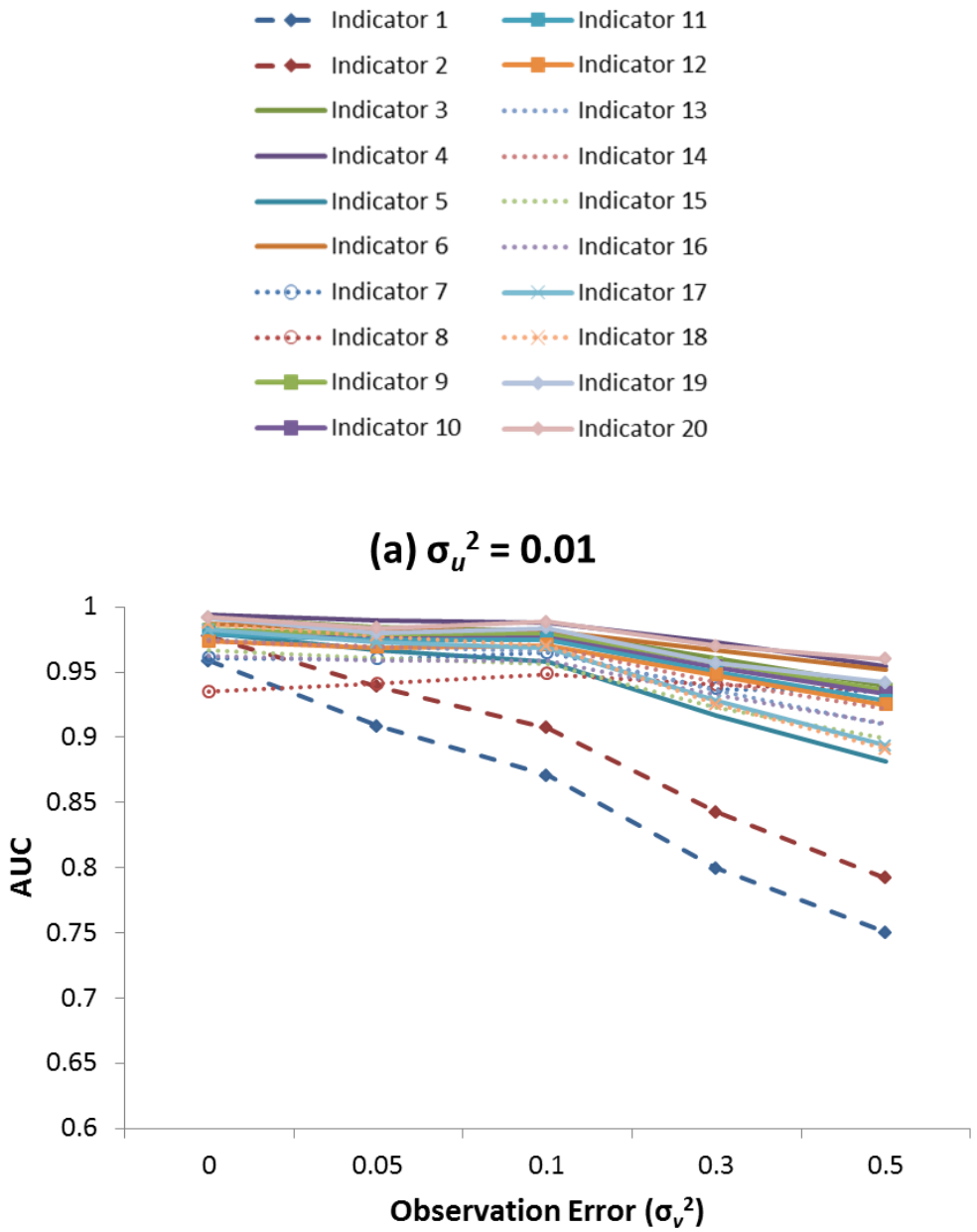

(e) $\sigma_v^2 = 0.5$
AUC
1
0.95
0.9
0.85
0.8
0.75
0.7
0.65
0.6
0.01
0.05
0.1
0.3
0.5
Process Variation (σ_u^2)

Figure S2. Changes in AUC as a function of increasing observation error (σ_v^2) at different levels of process variation (σ_u^2 = 0, 0.05, 0.1, 0.3, 0.5 in panels a-e, respectively) for all indicators. Higher AUC values reflect a better ability of an indicator to differentiate between a declining and non-declining population. Table 2 and Supporting Information (SI) part B define the indicators. Thick broken dashed lines with solid symbols are for indicators that are based on decline in the last 3 generations (indicators 1 and 2). Solid lines are for indicators that are based on decline from some historical baseline (indicators 3-6, 9-12, 17, and 19-20). Thin dotted lines are for indicators based on decline from a maximum abundance (indicators 7-8, 13-16, and 18).

263 Figure S2

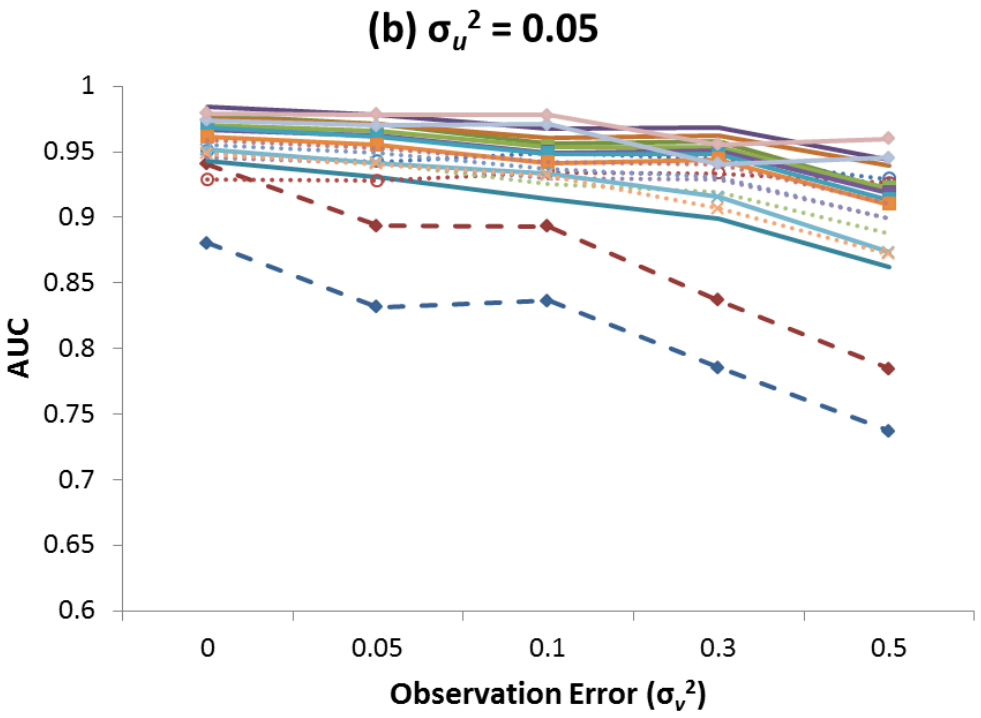


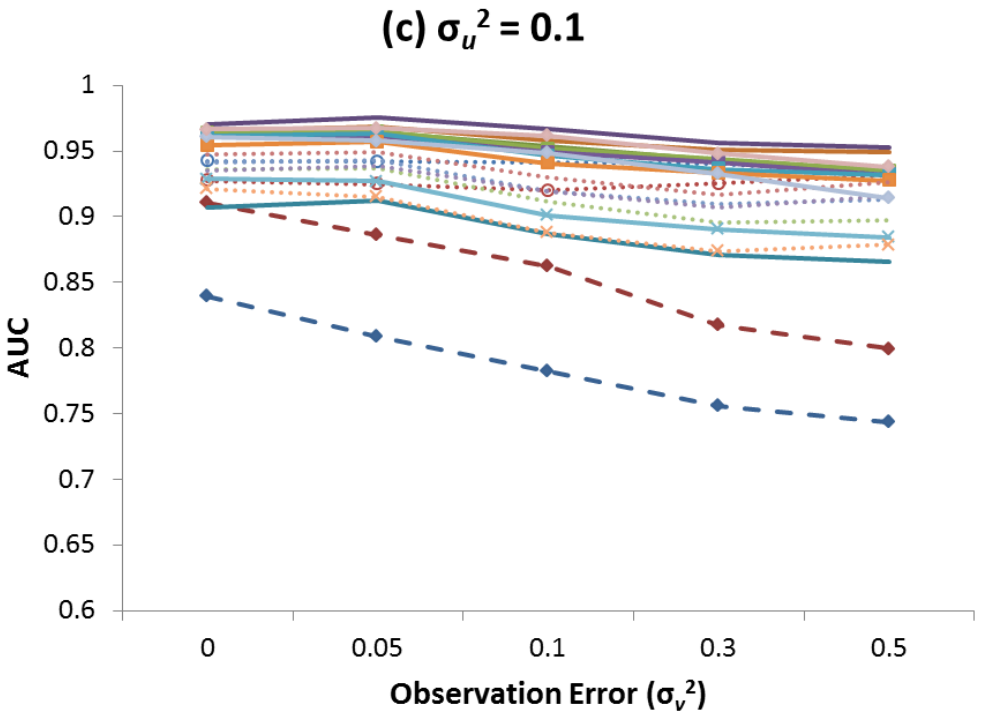
264

265

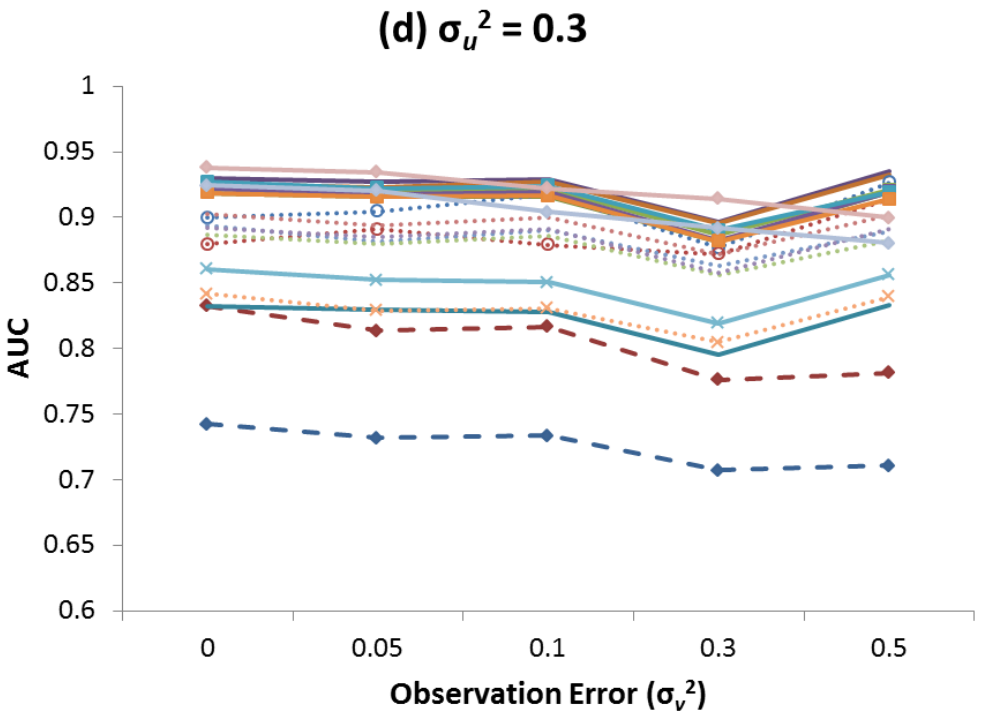
266

Figure S2

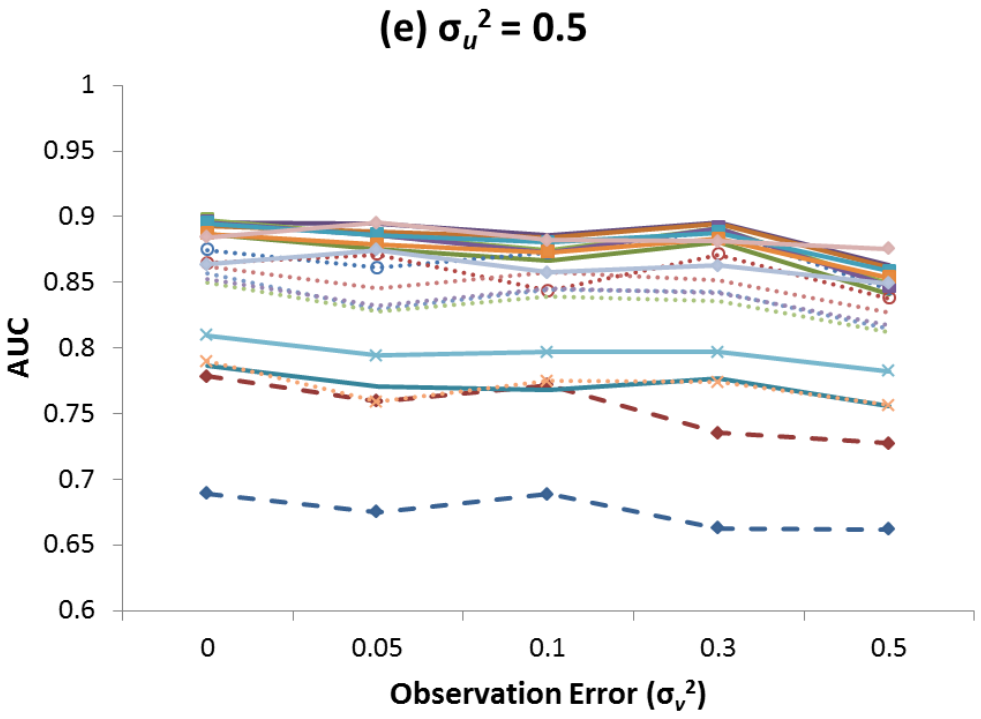

(b) $\sigma_u^2 = 0.05$
AUC
1
0.95
0.9
0.85
0.8
0.75
0.7
0.65
0.6
0
0.05
0.1
0.3
0.5
Observation Error (σ_v^2)


(c) $\sigma_u^2 = 0.1$
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 (σ_v^2)

271 Figure S2



272



273

274

Figure S3. Changes in AUC as a function of the boundary used to define a long-term status of a declining population for all indicators, with no observation error ($\sigma_v^2 = 0$) and low process variation ($\sigma_u^2 = 0.01$). Higher AUC values reflect a better ability of an indicator to differentiate between a declining and non-declining population. Table 2 and Supporting Information (SI) part B define the indicators. Thick broken dashed lines with solid symbols are for indicators that are based on decline in the last 3 generations (indicators 1 and 2). Solid lines are for indicators that are based on decline from some historical baseline (indicators 3-6, 9-12, 17, and 19-20). Thin dotted lines are for indicators based on decline from a maximum abundance (indicators 7-8, 13-16, and 18).

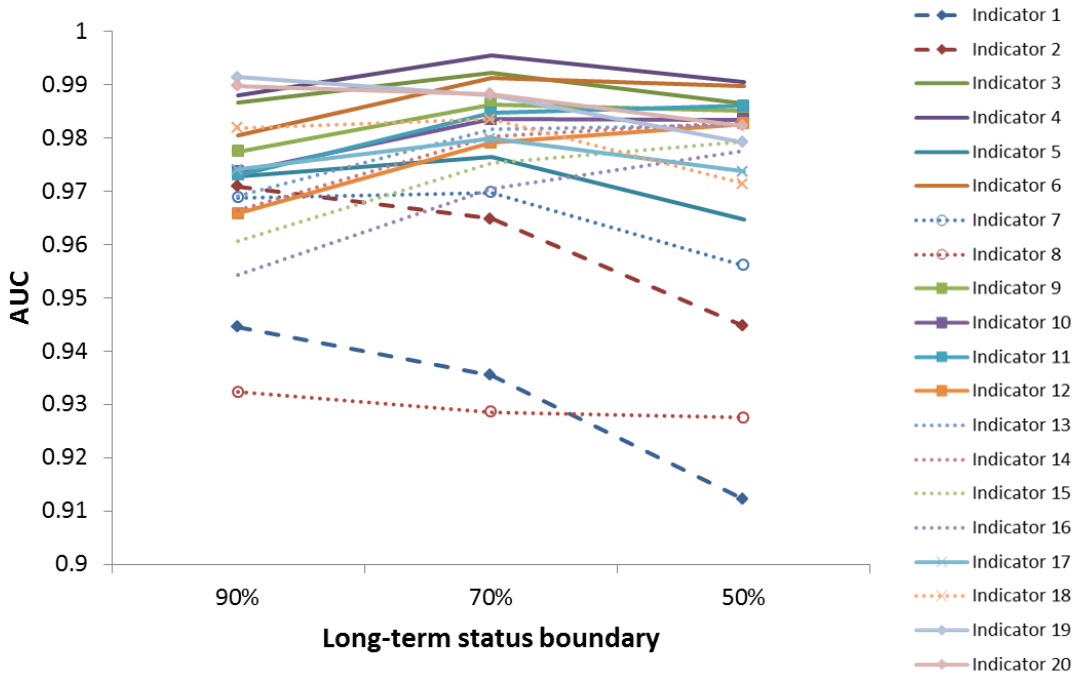


Figure S4. Changes in AUC as a function of the values of the *a* (panel a) and *b* (panel b) parameters in equation 1 of the main paper for all indicators. Higher AUC values reflect a better ability of an indicator to differentiate between a declining and non-declining population. Increasing *a* or decreasing *b* results in more productive populations, leading to fewer populations being classified as declining using our initial boundaries. Without both declining and non-declining populations, it is impossible to calculate an ROC curve. To avoid this problem, we used, for only this particular sensitivity analysis, the mean percentage decline generated by each set of populations as the boundary that defines a declining population. Table 2 and Supporting Info. (SI) part B define the indicators. Thick broken dashed lines with solid symbols are for indicators that are based on decline in the last 3 generations (indicators 1 and 2). Solid lines are for indicators that are based on decline from some historical baseline (indicators 3-6, 9-12, 17, and 19-20). Thin dotted lines are for indicators based on decline from a maximum abundance (indicators 7-8, 13-16, and 18).

Figure S4

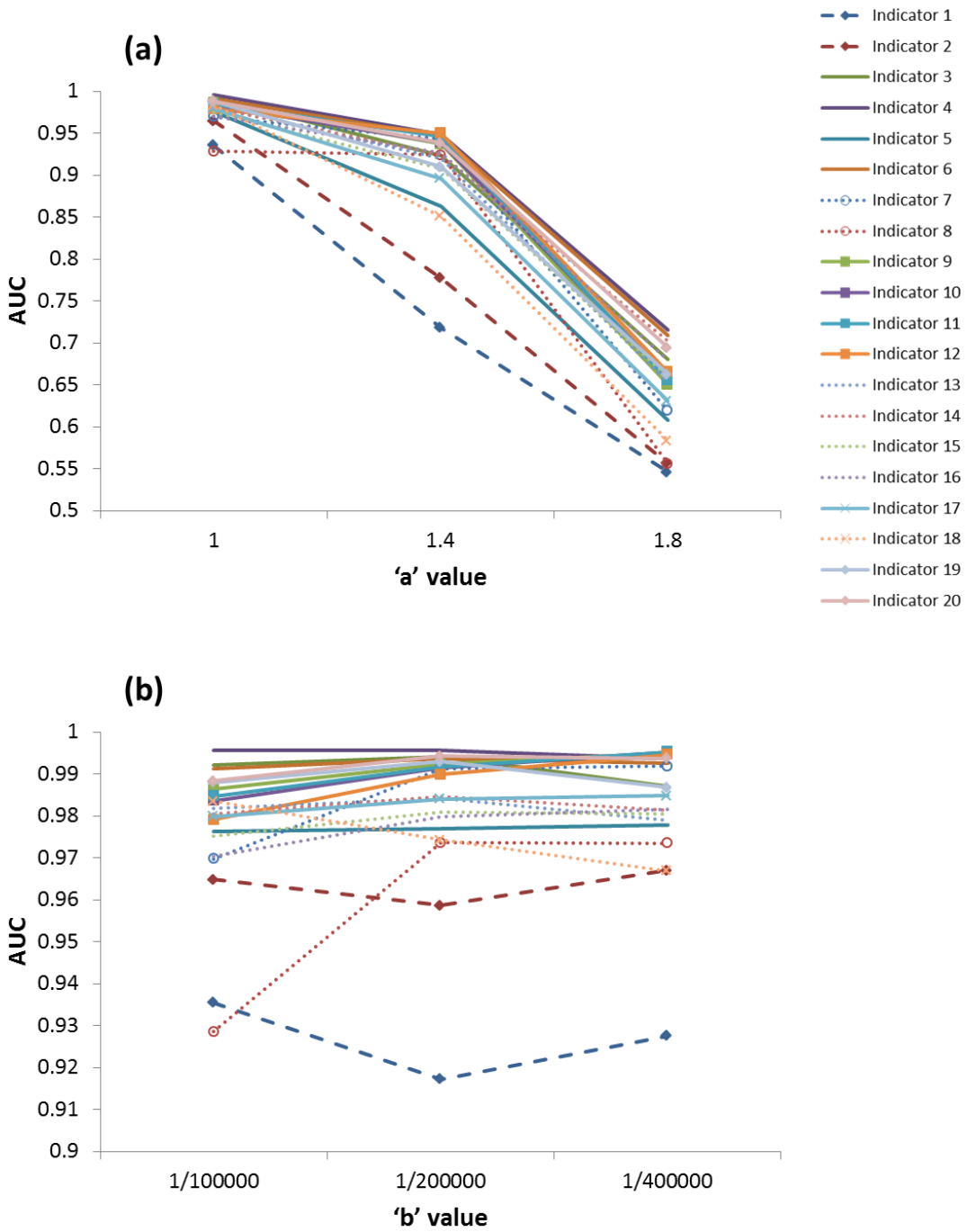

(a)
AUC
1
0.95
0.9
0.85
0.8
0.75
0.7
0.65
0.6
0.55
0.5
1
1.4
1.8
'a' value
Indicator 1
Indicator 2
Indicator 3
Indicator 4
Indicator 5
Indicator 6
Indicator 7
Indicator 8
Indicator 9
Indicator 10
Indicator 11
Indicator 12
Indicator 13
Indicator 14
Indicator 15
Indicator 16
Indicator 17
Indicator 18
Indicator 19
Indicator 20
(b)
AUC
1
0.99
0.98
0.97
0.96
0.95
0.94
0.93
0.92
0.91
0.9
1/100000
1/200000
1/400000
'b' value

Figure S5. The false positive rate (FPR, dark blue) and false negative rate (FNR, red) across threshold levels (ranging from 0 to 100%) for classifying a population as declining during the evaluation period, given a boundary condition of 90% for classification of the long-term trend. Results are for the base-case scenario with low process variation and no observation error (σ_u^2 = 0.01, σ_v^2 = 0). The purple 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). If avoiding a false negative is considered twice as important as avoiding a false positive (i.e., that the false negative rate must be half the false positive rate), then the lowest false negative error rate possible is at the threshold of population decline indicated by where the left-hand orange vertical line intersects the red curve showing the false negative rate. The converse situation is shown at the right-hand orange line. Table 2 and Supporting Information (SI) part B define the indicators.

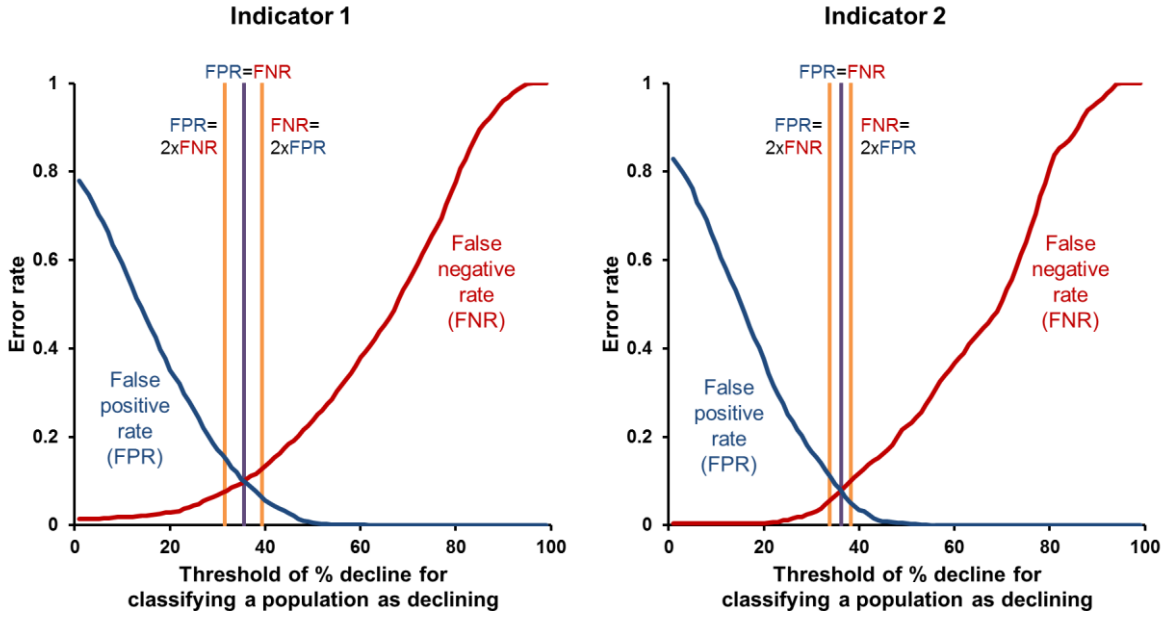


Figure S5

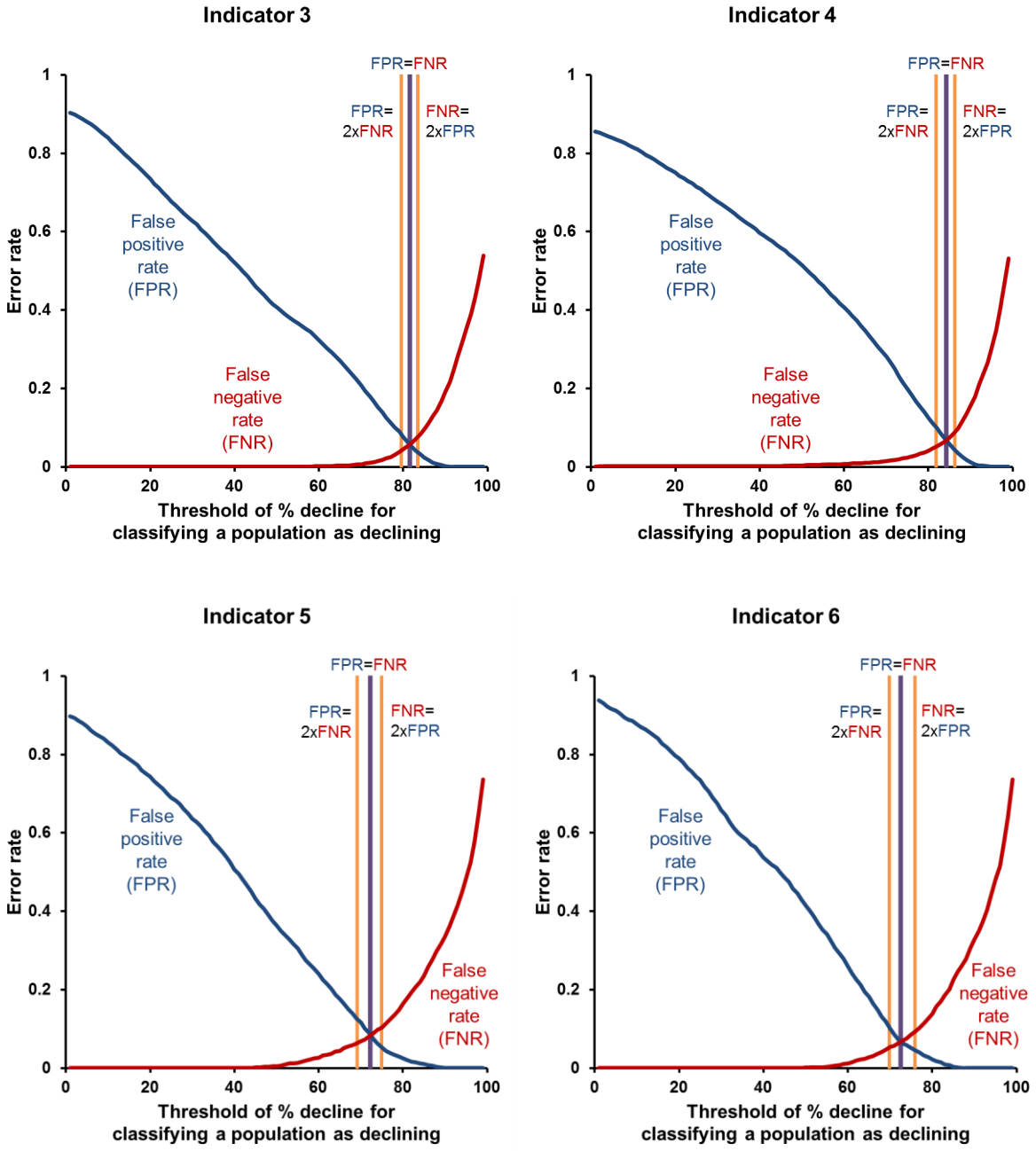

Indicator 3
Indicator 4
Indicator 5
Indicator 6
FPR=FNR
FPR= 2xFNR
FNR= 2xFPR
False positive rate (FPR)
False negative rate (FNR)
Error rate
Threshold of % decline for classifying a population as declining
0
0.2
0.4
0.6
0.8
1
0
20
40
60
80
100

Figure S5

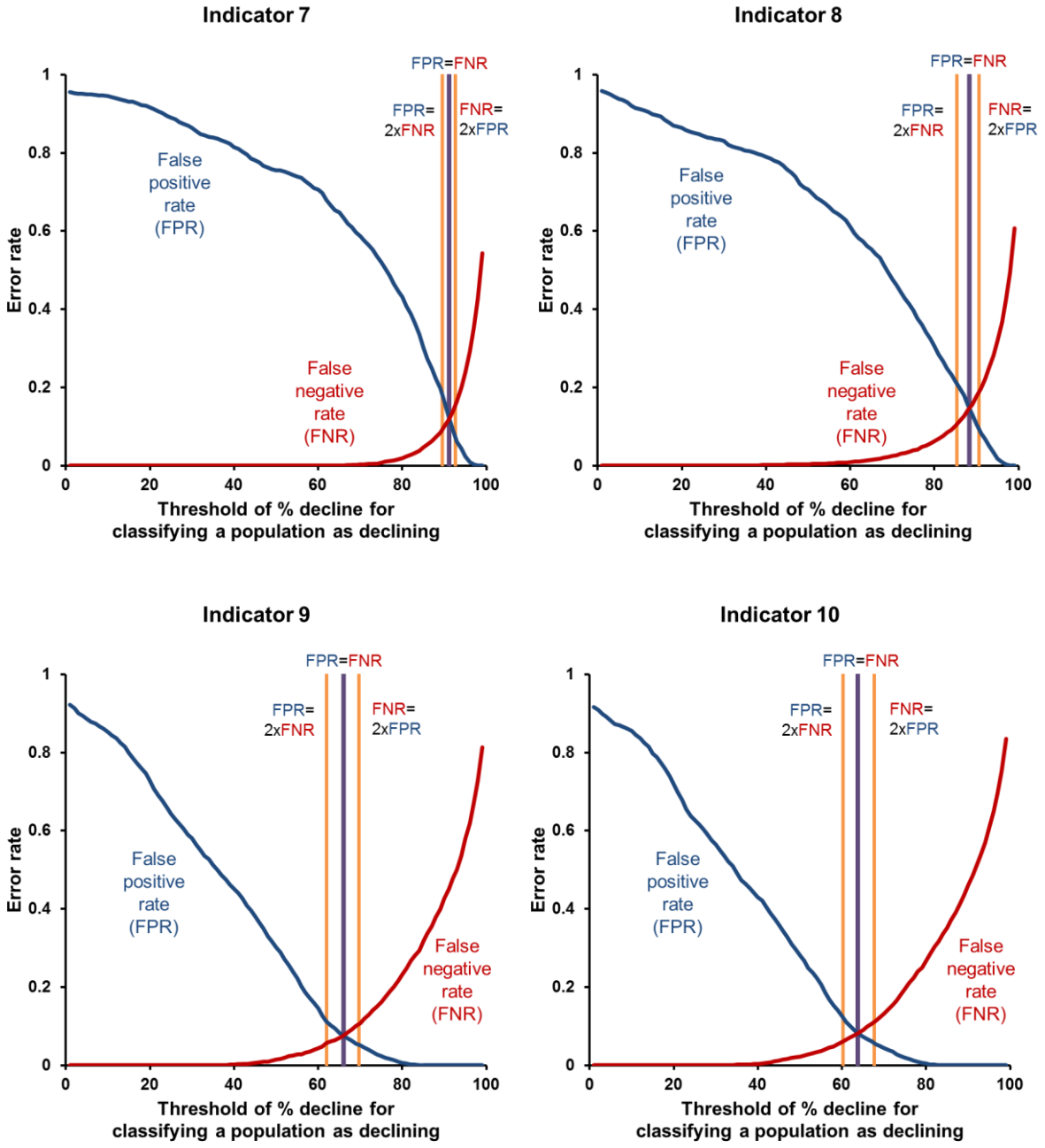
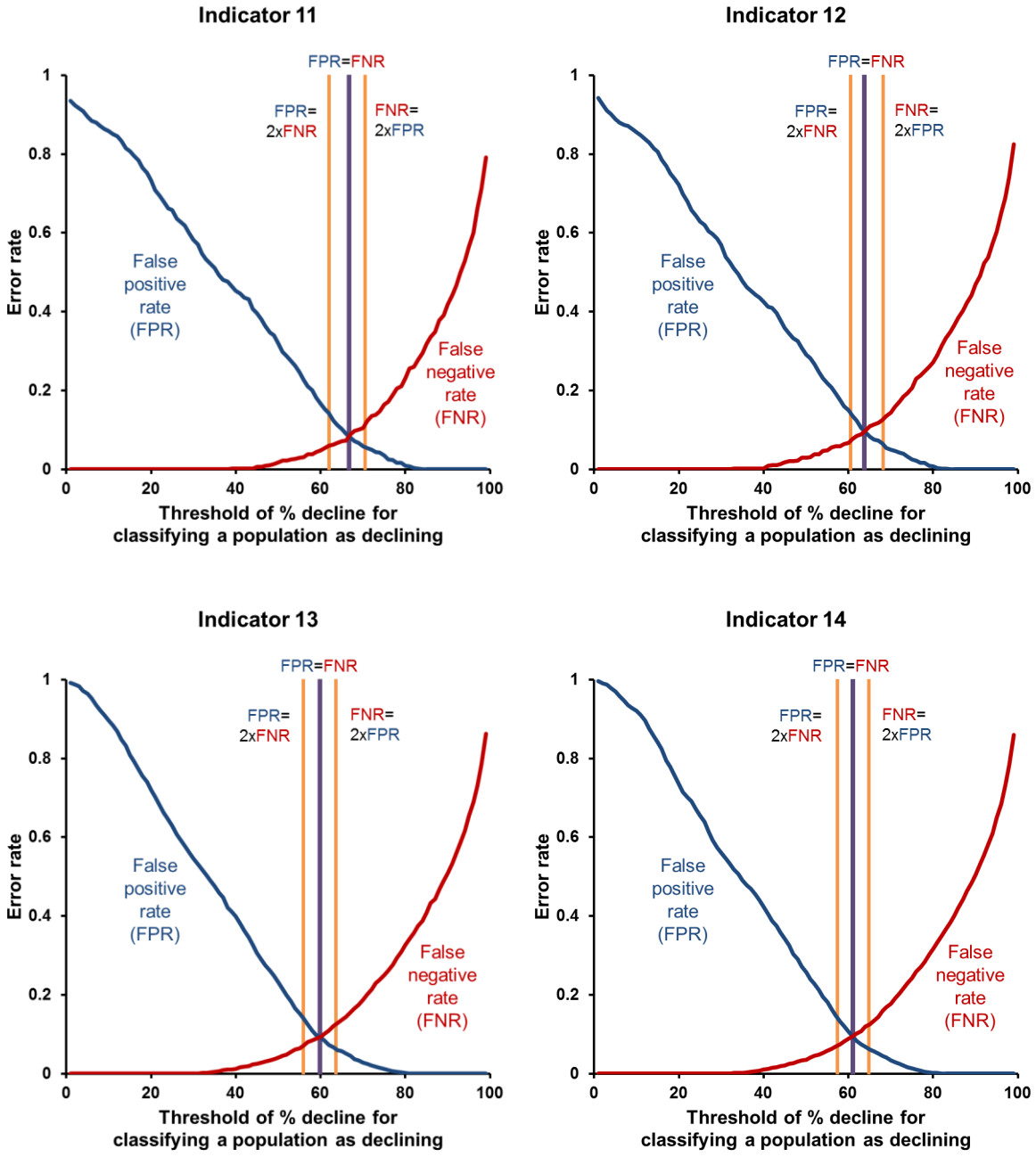
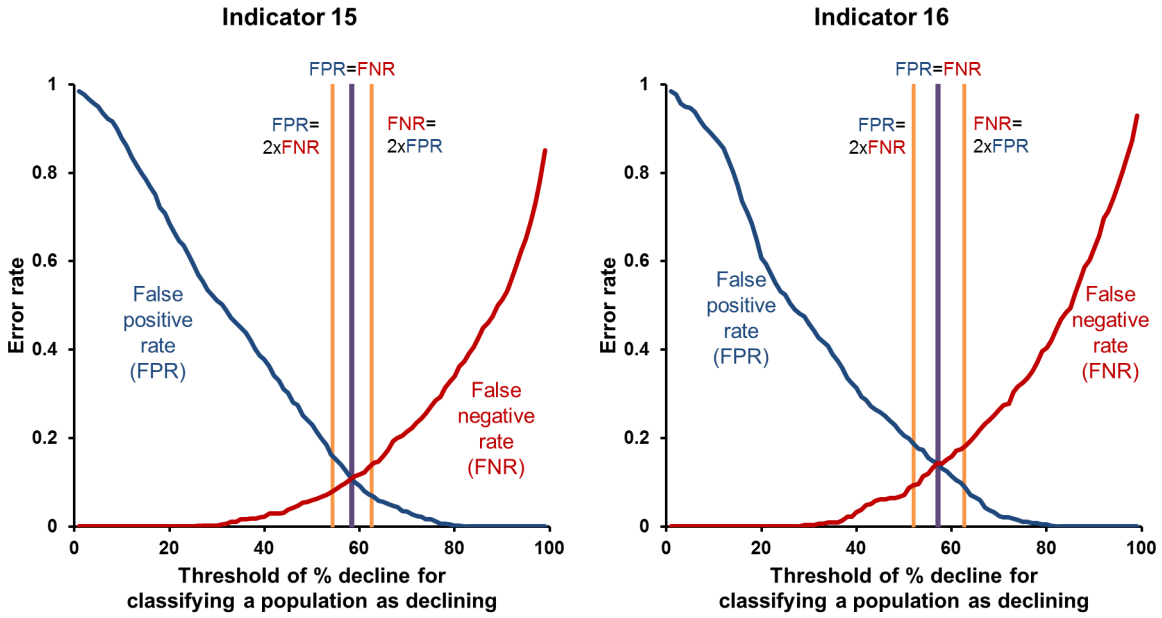

Indicator 7
Indicator 8
Indicator 9
Indicator 10
FPR=FNR
FPR= 2xFNR
FNR= 2xFPR
False positive rate (FPR)
False negative rate (FNR)
Error rate
Threshold of % decline for classifying a population as declining

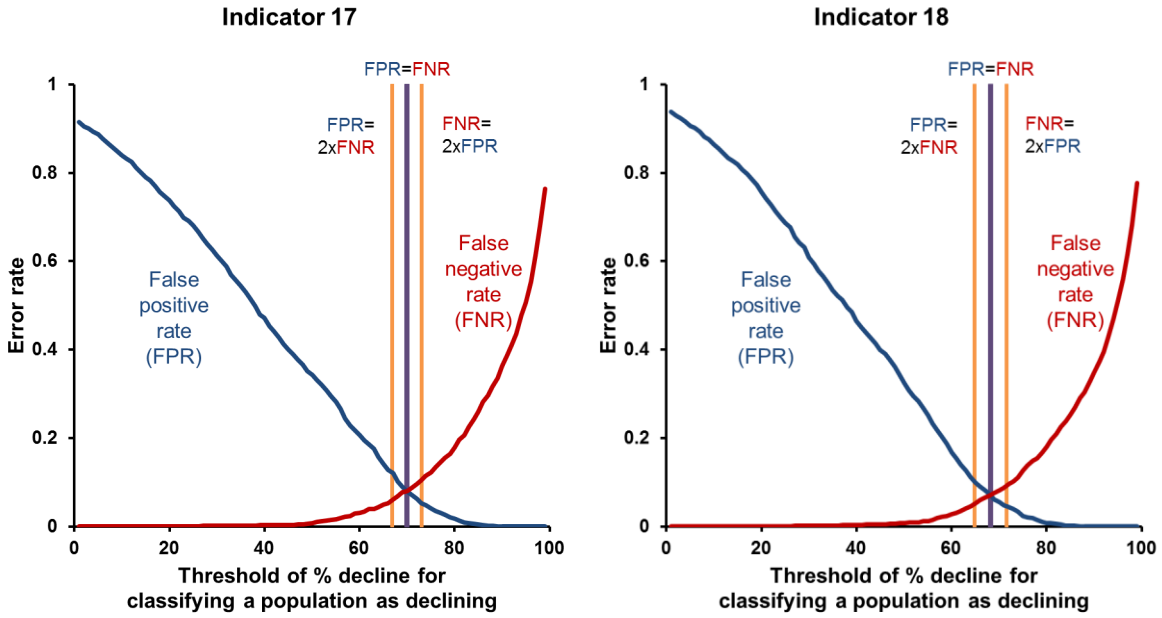
Figure S5


Indicator 11
Indicator 12
Indicator 13
Indicator 14
FPR=FNR
FPR= 2xFNR
FNR= 2xFPR
False positive rate (FPR)
False negative rate (FNR)
Error rate
Threshold of % decline for classifying a population as declining

333 Figure S5



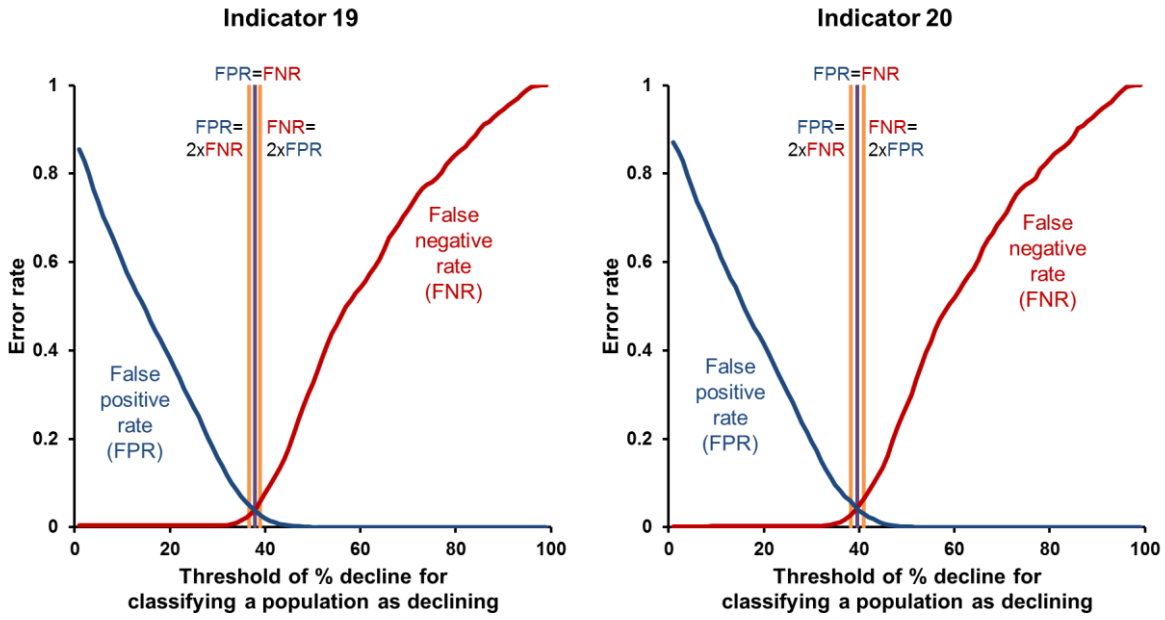
334



335

336

337 Figure S5



338
339

Figure S6. Plots of the minimum rate of false negative (FN) errors (probability values on contours) that can be obtained if the false positive (FP) rate is constrained to be below the value specified on the Y-axis. False negative rates are shown as a function of observation error (σ_v^2, left panels) and process variation (σ_u^2, right panels). Left-hand panels were generated assuming low process variation ($\sigma_u^2 = 0.01$); right-hand panels were generated assuming no observation error ($\sigma_v^2 = 0$). Table 2 and Supporting Information (SI) part B define the indicators.

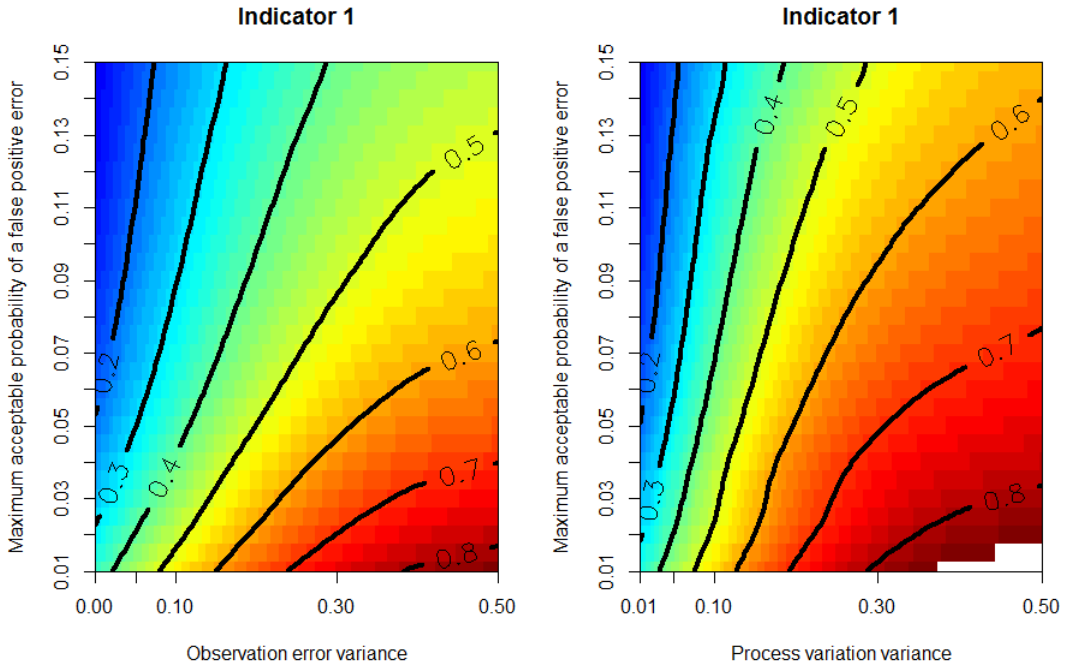
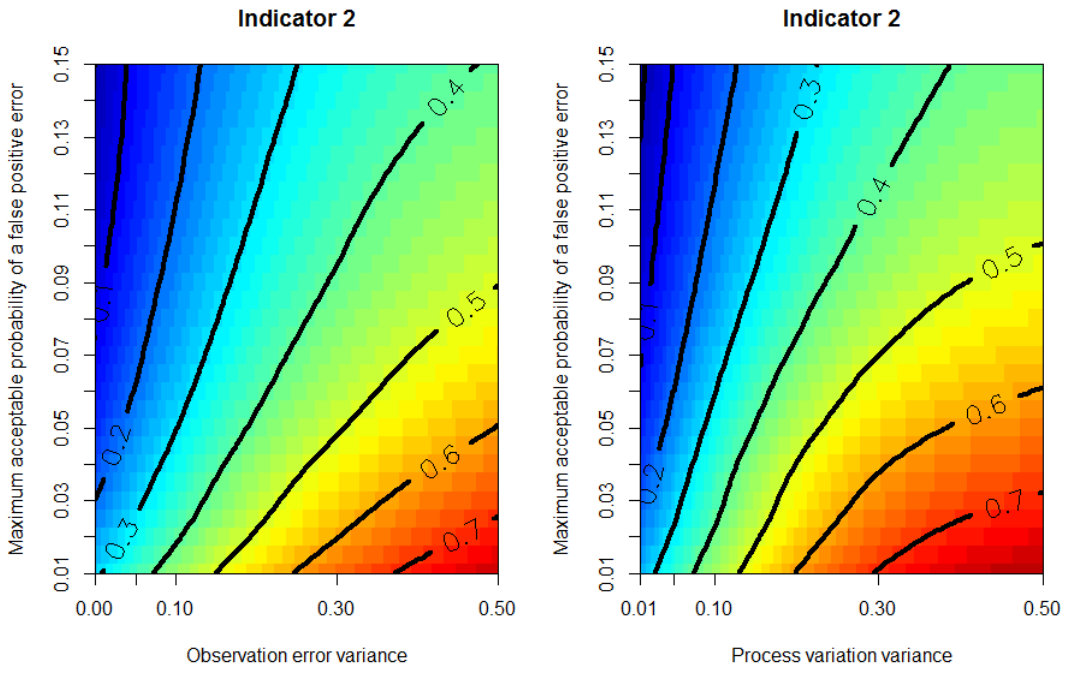


Figure S6


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
0.00
0.10
0.30
0.50
Observation error variance
Indicator 2
Maximum acceptable probability of a false positive error
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.01
0.10
0.30
0.50
Process variation variance

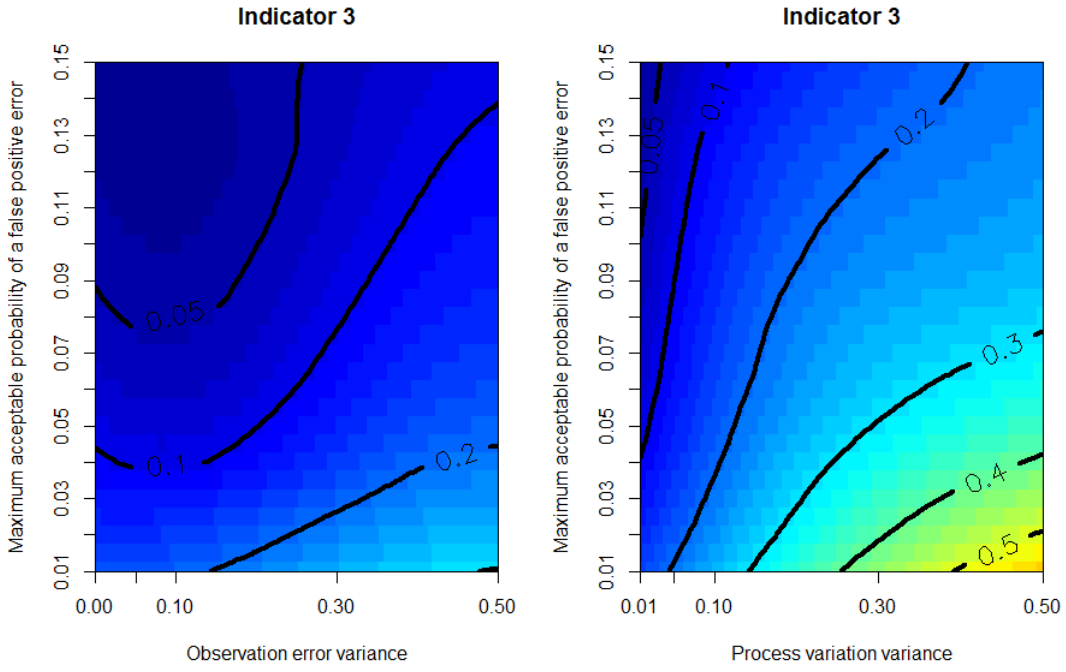
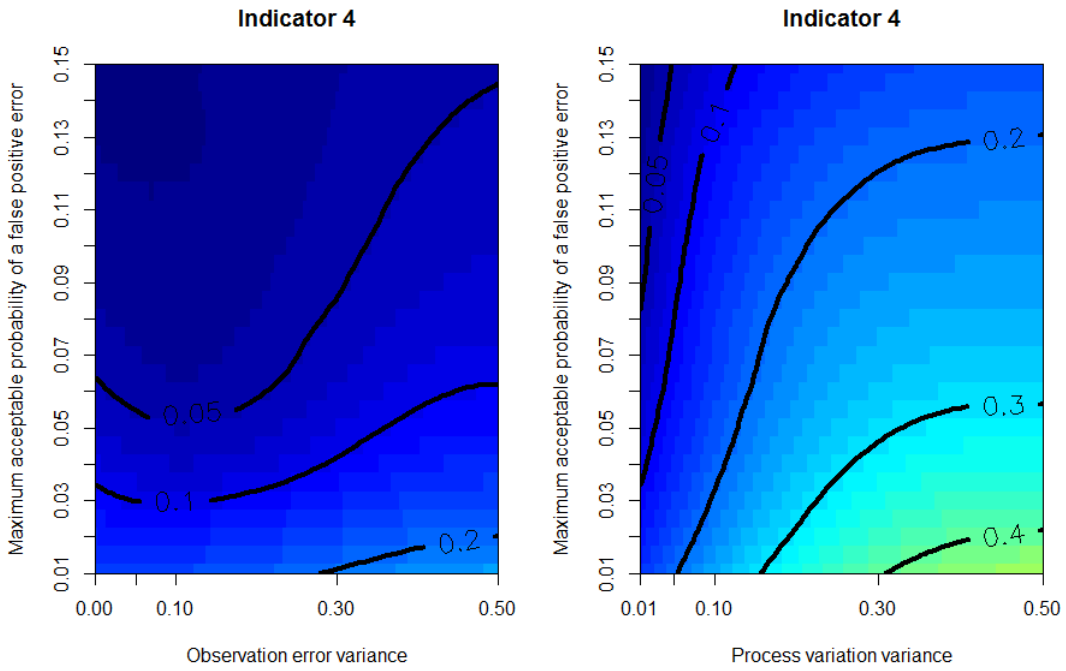
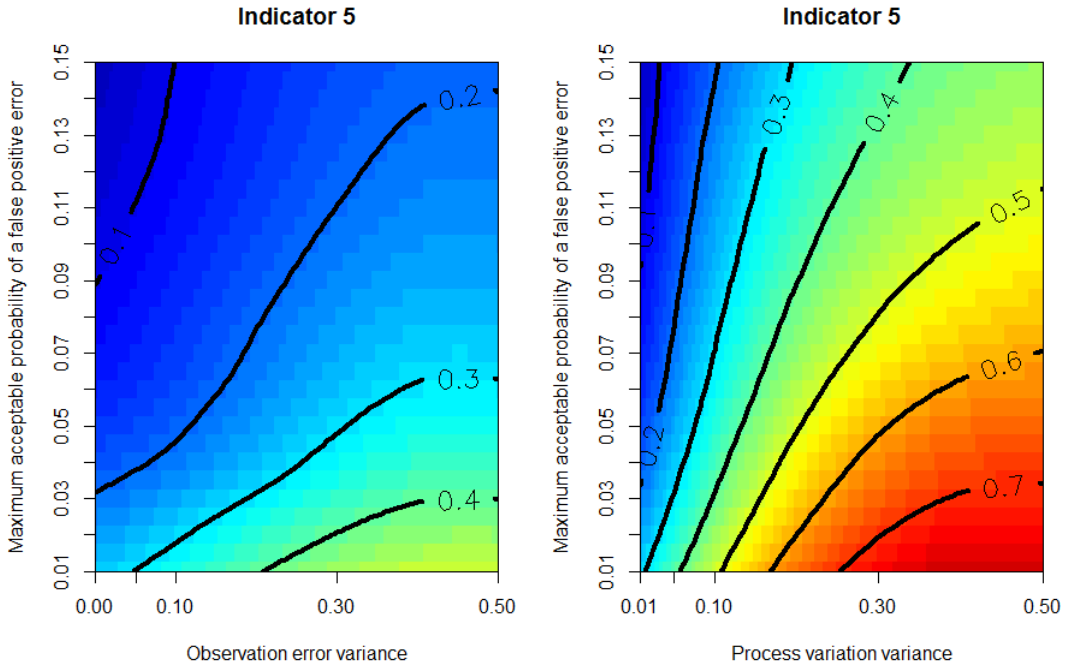
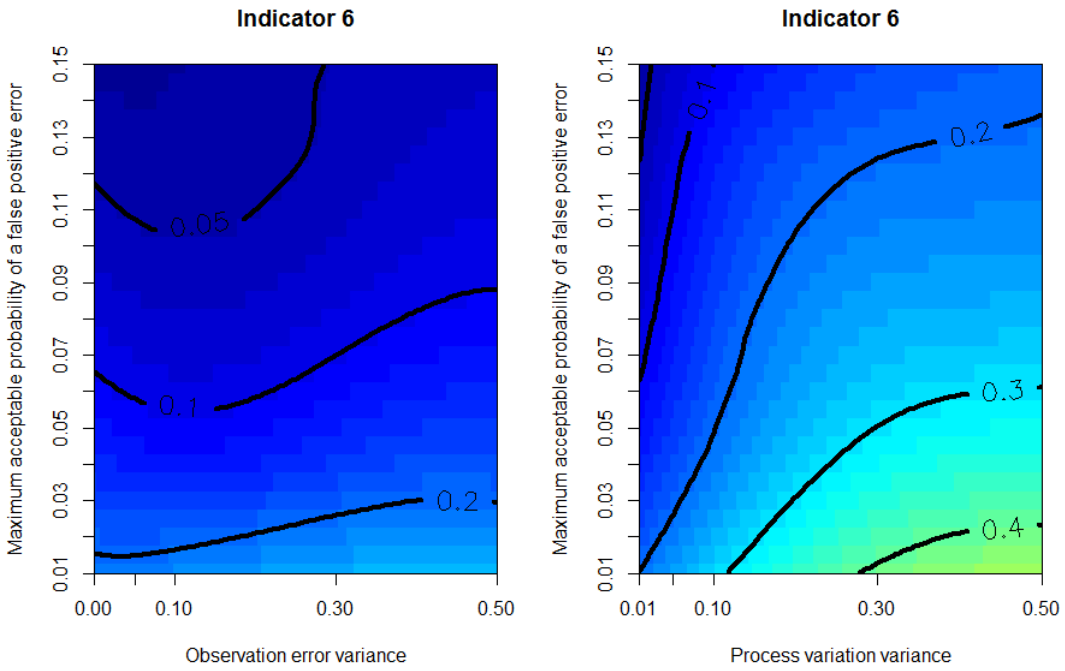

Indicator 3
Maximum acceptable probability of a false positive error
0.01
0.03
0.05
0.07
0.09
0.11
0.13
0.15
0.05
0.1
0.2
0.00
0.10
0.30
0.50
Observation error variance
Indicator 3
Maximum acceptable probability of a false positive error
0.05
0.1
0.2
0.3
0.4
0.5
0.01
0.10
0.30
0.50
Process variation variance

Figure S6

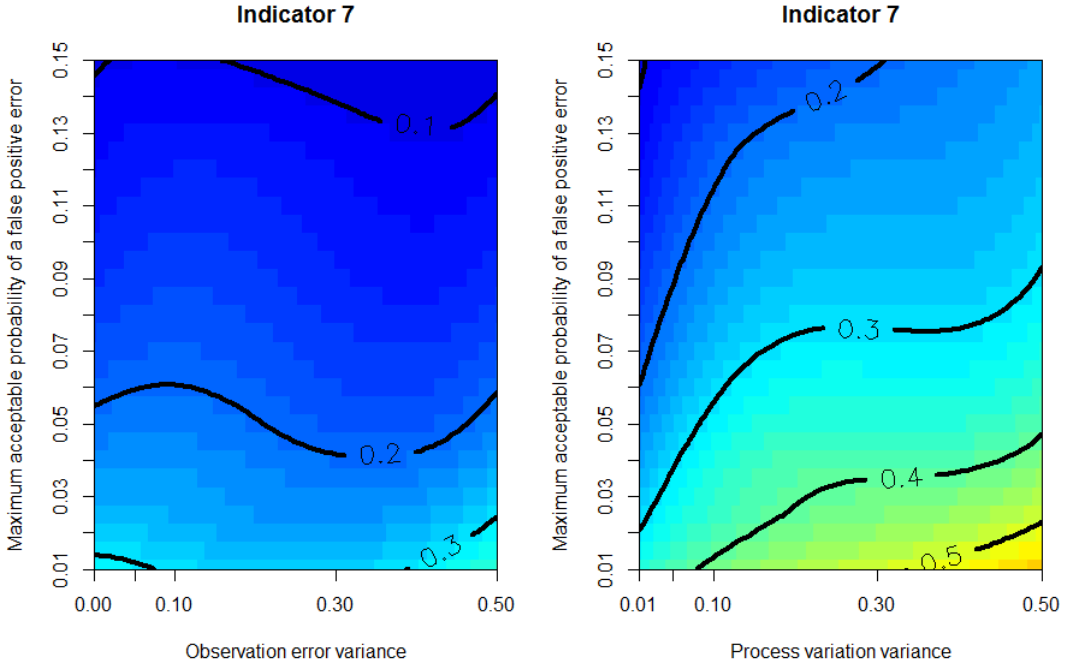

Indicator 4
Maximum acceptable probability of a false positive error
0.01
0.03
0.05
0.07
0.09
0.11
0.13
0.15
0.05
0.1
0.2
0.00
0.10
0.30
0.50
Observation error variance
Indicator 4
0.05
0.1
0.2
0.3
0.4
0.01
0.10
0.30
0.50
Process variation variance


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
0.00
0.10
0.30
0.50
Observation error variance
Indicator 5
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.01
0.10
0.30
0.50
Process variation variance

357 Figure S6



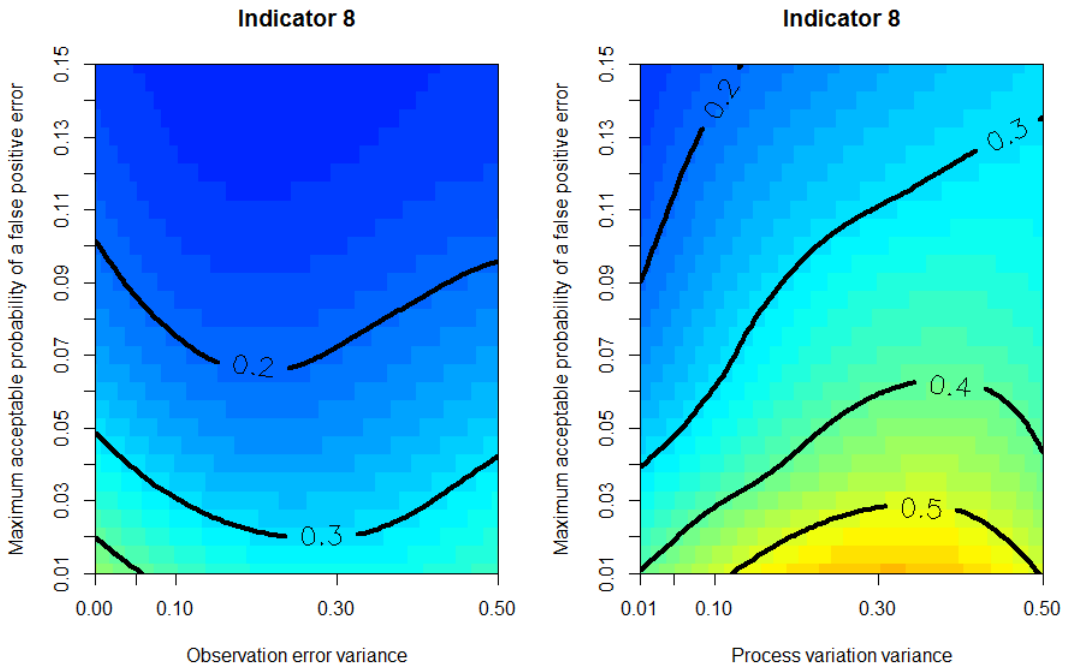
358



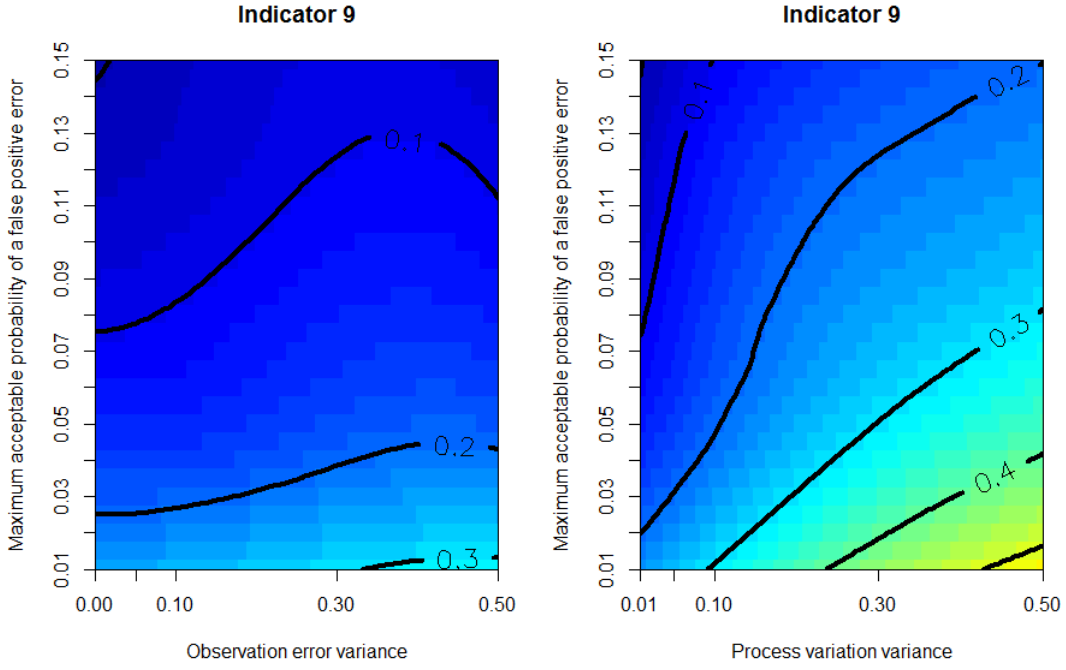
359

360

361 Figure S6



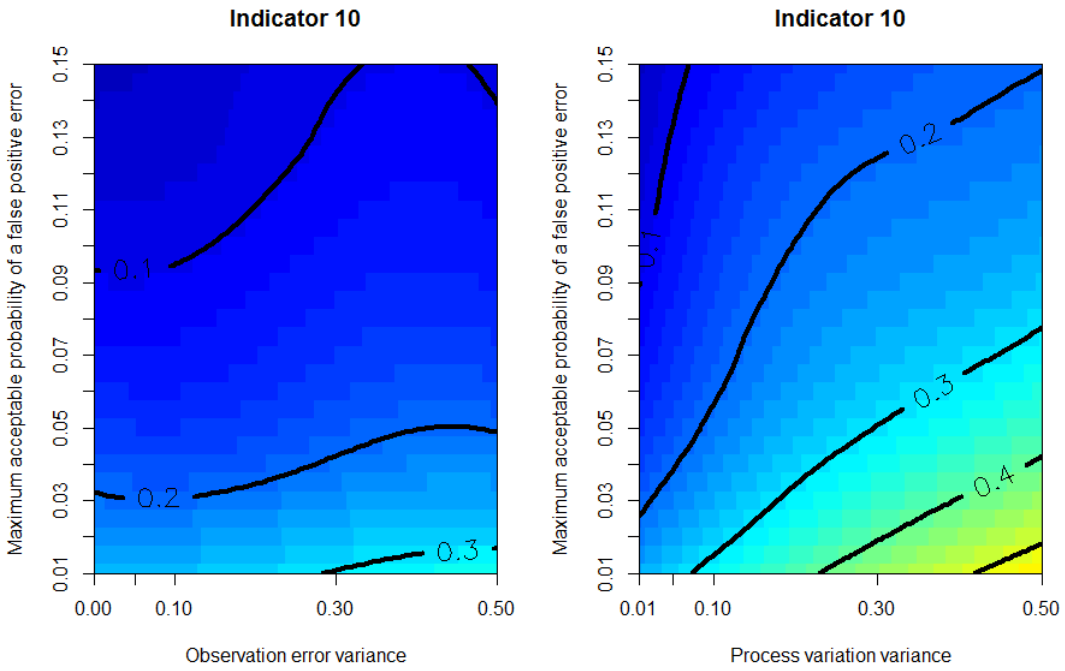
362

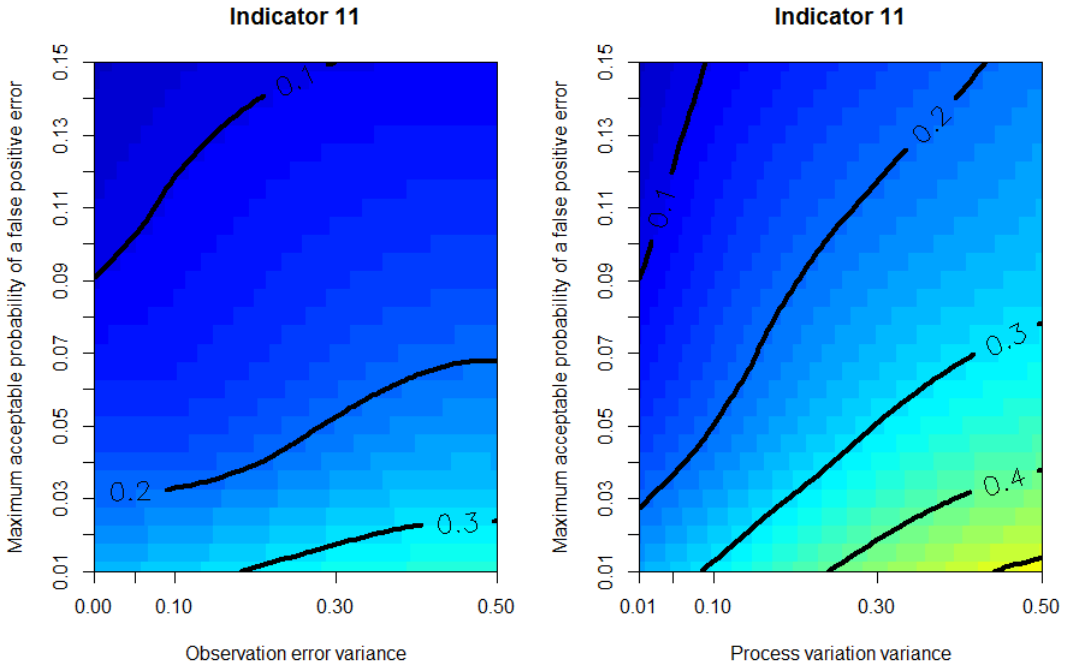


363

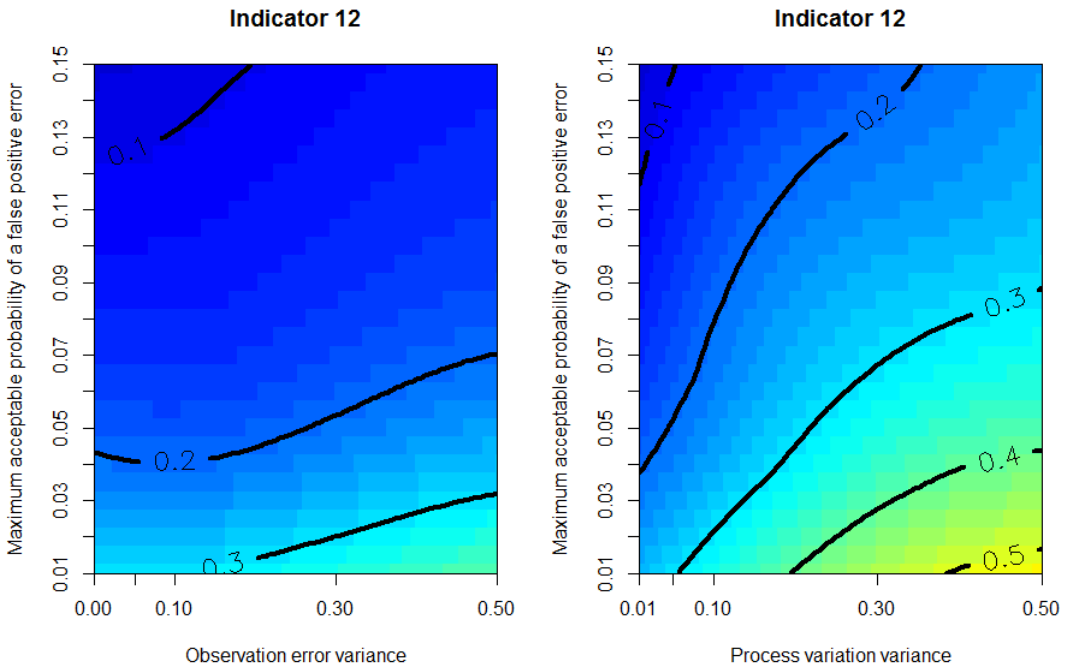
364

Figure S6

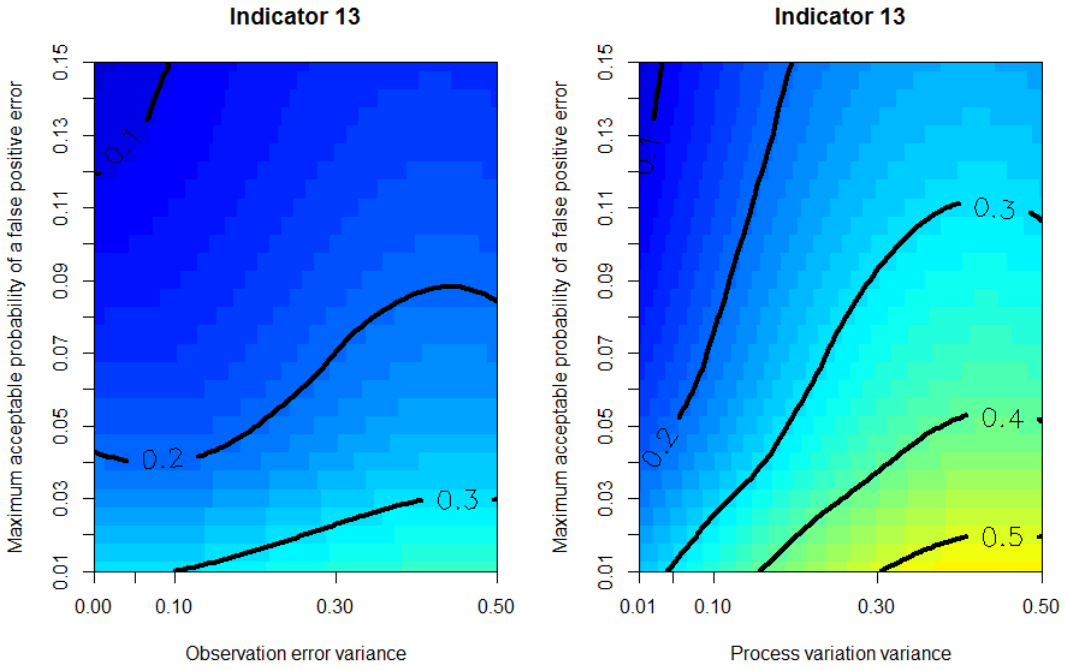

Indicator 10
Maximum acceptable probability of a false positive error
0.01
0.03
0.05
0.07
0.09
0.11
0.13
0.15
0.1
0.2
0.3
0.00
0.10
0.30
0.50
Observation error variance
Indicator 10
Maximum acceptable probability of a false positive error
0.1
0.2
0.3
0.4
0.01
0.10
0.30
0.50
Process variation variance


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
0.00
0.10
0.30
0.50
Observation error variance
Indicator 11
Maximum acceptable probability of a false positive error
0.1
0.2
0.3
0.4
0.01
0.10
0.30
0.50
Process variation variance

369 Figure S6



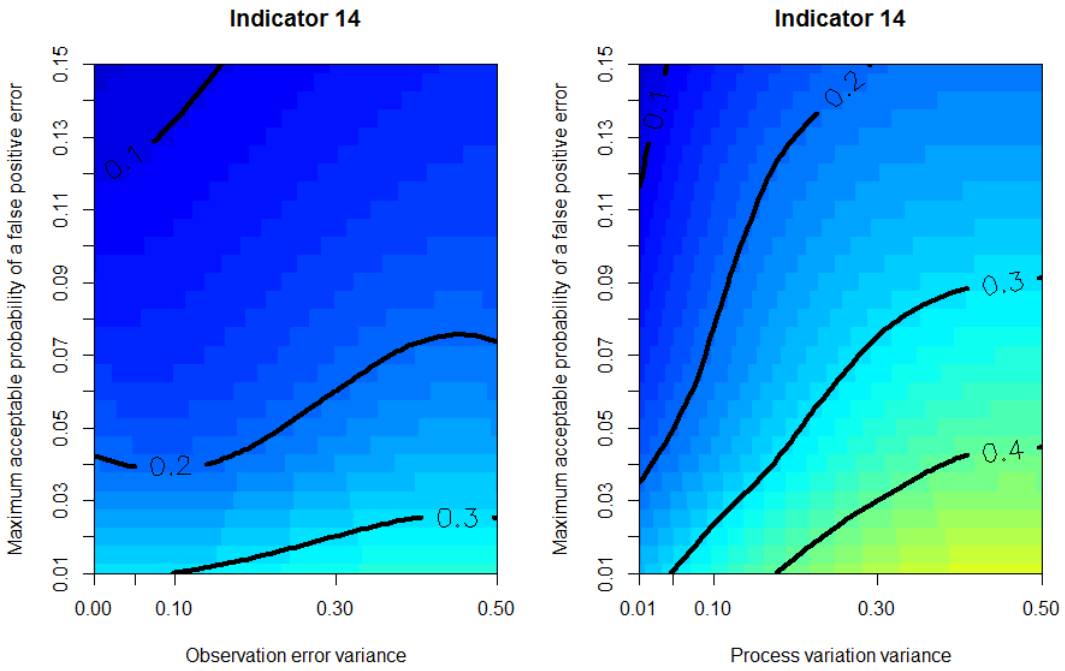
370

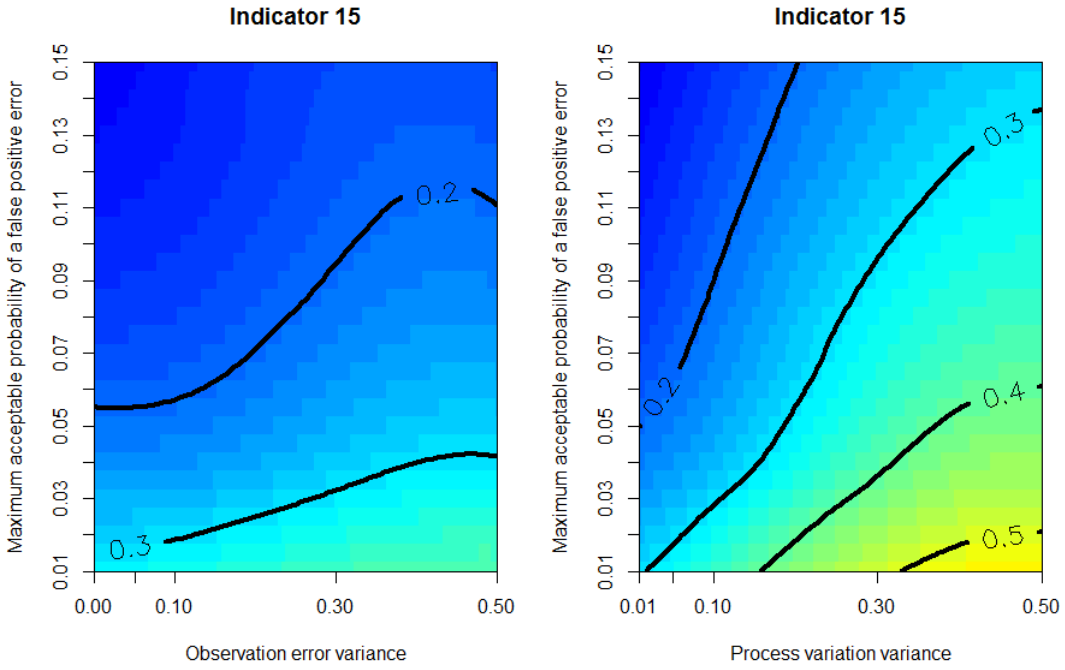


371

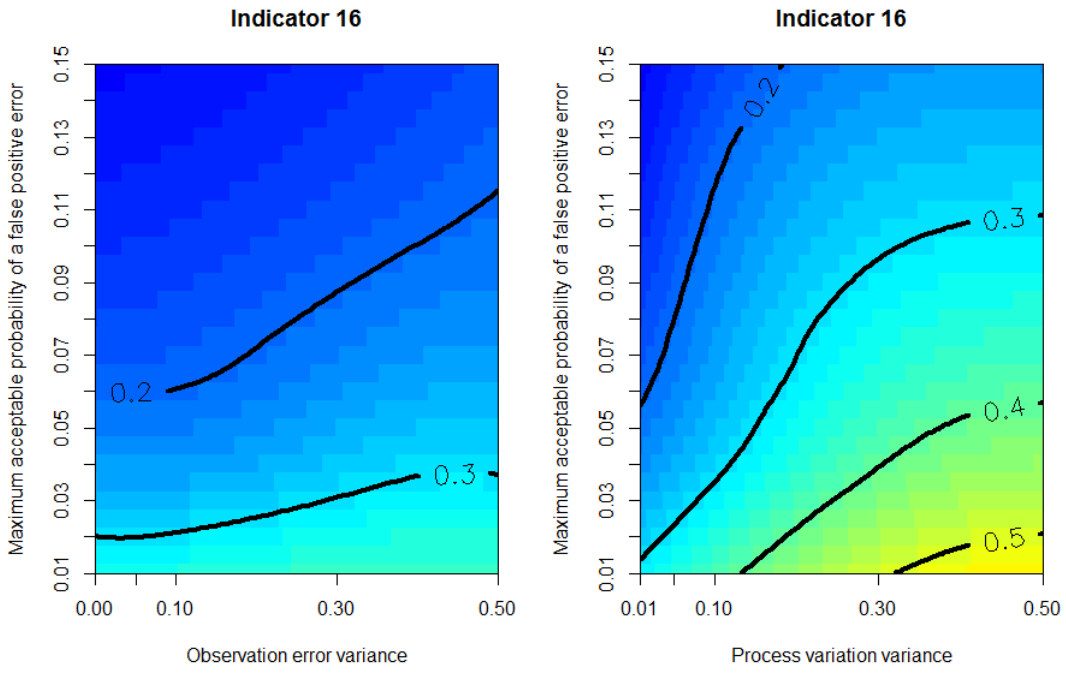
372

Figure S6

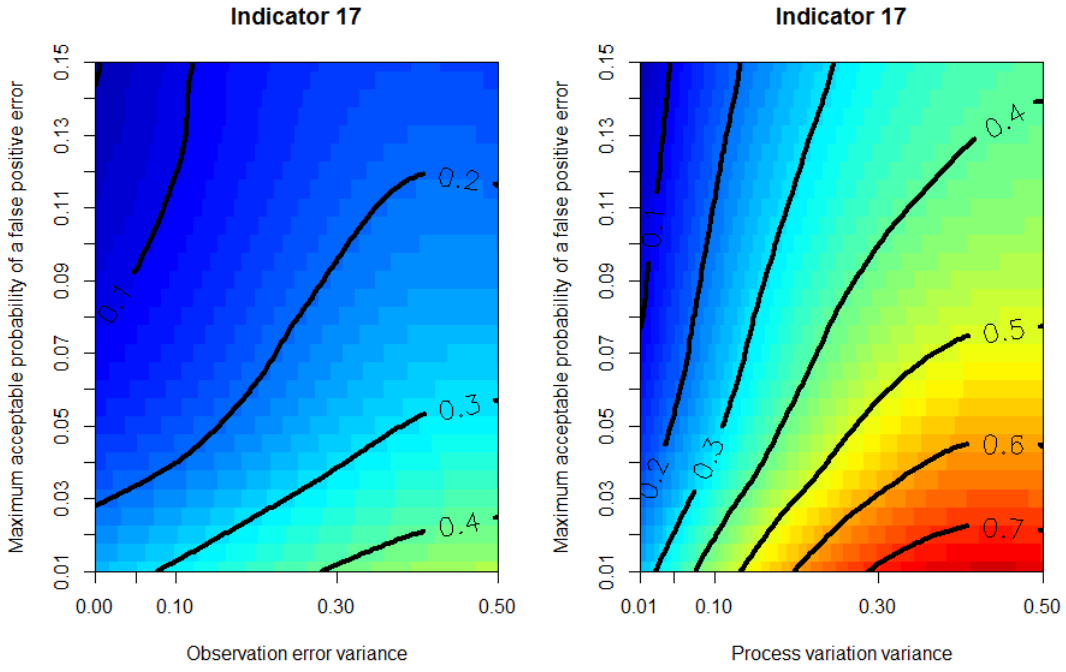

Indicator 14
Maximum acceptable probability of a false positive error
0.01
0.03
0.05
0.07
0.09
0.11
0.13
0.15
0.1
0.2
0.3
0.00
0.10
0.30
0.50
Observation error variance
Indicator 14
Maximum acceptable probability of a false positive error
0.1
0.2
0.3
0.4
0.01
0.10
0.30
0.50
Process variation variance


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
0.00
0.10
0.30
0.50
Observation error variance
Indicator 15
Maximum acceptable probability of a false positive error
0.2
0.3
0.4
0.5
0.01
0.10
0.30
0.50
Process variation variance

377 Figure S6



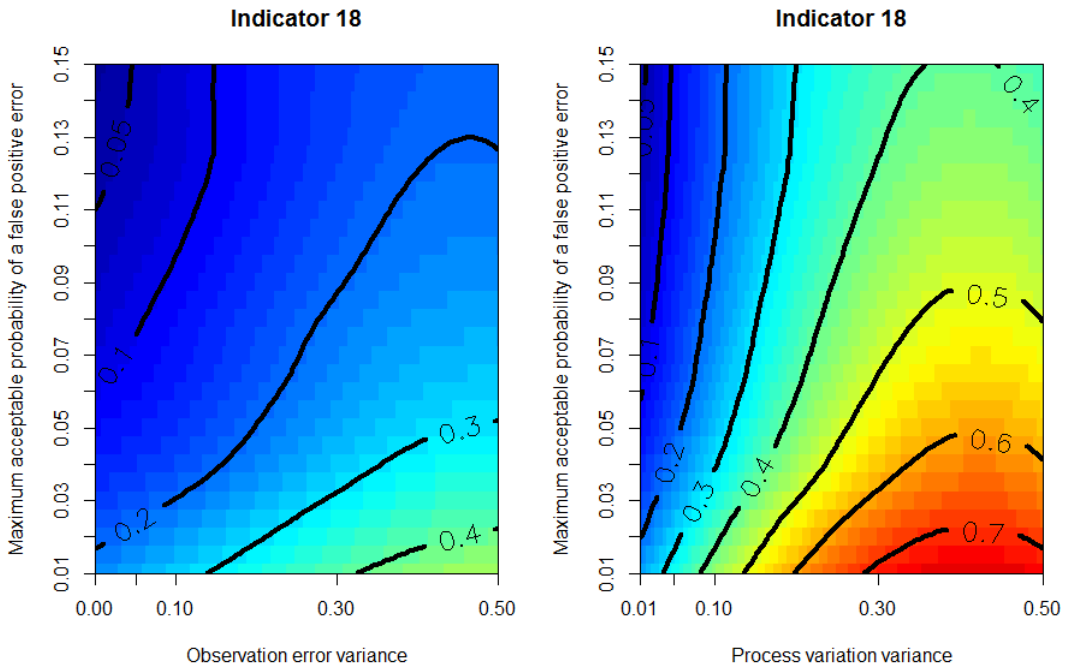
378



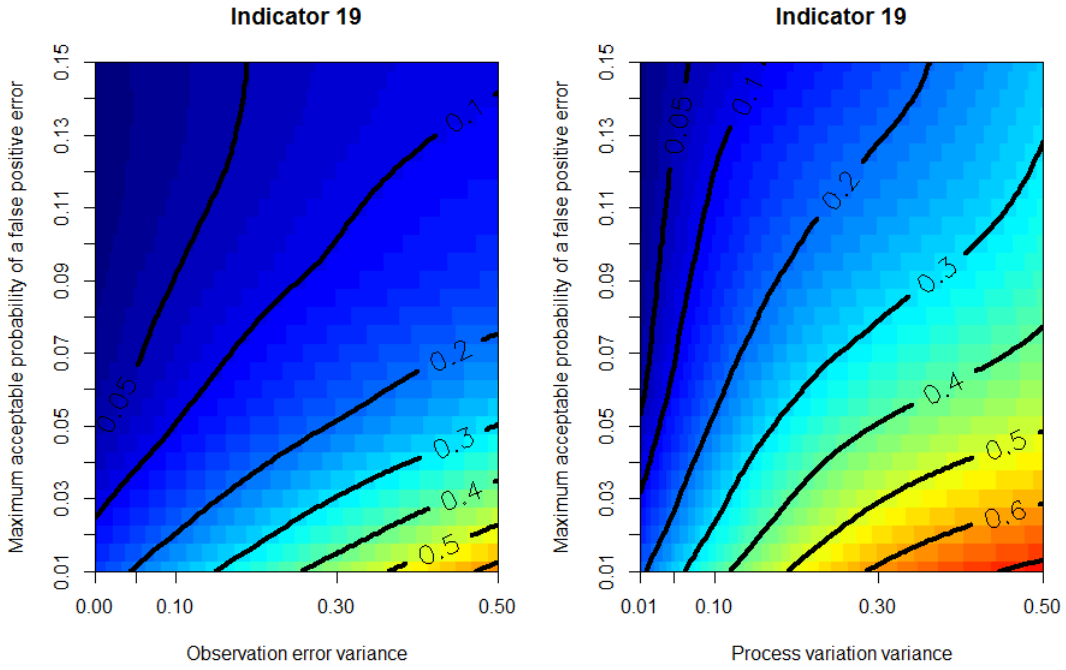
379

380

381 Figure S6



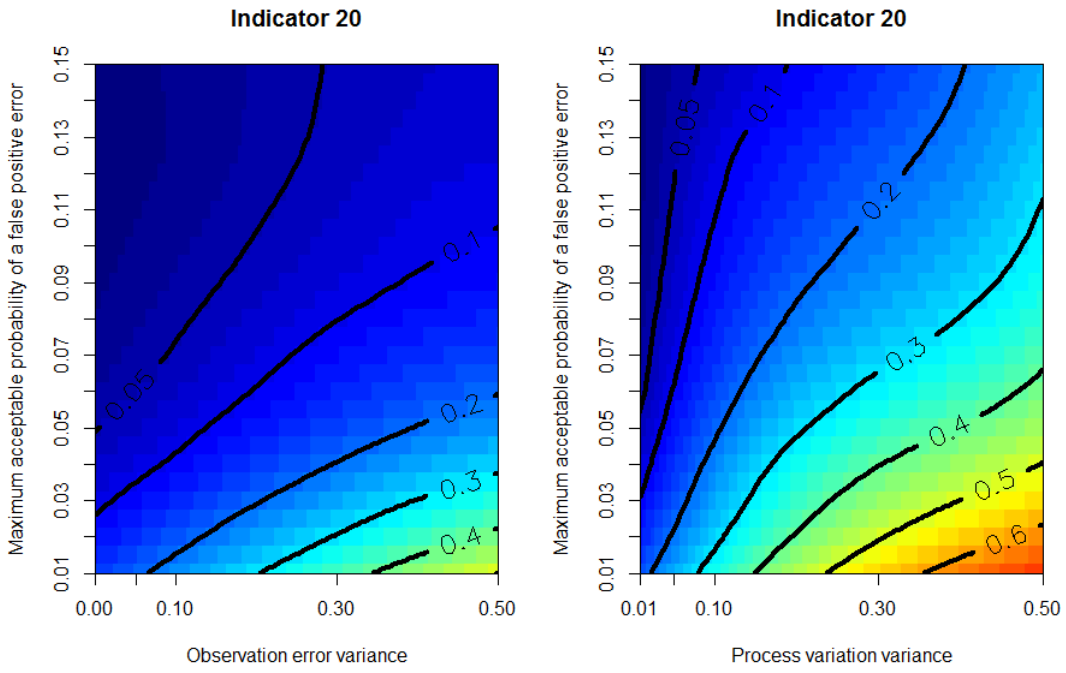
382



383

384

385 Figure S6



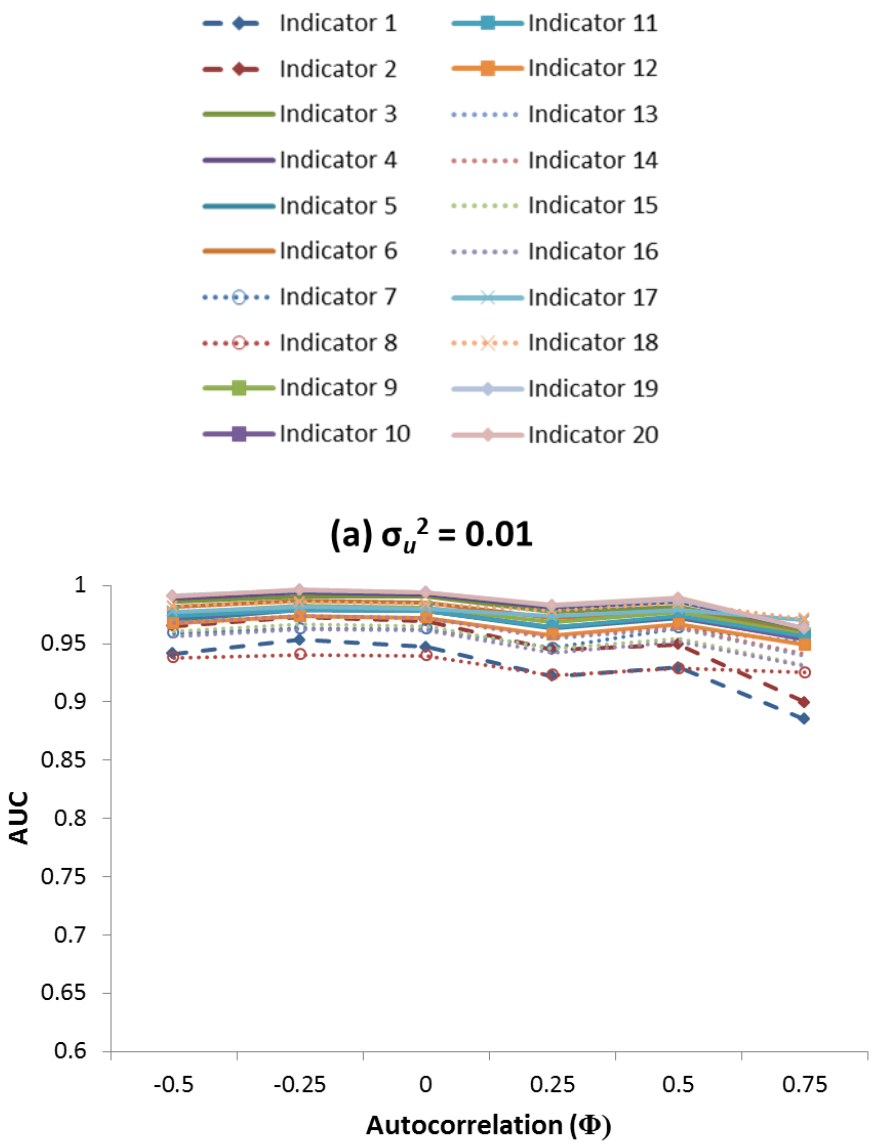
386

387

388 **Figure S7.** Changes in AUC as a function of the magnitude of temporal autocorrelation (Φ) in
389 process variation at different levels of process variation (σ_u^2 = 0.01, 0.05, 0.1, 0.3, 0.5 in panels
390 a-e, respectively) with no observation error (σ_v^2=0) for all indicators. Higher AUC values reflect
391 a better ability of an indicator to differentiate between a declining and non-declining population.
392 Table 2 and Supporting Information (SI) part B define the indicators. Thick broken dashed lines
393 with solid symbols are for indicators that are based on decline in the last 3 generations
394 (indicators 1 and 2). Solid lines are for indicators that are based on decline from some historical
395 baseline (indicators 3-6, 9-12, 17, and 19-20). Thin dotted lines are for indicators based on
396 decline from a maximum abundance (indicators 7-8, 13-16, and 18).
397

398 Figure S7

399



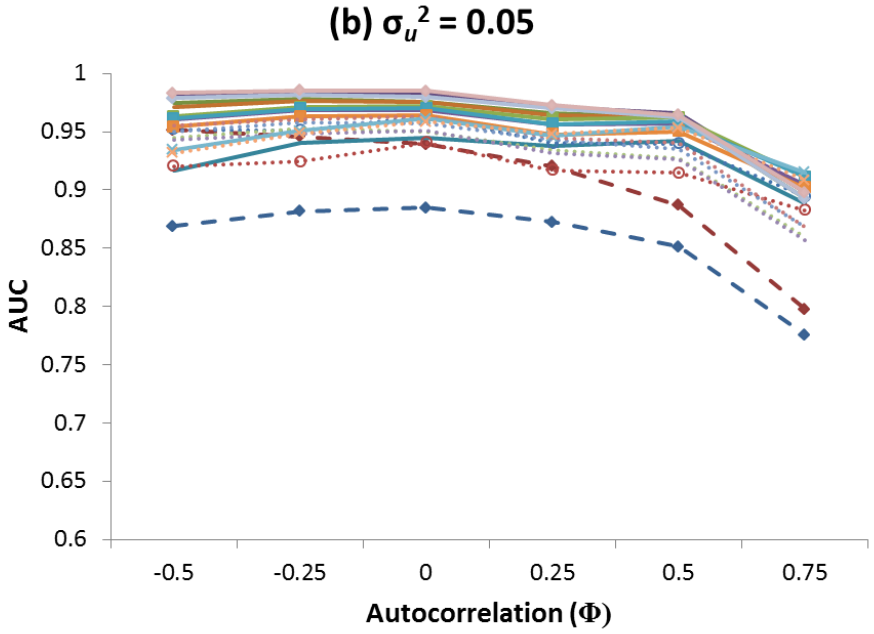
400

401

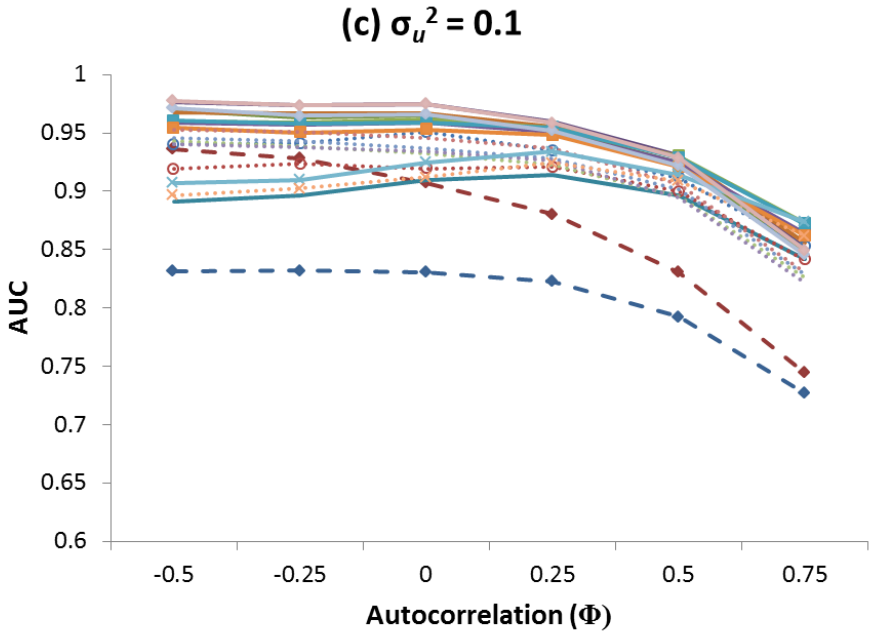
402

403 Figure S7

404



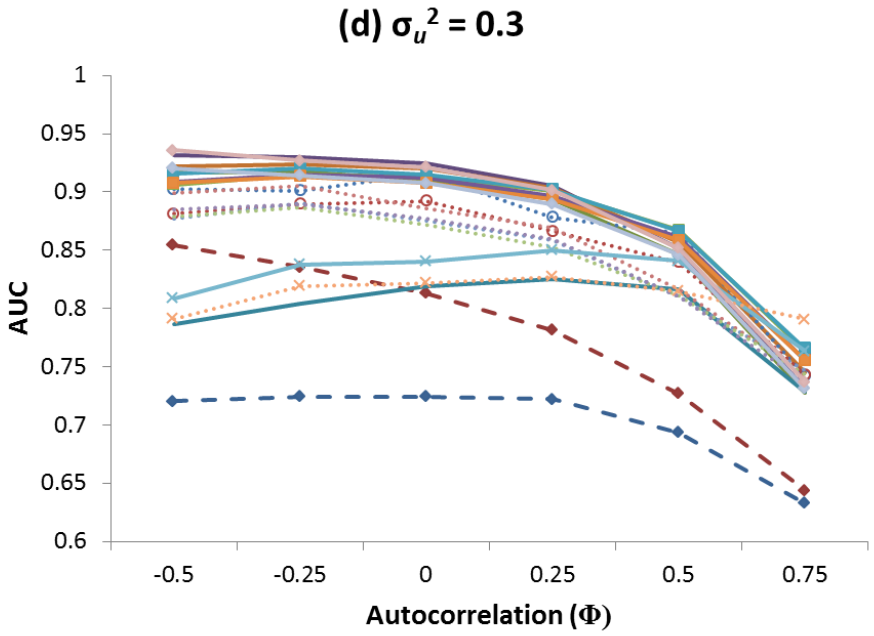
405

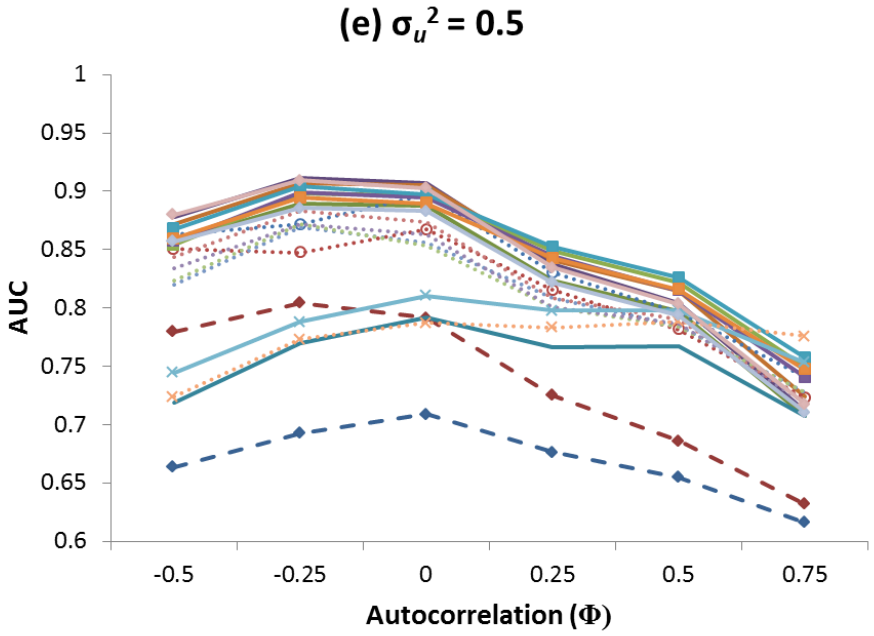


406

407

Figure S7

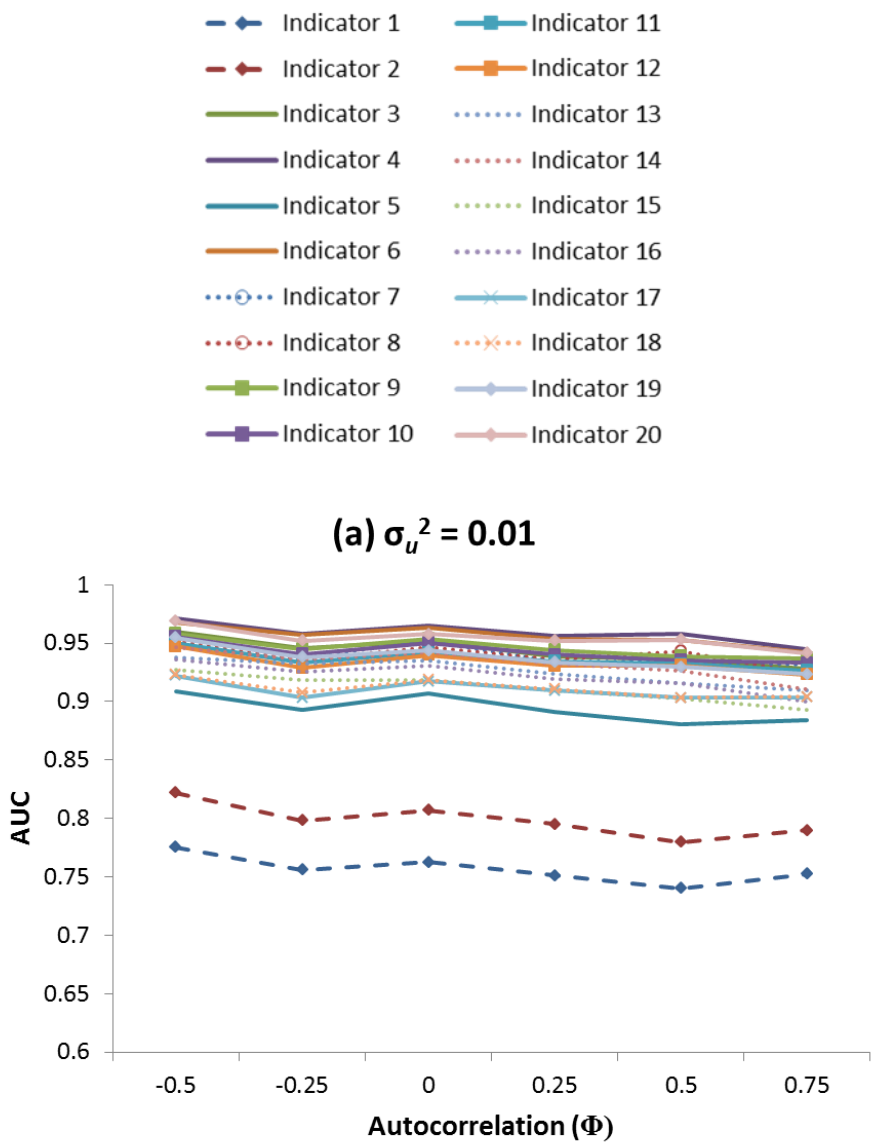

(d) $\sigma_u^2 = 0.3$
AUC
1
0.95
0.9
0.85
0.8
0.75
0.7
0.65
0.6
-0.5
-0.25
0
0.25
0.5
0.75
Autocorrelation (Φ)


(e) $\sigma_u^2 = 0.5$
AUC
1
0.95
0.9
0.85
0.8
0.75
0.7
0.65
0.6
-0.5
-0.25
0
0.25
0.5
0.75
Autocorrelation (Φ)

414 **Figure S8.** Same as Figure S7, except these results are from simulations with high observation
415 error ($\sigma_v^2 = 0.5$) for all indicators, instead of no observation error. Higher AUC values reflect a
416 better ability of an indicator to differentiate between a declining and non-declining population.
417 Table 2 and Supporting Information (SI) part B define the indicators. Thick broken dashed lines
418 with solid symbols are for indicators that are based on decline in the last 3 generations
419 (indicators 1 and 2). Solid lines are for indicators that are based on decline from some historical
420 baseline (indicators 3-6, 9-12, 17, and 19-20). Thin dotted lines are for indicators based on
421 decline from a maximum abundance (indicators 7-8, 13-16, and 18).
422

423 Figure S8

424

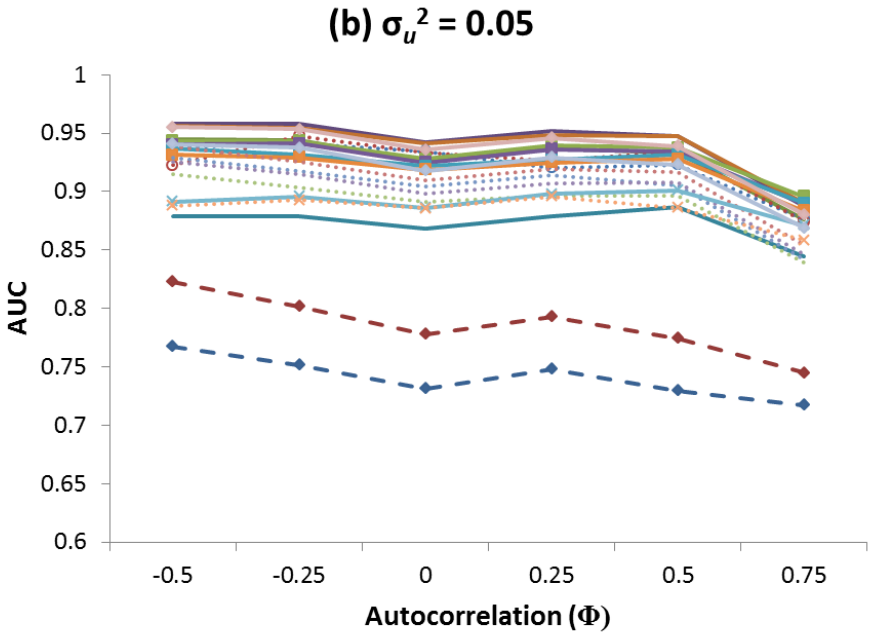


425

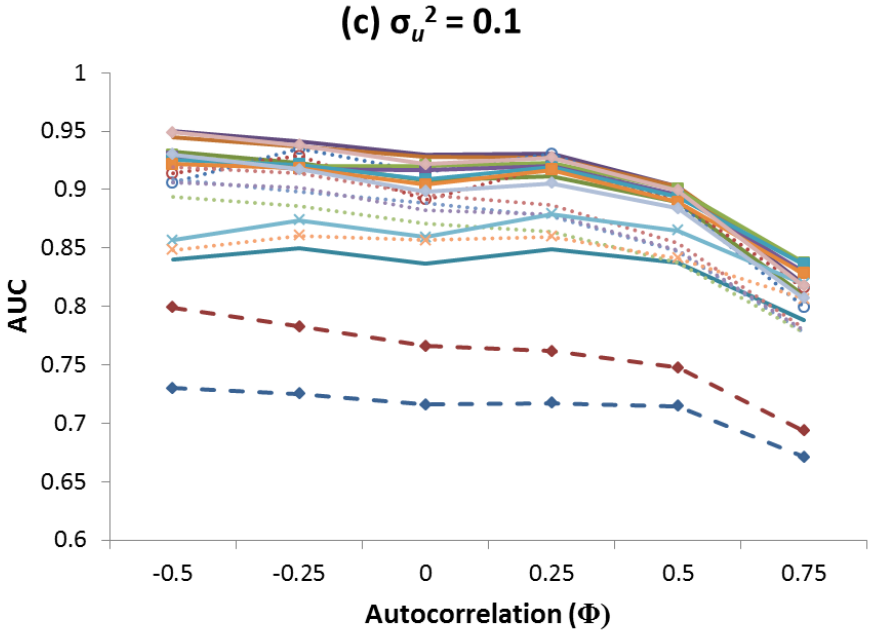
426

427

428 Figure S8



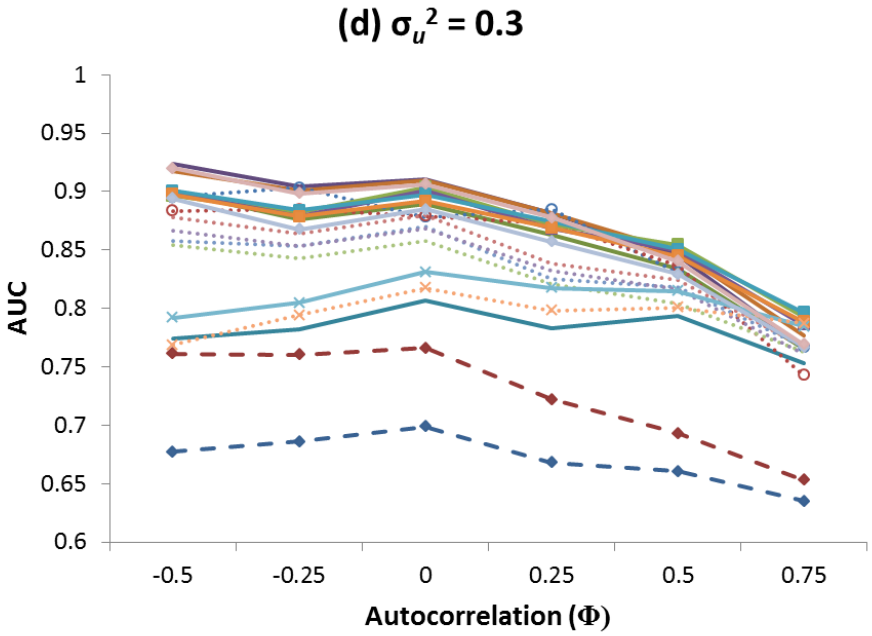
429

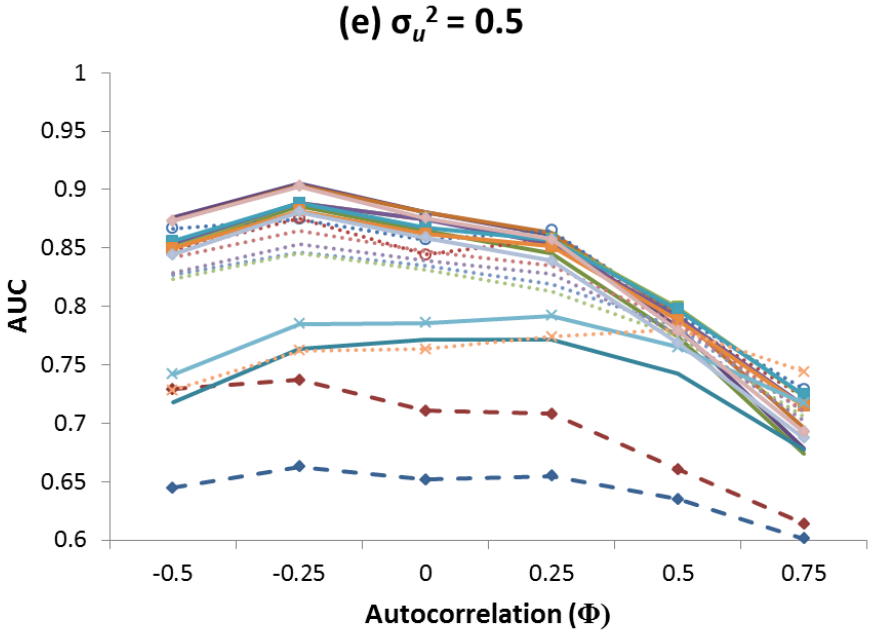


430

431

Figure S8


(d) $\sigma_u^2 = 0.3$
AUC
1
0.95
0.9
0.85
0.8
0.75
0.7
0.65
0.6
-0.5
-0.25
0
0.25
0.5
0.75
Autocorrelation (Φ)


(e) $\sigma_u^2 = 0.5$
AUC
1
0.95
0.9
0.85
0.8
0.75
0.7
0.65
0.6
-0.5
-0.25
0
0.25
0.5
0.75
Autocorrelation (Φ)

F. References for Supporting Information

Baxter, P.W.J. & Possingham, H.P. (2011). Optimizing search strategies for invasive pests: learn before you leap. *Journal of Applied Ecology*, **48**, 86-95.

Connors, B.M., Cooper, A.B., Peterman, R.M., & Dulvy, N.K. (2014). The false classification of extinction risk in noisy environments. *Proceedings of the Royal Society B*, **281**, 20132935. DOI:10.1098/rspb.2013.2935.

Cooke, S.J., Hinch, S.G., Farrell, A.P., Lapointe, M.F., Jones, S.M.R., Macdonald, J.S., Patterson, D.A., Healey, M.C. & Van Der Kraak, G. (2004). Abnormal migration timing and high en-route mortality of sockeye salmon in the Fraser River, British Columbia. *Fisheries*, **29**, 22–33.

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).

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

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

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

460 Korman, J., Peterman, R.M. & Walters, C.J. (1995). Empirical and theoretical analyses of
461 correction of time-series bias in stock–recruitment relationships of sockeye salmon. *Can.*
462 *J. Fish. Aquat. Sci*., **52**, 2174–2189.
463 Paulsen, C.M., Hinrichsen, R.A. & Fisher, T.R. (2007). Measure twice, estimate once: Pacific
464 salmon population viability analysis for highly variable populations. *Transactions of the*
465 *American Fisheries Society*, **136**, 346-364.
466 Pearce, J. & Ferrier, S. (2000). Evaluating the predictive performance of habitat models
467 developed using logistic regression. *Ecological Modelling*, **133**, 225-245.
468 Peterman, R.M., Dorner, B. (2012). A widespread decrease in productivity of sockeye salmon
469 populations in western North America. *Canadian Journal of Fisheries and Aquatic*
470 *Sciences* **69**, 1255–1260.
471 Peterman, R.M., Pyper, B.J., Mueter, F.J., Haeseker, S.L., Su, Z. & Dorner, B. (2009). Statistical
472 Models of Pacific Salmon that Include Environmental Variables. *American Fisheries*
473 *Society Symposium*, **71**, 125–146.
474 Porszt, E.J., Peterman, R.M., Dulvy, N.K., Cooper, A.B. & Irvine, J.R. (2012). Reliability of
475 indicators of decline in abundance. *Conservation Biology*, **26**, 894-904.
476 Rand, P.S. (2011). *Oncorhynchus nerka*. IUCN Red List of Threatened Species. Version 2011.2.
477 International Union for Conservation of Nature, Gland, Switzerland. Available from
478 http://www.iucnredlist.org/apps/redlist/details/135301/0 (accessed January 2012).
479 Ricker, W.E. (1997). Cycles of abundance among Fraser River sockeye salmon (*Oncorhynchus*
480 *nerka*). *Canadian Journal of Fisheries and Aquatic Sciences*, **54**, 950-968.
481 Venables, W.N. & Ripley, B.D. (2002). Modern applied statistics with S. 4th edition. Springer
482 Science, New York.

Walters, C. J. & Ludwig, D. (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., Kendall, B.E. & Possingham, H.P. (2011). Variability in population abundance and the classification of extinction risk. *Conservation Biology*, **25**, 747-757.