Further development of the 'Acceptable Level of Impact' framework for effects of offshore wind farms on seabirds

Extended method development and sensitivity analysis

Author(s): Vincent Hin, Tobias van Kooten, Astrid Potiek & Jente Kraal

Wageningen University & Research report: C088/23

Further development of the 'Acceptable Level of Impact' framework for effects of offshore wind farms on seabirds

Extended method development and sensitivity analysis

Authors: Vincent Hin[1], Tobias van Kooten[1], Astrid Potiek[2] & Jente Kraal[2]

[1]Wageningen Marine Research

[2]Waardenburg Ecology

Wageningen Marine Research
IJmuiden, December 2023

Wageningen Marine Research report C088/23

Keywords: offshore wind farms, acceptable population effects, seabirds, sensitivity analysis

Client: Rijkswaterstaat Zee en Delta
Attn.: Martine Graafland
Lange Kleiweg 34
2288 GK, Rijswijk

This report can be downloaded for free from https://doi.org/10.18174/644082
Wageningen Marine Research provides no printed copies of reports

Wageningen Marine Research is ISO 9001:2015 certified.

KvK nr. 09098104,
WMR BTW nr. NL 8113.83.696.B16.
Code BIC/SWIFT address: RABONL2U
IBAN code: NL 73 RABO 0373599285

A_4_3_2 V32 (2021)

Contents

Summary **4**

1 Assignment **5**

1.1 Background 5
1.2 ALI methodology 5
1.3 Scope & outline 6

2 Materials and Methods **8**

2.1 Northern gannet population model 8
2.2 Sensitivity analysis 9
2.2.1 Parameter sampling method 9
2.2.2 Parameter standard deviation 10
2.2.3 X threshold and ALI calculation 10
2.2.4 Variation of casualties through time 11

3 Results **13**

3.1 Derivation of P_C and Y 13
3.1.1 Background 13
3.1.2 Process and derivation of probabilities 13
3.2 Results sensitivity analysis 16
3.2.1 Parameter sampling method 16
3.2.2 Effect of parameter sampling method on ALI parameters 18
3.2.3 Effect of standard deviation 19
3.2.4 Effect of X threshold value 20
3.2.5 Variation of causalities through time 22

4 Conclusions and recommendations **23**

4.1 Method and definition P_C 23
4.2 Sensitivity analysis 23
4.2.1 Parameter sampling method 23
4.2.2 Standard deviation of parameter estimates 24
4.2.3 X' threshold level 24
4.2.4 Time varying mortality sequences 24
4.3 The effect of variation and uncertainty within the ALI framework 24
4.4 Recommendations 25

5 Quality Assurance **26**

References **27**

Justification **28**

Summary

The Acceptable Level of Impact (ALI) methodology defines acceptable limits for the predicted population effects of mortality imposed by offshore wind farms (OWFs) for marine birds. These population effects are quantified using stochastic population models tailored for specific seabird species. The ALI is defined as: '*The probability of a population decline of X% or more, 30 years after the impact, cannot exceed Y'.* In this definition, the X threshold determines the reduction in population abundance due to OWFs that is considered acceptable, evaluated 30 years after the onset of the impact. The Y threshold of the ALI is derived from a threshold value for the 'causality probability' (P_C). P_C is the probability that a violation of X is caused by the impact from OWFs, instead of being caused by variability or uncertainty inherent to predictions of future population abundance.

The ALI methodology was reviewed in November 2021 – January 2022. Reviewers concluded that the approach and assumptions of the proposed ALI methodology were a considerable improvement in relation to existing impact-evaluation frameworks (ORNIS, PBR). However, reviewers also raised concerns about the legal tenability and definition of the causality level P_C.

Several points put forward in the reviews were addressed within the current report, namely a derivation of the causality threshold P_C and results of a sensitivity analysis of the ALI methodology, which contained several components. First, an analysis of the effect of uncertainty on the outcome of the ALI was performed. This analysis was done by: 1) changing the way in which parameters of the population model were sampled to simulate year-to-year variation in demographic rates and, 2) changing the standard deviation of the parameters of the population model. Second, the effect of changing the X threshold value on the outcome of the ALI was evaluated. Lastly, the effect of time-varying, as opposed to constant, mortality levels on the outcome of the ALI was explored.

We also address concerns about the definition of P_C. We derive P_C from conditional probabilities. This derivation shows that P_C in itself is correctly defined, and is consistently drawn from other commonly used risk measures, such as the 'attributable fraction among the exposed' and the 'relative risk ratio.'

The sensitivity analysis reveals that sampling population parameters once at the start of each simulation (initial sampling) leads to considerable more variation in predicted population abundance than sampling population parameters each year (annual sampling). More variation, resulting in more uncertainty of the population projection leads to an increased probability of an X threshold violation in the scenario without OWF impact. In addition, a more strict X threshold also results in an increased probability of an X threshold violation in the scenario without OWF impact. The sensitivity analysis reveals that in the current framework, with the Y threshold being dependent on the causality threshold P_T and the probability of violation of the X threshold in the unimpacted scenario, a more strict X threshold leads to a less strict Y threshold, which is considered undesirable. This relationship stems from the use of the causality measure P_C, which attempts to correct for false positive outcomes (threshold violation without impact) that are caused by variation (environmental stochasticity) or uncertainty. In principal, the relationship between X and Y can be accounted for in the choice of X and the threshold value for the causal probability, from which Y is derived. However, this would further complicate the methodology and create a potential problem in the application of the framework, because correctly choosing X and Y values would require a thorough methodological understanding of the framework. The sensitivity analysis also showed that the outcome of the ALI was not affected when using time-varying, as opposed to constant, mortality levels.

An essential property of any framework is that it is fit for purpose. This includes that it can be applied without risk of accidental misuse by the end users. This is not the case for the current framework, and we therefore strongly advise to further develop the current ALI methodology. We recommend to revise the methodology to increase its simplicity and avoid use of the causality measure P_C, which is at the root of the current problem.

1 Assignment

1.1 Background

The Acceptable Level of Impact (ALI) methodology (Potiek et al., 2022) defines acceptable limits for the predicted population effects of mortality imposed by offshore wind farms (OWFs) for marine birds. The methodology was developed in 2021 as a replacement for the Potential Biological Removal (PBR) method. Unlike the PBR method, the ALI methodology makes a clear distinction between the uncertainty associated with the predicted population effect of OWFs, and the societal and legal considerations about which effects are acceptable. The uncertainty stems from the ecology and available knowledge on the species in question and lies in the domain of researchers, while the question of which population effects are admissible is considered by management and policy-makers. This distinction ensures that researchers are exempted from making societal evaluations, while policy-makers can focus on such considerations without requiring specialist ecological knowledge of the subject species.

The ALI methodology was first used for the calculations of the "*Aanvullend Ontwerp bij het Programma Noordzee 2022-2027*" (Potiek et al., 2021; Rijksoverheid, 2021; Soudijn et al., 2022). After this report, the methodology was reviewed in November 2021 – January 2022. The general conclusion from the reviews was that the approach and assumptions of the proposed ALI methodology were a considerable improvement in relation to existing impact-evaluation frameworks (ORNIS, PBR). However, the reviews also raised concerns about the legal tenability and definition of the causality level P_C used within the ALI framework, among some others points. The issues raised in the reviews, together with feedback from government officials were addressed within the project "*Doorontwikkeling ALI methodologie*". The project further develops the ALI framework and also includes a sensitivity analysis, which was not performed for the original report (Potiek et al., 2021) due to time constraints. In the next sections we summarize the ALI methodology as proposed by Potiek et al. (2022) and introduce a more detailed scope and outline for further development of the methodology.

1.2 ALI methodology

The ALI methodology uses population models to compare scenarios with mortality impact from OWFs, against scenarios without such impact (unimpacted). The methodology accounts for uncertainty about the future development of populations by means of a Monte Carlo approach. This means running many replicate simulations of the population model, and for each replicate using slightly different parameter values. This leads to a plume of population trajectories per scenario (impacted and unimpacted).

The ALI is currently formulated as: '*The probability of a population decline of X% or more, 30 years after the impact, cannot exceed Y'*. (Potiek et al., 2022). The threshold values X and Y are chosen by policy makers and together determine whether the ALI is violated. A violation of the ALI occurs if the probability that the population declines by more than X%, 30 years after the onset of impact, is larger than Y. The time frame is 30 years is derived from the expected lifetime of OWFs, but this can be easily extended to longer time frames. A possible extension of the ALI to 40 years will be dealt with in the near future.

The X threshold determines the reduction in population abundance due to OWFs that is considered acceptable, evaluated 30 years after the onset of the impact. This reduction is measured as the relative difference between the impacted population abundance after 30 simulated years and the *median* population abundance in the unimpacted scenario after 30 simulated years. For example, an X value of 0.25 means that the impacted population abundance can at maximum be 25% lower than the median unimpacted population abundance. Under both scenarios, some replicate simulations will exceed the X threshold. Under the unimpacted scenario this is solely the effect of uncertainty and under the impacted scenario this is the joined effect of uncertainty and impact. The X threshold value (acceptable decline over the time frame of 30 years) is derived from a policy decision for X', which is the acceptable decline over three generations (Potiek et al., 2022).

The Y threshold value of the ALI is derived from a threshold value for the 'causality level' (P_C). P_C quantifies the probability that a violation is caused by the impact from OWFs, instead of being caused by the variability or uncertainty inherent to the future development of populations. For example, when $P_C = 0.66$, it is twice as likely that a threshold violation is caused by an impact than not (0.66 vs 0.33). Policy makers decide on a threshold value for P_C, which is referred to as P_T. P_T is used subsequently to calculate the threshold value Y. Note that Y is a threshold value for the total probability of violation in the impacted scenario, both due to impact as well as due to uncertainty/variation.

1.3 Scope & outline

The main points of criticism to the ALI method raised in the reviews pertain to:

1. The legal tenability of the ALI method.
2. The use and definition of P_C and its derivation from statistical theory and its relationship to alternative measures.
3. The Monte Carlo approach used to incorporate parameter uncertainty.
4. The sensitivity of the ALI method to changes in the threshold parameters (X and Y)
5. The ability of the ALI method to account for changes in number of causalities through time.
6. The differences between the ALI framework and that used for OWF effect assessments of marine mammals.

In this report, we address points 2, 3, 4 and 5. Point 1 has been addressed in a discussion with legal advisors and reported separate from the methodological developments addressed here. Point 6 will be treated later. Points 3, 4 and 5 were addressed using sensitivity analysis. In general terms, a sensitivity analysis can be used to study the effect of model assumptions on the outcome of a model. Sensitivity analysis is often used to study the response of model outcomes to changes in model parameters, but can also be used to study the effect of more structural changes to a model. This is how sensitivity analysis was used in the current study.

The reviews of the ALI methodology raised concerns about the use and definition of P_C. For instance, it was questioned whether P_C was defined correctly and how it relates to other statistical measures, such as the risk ratio. In response to this point (point 2 above), we included in this report a formal derivation of the ALI method, based on the theory of conditional probabilities (section 3.1). We also show the relationship between P_C and the Risk Ratio, which was suggested to be a potentially more appropriate alternative for defining risk of an adverse outcome.

The reviews also questioned the method of parameter sampling used in the ALI, and how this method influences the amount of uncertainty concerning future population trajectories and the consequences for the probability of ALI violation (point 3 above). To address point 3, we use an alternative parameter sampling method that generates less variation in population trajectories and we compare the likelihood of ALI violation between the two methods. This alternative method has been explored earlier for the population models, but not in relation to the ALI thresholds (Van Kooten et al, 2019). The details of this analysis are outlined in section 2.2.1.

In addition to using a different parameter sampling method, we also vary the standard deviation of parameter estimates for each sampling method. Varying the standard deviation of parameter estimates will also change the amount of variation in population trajectories. This analysis therefore serves as an alternative method to study the impact of variation in population abundance on the outcome of the ALI method. The details of this analysis are outlined in section 2.2.2.

Point 4 is derived from the question of how the threshold values X and Y should be chosen and how this decision affects the outcome of the ALI (likelihood of ALI violation). To address point 4, we show how changes in ALI threshold values affect the probability of an ALI violation for a range of simulated mortality impacts from OWFs.

The ALI was originally used in assessments that assumed that added mortality from OWFs remained constant through time (Potiek et al., 2021; Soudijn et al., 2022). This raised the question whether the ALI could be modified to account for changes in the number of casualties through time. Here we test for the effect of such a modification, by comparing constant mortality with time-dependent mortality. Details of this analysis are outlined in section 2.2.4.

2 Materials and Methods

The ALI methodology uses population models to predict trajectories of future population abundance. For the analysis performed in this study, we use the population model of Northern gannets from Van Kooten et al. (2019) as a test case. In the next section we first describe the details of this population model, before outlining the details of each specific analysis.

2.1 Northern gannet population model

The analyses were performed using the population model for the northern gannet (*Morus bassanus*) as a test case (Van Kooten et al., 2019). The model for this species was chosen because the ALI threshold was violated for this species in the last cumulative assessment of OWF impacts for the Dutch continental shelf (Potiek et al., 2021; Soudijn et al., 2022). The population model of the northern gannet consists of six stages: four immature stages (age 0 to 3), one pre-breeding adult stage (age 4) and one adult stage (age > 4). The pre-breeding adult stage and the breeding adult stage share the same survival probability S_A, but – as the name suggests – the pre-breeding adults do not breed. The annual projection matrix of the northern gannet population model reads:

$$\mathbf{A} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ S_0 & 0 & 0 & 0 & 0 & S_0 \frac{F_A}{2}(1 - P_F) \\ 0 & S_1 & 0 & 0 & 0 & 0 \\ 0 & 0 & S_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & S_3 & 0 & 0 \\ 0 & 0 & 0 & 0 & S_A & S_A \end{pmatrix} \qquad \text{eq. 1}$$

To model the effect of parameter uncertainty and (environmental) stochasticity the model includes variation in parameter values. All parameter values are assumed to follow a beta distribution with mean and standard deviations as indicated in Table 1. The beta distribution was chosen mainly because its range is between zero and one. The distributions of all parameter values are shown in Figure 1.

***Table 1**: Parameters of the northern gannet population model (Potiek et al., 2021; Soudijn et al., 2022).*

parameter	*mean*	*SD*	*description*
F_A	0.700	0.0820	Breeding success
P_F	0.050	0.1250	Prob. floater
S_0	0.481	0.0853	Survival age 0
S_1	0.816	0.0393	Survival age 1
S_2	0.884	0.0293	Survival age 2
S_3	0.887	0.0301	Survival age 3
S_A	0.918	0.0199	Adult survival

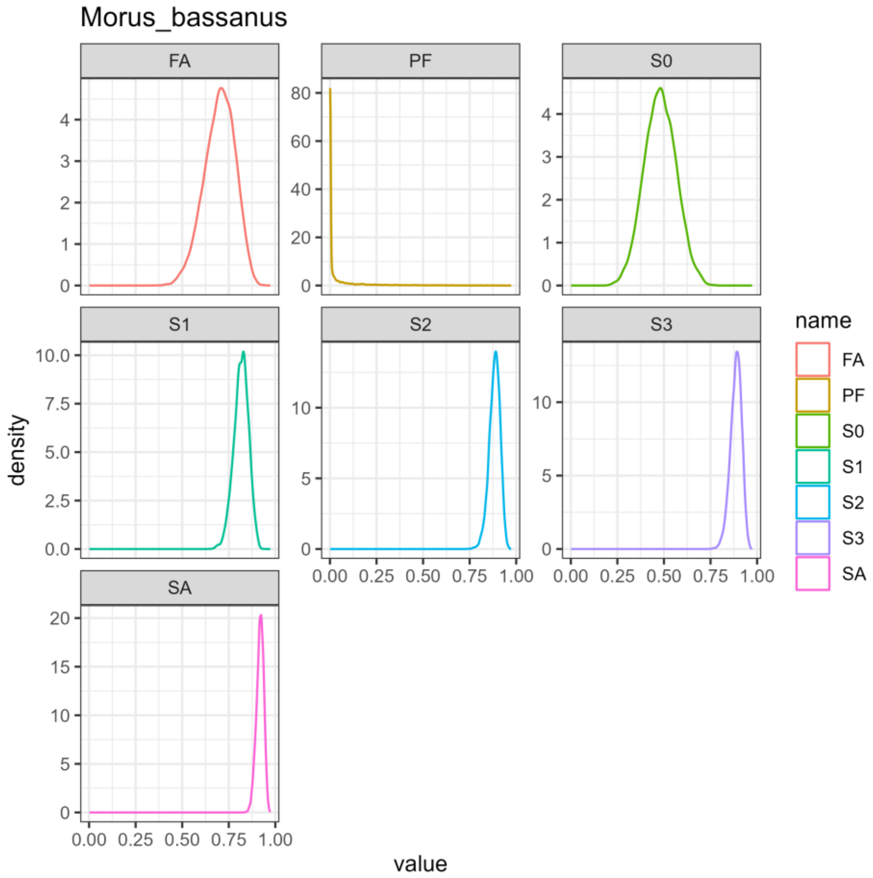

Figure 1*: Beta distributions of the parameters of the northern gannet population model. Mean and standard deviations are shown in Table 1.*

2.2 Sensitivity analysis

2.2.1 Parameter sampling method

The method of parameter sampling determines the amount of variation in the prediction of future population abundance. More variability is associated with less certainty about the predicted population trajectory. By changing the parameter sampling method we study how uncertainty affects the outcome of the ALI and aim to answer the question whether more uncertainty will increase or decrease the likelihood of an ALI violation.

The original ALI method uses sampled parameter values that do not change over time. This method was original adopted by (Van Kooten et al., 2019), because it created more variation in population abundance, which was considered to be precautionary. At the initial time point (t = 0), many (100,000) parameter sets are sampled, and each set is used to generate a population projection matrix $\mathbf{A}$. This projection matrix describes, for each life stage, the probability of an individual to transition to another life stage, or to remain in its current life stage. The asymptotic population growth rate λ describes the annual relative increase in population abundance and can be calculated directly from the projection matrix (Caswell, 2001). Population abundance *t* years later (N_t) can directly be derived from λ and the initial population abundance (N_0), according to

$$N_t = N_0 \lambda^t \qquad \text{eq. 2}$$

This equation describes an exponentially growing population when $\lambda > 1$ and an exponentially declining population when $\lambda < 1$.[1] All generated values of λ will be slightly different and λ follows a distribution that is determined by the distributions of all underlying parameters of the population model (Table 1).

[1] Eq. (1) assumes that the population grows according to the asymptotic growth rate λ, which furthermore requires that the abundances of different stages of the population are distributed according to the stable stage distribution.

A possible alternative approach would be to draw parameter values each year, which means that the population growth rate varies from year-to-year within a simulation (Van Kooten et al., 2019). This means that a year with poor conditions can be followed by a year with better conditions. With annual parameter sampling, the relationship between N_t and λ as described by eq. 2 no longer holds, because the value of λ changes each year. This also means that the population stage distribution is no longer constant within a simulation. We must therefore use the recursive relationship:

$$n_{t+1} = \mathbf{A}_t n_t, \qquad \text{eq. 3}$$

which projects the population state vector *n* from year *t* to year *t+1*, using the current sample of the population projection matrix $\mathbf{A}_t$.

To study the effect of the magnitude of variation in future population abundance, we study effect of annual parameter sampling on the outcome of the ALI. We term the method of annual parameter sampling the 'annual' method, as opposed to the 'initial' parameter sampling method that samples parameters at $t = 0$. With the annual sampling method, the population growth rate changes from year to year and it is highly unlikely that the value of λ will be high or low across many consecutive years. We therefore expect that variation in population abundance will still increase over time, but not as fast as in the original method with parameter sampling occurring only at $t = 0$.

2.2.2 Parameter standard deviation

Apart from using two different parameter sampling methods which generate a different degree of variation in population abundance, we also varied the standard deviation (SD) of all model parameters, except parameter P_F (breeding probability), because duplicating the standard deviation of this parameter resulted in unidentifiable beta distribution. This was done by a SD multiplier, which was either 0.5 (half the default SD), 1.0 (default SD) or 2.0 (twice the default SD). We expect that varying this SD multiplier leads to qualitatively similar effects as using different methods for parameter sampling (section 2.2.1), as both will impact the variation in population abundance at t = 30 years.

2.2.3 X threshold and ALI calculation

In the ALI methodology, X represents the threshold value for the acceptable decline in population abundance after 30 years of the impacted scenario, relative to the unimpacted case (Potiek et al., 2022). For example, X = 0.3 means that any decline in population abundance due to OWF impacts greater than 30% is deemed unacceptable. Lower X threshold values should be considered more strict. For example, a 20% population decline would be deemed acceptable for X = 0.3, but not for X = 0.15.

The value of X is recalculated from the acceptable decline over a period of T_{ref} years, termed X' (with apostrophe) which is decided by policy. Here, T_{ref} is the maximum of 10 years and three times the generation time, which is defined as the expected age at which a parent has produced its lifetime reproductive output (Caswell, 2001; Potiek et al., 2022). The equation to calculate X (30 years) from X' (T_{ref} years) is:

$$X = 1 - (1 - X')^{\frac{30}{T_{ref}}} \qquad \text{eq. 4}$$

From the value of X, the absolute threshold population abundance after 30 years is calculated as:

$$N_T = (1 - X)N_0(30) \qquad \text{eq. 5}$$

where $N_0(30)$ is the median unimpacted population abundance after 30 years. Population projections that drop below the abundance threshold N_T are considered a threshold violation. This holds for both the unimpacted (reference) scenario, and the impacted OWF scenario. The probability of a violation without impact ($P_{v,u}$) is the fraction of unimpacted population projections that result in a population abundance at $t = 30$ below N_T. Similarly, the probability of a violation with impact ($P_{v,i}$), is equal to the fraction of impacted population projections that drops below N_T. Both probabilities ($P_{v,u}$ and $P_{v,i}$) therefore depend on the X threshold value. A lower (more strict) X threshold value will lead to more threshold violations and will increase both $P_{v,u}$ and $P_{v,i}$. The question addressed in this analysis, is how this affects the likelihood of an ALI violation.

The outcome of the ALI is determined by the level of causality P_C, which is calculated as:

$$P_c = \frac{P_{v,i} - P_{v,u}}{P_{v,i}} = 1 - \frac{P_{v,u}}{P_{v,i}} \qquad \text{eq. 6}$$

The ALI is violated if the P_C exceeds the threshold causality value P_T, which is set by policy makers. Potiek et al. (2022) used an alternative method to determine whether the ALI is violated, based on Y, the threshold value for $P_{v,i}$. Y can be calculated by substitute P_T for P_C in eq. 5 and rearranging, which leads to:

$$Y = P_{v,i}^{T} = \frac{P_{v,u}}{1 - P_T} \qquad \text{eq. 7}$$

When using Y, the ALI is violated if $P_{v,i} > Y$. The advantage of using eq. 6, instead of eq. 7, is that one can directly judge the outcome of the ALI for different values of P_T (Table 2), without the need of recalculating each P_T threshold to an Y threshold. We therefore prefer to present the causality level P_C, instead of multiple Y threshold value for each value of P_T.

From eq. 6, we see that P_C increases with $P_{v,i}$ and decreases with $P_{v,u}$. It is not straightforward to tell whether P_C will increase or decrease with lower X values, because this depends on the how fast both $P_{v,u}$ and $P_{v,i}$ respond to changes in X. We therefore numerically explored the effect of different X threshold values upon the chance of threshold violation with and without impact and the resulting causality level. We did this for three values of X': 0.3, 0.15 and 0.01, corresponding to an acceptable relative decline in population abundance over T_{ref} years of, respectively, 30%, 15% and 1% (Table 2). We repeated this analysis for a range of simulated OWF impacts as additional mortality of individuals over 2 years of age. The mortality impact of OWFs were modelled by reducing survival parameters S_2, S_3, and S_A, ranging from 0.0% reduction (no impact) to 5% survival reduction. In total, 11 levels of OWF impacts were used (Table 2).

For each different method, X threshold, survival impact multiplier and standard deviation multiplier (Table 2; 198 combinations), we calculate the probability of an impact ($P_{v,i}$) and the causality level (P_C; eq. 6). This causality level can be compared with any of the four P_T threshold values used (Table 2).

Table 2*: Overview of methods, variables and parameter explored in the sensitivity analysis. The different parameter sampling methods, X threshold values, OWF impacts and SD multiplier values resulted in 198 unique values for P_C, which were compared to all threshold values (P_T).*

Factor	**Values**	**Number of levels**	**Description**
Parameter sampling method	'initial' vs 'annual'	2	Parameters are sampled at t = 0 ('initial') and remain constant through time or parameter are sampled each year ('annual')
X threshold	0.3, 0.15, 0.01	3	Acceptable population decline as proportion of unimpacted population. Corresponding to acceptable decline of, respectively, 30%, 15% and 1%
OWF impact (% survival reduction)	0%, 0.25%, 0.5%, 0.75%, 1%, 1.5%, 2%, 2.5%, 3%, 4%, 5%	11	To simulate OWF impact, 11 different impacts were applied to survival parameters S_2, S_3 and S_A. An OWF impact of 2.5% corresponds to a 2.5% decrease in baseline survival. An impact of 0% represents the unimpacted scenario.
SD	0.5, 1.0, 2.0	3	Standard deviation multiplier applied to all model parameters except PF (breeding probability)
P_T threshold	0.10, 0.33, 0.5, 0.66	4	Threshold causality levels for ALI violation.

2.2.4 Variation of casualties through time

To explore the effect of variation in the number of casualties through time we apply four different time-dependent mortality sequences (Figure 2). These sequences all have the same mean mortality across the entire period (30 years), and only vary with respect to when mortality is applied. Obviously, a higher mortality rate will increase the chance of an ALI violation. As opposed to the survival multiplier described above, the mortality rates are subtracted from baseline parameters S_2, S_3 and S_A. Using the different sequences, we study whether there are clear differences in the distribution of population abundance at t = 30, because this distribution determines the outcome of the ALI.

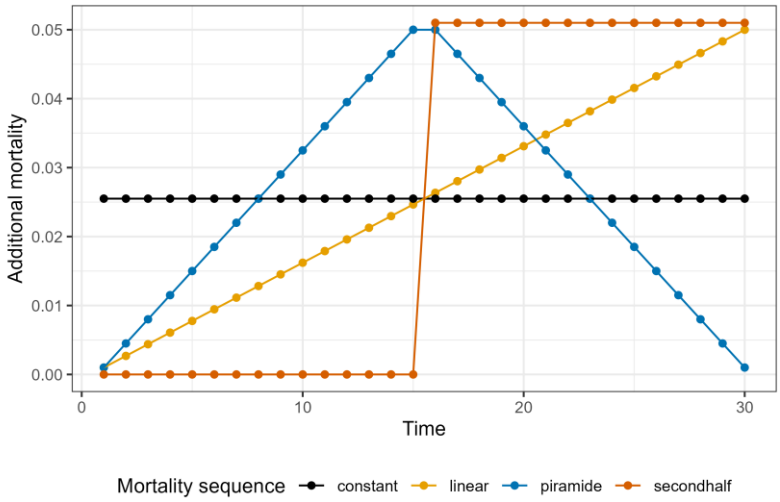

Figure 2*: Mortality sequences used to study the effect of time-dependent variation in number of casualties*

3 Results

3.1 Derivation of P_C and Y

3.1.1 Background

In this section, we show how the quantities P_C and Y (the threshold value) are derived from general statistics, and we relate them to other commonly used statistical measures.

The currently used ALI formulation is *'The probability of a population decline of X% or more, 30 years after the impact, cannot exceed Y'*. In this formulation, there are two threshold values which together determine the ALI. The value of *X* defines what is considered an unacceptable outcome. Considerations for determining appropriate values for X can be found in the main ALI report (Potiek et al., 2022). The value of *Y* is the threshold probability that the unacceptable outcome (a decline larger than X%) occurs. Defined this way, the ALI both specifies what is deemed an unacceptable outcome and considers the inherent uncertainty in the future development of natural populations. The ultimate consequence of this uncertainty is that the unacceptable outcome (a population reduction exceeding X%) might occur even without any additional impact.

3.1.2 Process and derivation of probabilities

The process we use to derive our probabilities that the threshold X is violated, is as follows: Given a population dynamical model for a bird species, a threshold population decrease X and a distribution for each of the parameter values in the model, we draw 100,000 parameter sets for the population model. These are the 'unimpacted' parameter sets. For each of these, we calculate whether the population after 30 years remains above the threshold determined by X (eq. 5). The fraction of the total number of parameter sets for which this is the case, we call $P_{n,u}$, or the probability of no threshold violation when unimpacted. It follows that $1 - P_{n,u}$ is the probability that any of the sampled unimpacted parameter sets leads to a threshold violation. We call this $P_{v,u}$, the probability of a violation when unimpacted.

Next we use the sample of 100,000 unimpacted parameter sets and add to each the impact for which we want to study the effect. This impact takes the form of a higher mortality rate in the population model, which is added to the background mortality. In the same way as above, we calculate the probability of no threshold violation with impact $P_{n,i}$ and of a threshold violation with impact $P_{v,i}$.

It is important to realize that the probability of a violation with impact depends on the outcome (violation or not) without impact, because additional mortality derived from impact is added to the unimpacted simulations. Based on this, we can further specify these probabilities as conditional probabilities $P_{n,i|n,u}$, $P_{v,i|n,u}$ and $P_{n,i|v,u}$, $P_{v,i|v,u}$. The notation $P_{v,i|n,u}$ means the probability of a threshold violation (*v*) with an impact (*i*), given that there was no violation (*n*) without the impact (*u*). Other conditional probabilities follow the same notation. The probabilities $P_{n,i|v,u}$ and $P_{v,i|v,u}$ are 0 and 1 respectively, because of the nature of the impact: Any set of parameter values which results in a threshold violation without impact, will also result in a violation when the mortality is increased further, which is what the impact does. Therefore, $P_{v,i|v,u}$ equals 1 and, by necessity, $P_{n,i|v,u}$ equals 0.

The total probability that a set of parameters from our 100,000 sets falls below the threshold X after the impact then equals

$$P_{v,i} = P_{v,u} \cdot P_{v,i|v,u} + P_{n,u} \cdot P_{v,i|n,u} \qquad \text{eq. 8}$$

But because $P_{v,i|v,u}$ equals 1, this simplifies to:

$$P_{v,i} = P_{v,u} + P_{n,u} \cdot P_{v,i|n,u}. \qquad \text{eq. 9}$$

For the determination of our threshold values, we are specifically interested in $P_{n,u} \cdot P_{v,i|n,u}$, the probability that an observed violation of X with impact, occurs in a parameter set which did not yield a violation of X without impact. We can rearrange the equation above, to see that:

$$P_{n,u} \cdot P_{v,i|n,u} = P_{v,i} - P_{v,u} \qquad \text{eq. 10}$$

The left and right hand side of this equation are two ways to express the probability that if we draw one random set of parameters from all our 100,000 simulated sets, we obtain a set that leads to no violation without impact, and violation with impact. However, for our purpose we are interested in the probability that an *observed* violation after impact came from a parameter set which did *not* yield a violation *without* the impact. In other words, we have to correct for not drawing one parameter set from all 100,000, but only from among those that lead to a threshold violation with the impact. To obtain this we multiply by the total number of parameter sets (100,000), and then divide by the number of parameter sets (between 0 and 100,000) which led to a violation after the impact. $P_{v,i}$ is the inverse of this, the number of violations divided by the total number of simulations. Hence, we can divide the above by $P_{v,i}$. This, which we call P_C, is the probability that any observed threshold violation after the impact is caused by applying the impact, and did not already lead to a violation before the impact was applied:

$$P_c = \frac{P_{v,i} - P_{v,u}}{P_{v,i}} \qquad \text{eq. 11}$$

This quantity is strictly between 0 and 1, because $0 \leq P_{v,u} \leq P_{v,i}$, since a violation without impact *always* also results in a violation with impact.
The quantity we identify as P_C, and which we call 'causal certainty' is used in epidemiology, where it is called the 'attributable fraction among the exposed'. The principle of attributable fractions was first developed by Levin (1953). It is the proportion of incidents, such as patients having certain symptoms, that can be attributed to a risk factor such as a certain disease. In our case the presence of an OWF is the risk factor that potentially causes a threshold violation

Box 1
A commonly used alternative to attributable fractions is the Risk Ratio (RR), also known as relative risk. This is calculated as:

$$RR = \frac{\frac{N_{v,i}}{N_{total,i}}}{\frac{N_{v,u}}{N_{total,u}}} = \frac{P_{v,i}}{P_{v,u}} \qquad \text{eq. 12}$$

Where N indicates the number of simulated parameter sets resulting in the outcome specified in the subscript. Eq. (10) shows that RR is calculated as the ratio between the probability (or risk) of a violation with impact, and that without an impact. The RR, like the causal certainty we use, quantifies the relationship between a violation with and without the impact, but it does so as a factor: RR tells us how many times more likely we are to observe a violation with impact, compared to without impact. Because it is a ratio, RR varies between infinity (when there are almost no violations without impact, but more with impact) and zero (when there are no violations with impact, but there are violations without impact). This latter case implies that the impact reduces the probability of violation. In our case, where the impact consists of added mortality, this outcome is impossible, so our RR is limited between 1 (when there are exactly as many violations with and without impact, so no effect) and infinity. Ultimately, we want to use our indicator of the 'impact effect' (be it P_C or RR or another measure) as a choice criterium for policymakers, based on how much (un-)certainty about the cause of a possible violation is acceptable or problematic. While it is possible to use RR, we find it less intuitive, because RR=1 means no effect, and full certainty is obtained when RR equals infinity. To overcome this, we chose to use the attributable fraction, or causal certainty. However, the two are related. If we define RR_{adj} as RR but adjusted so that it is zero at no effect (by subtracting one) and then to be 1 at complete certainty when RR=infinity (by dividing by 1+RR), we obtain:

$$RR_{adj} = \frac{(RR-1)}{1+(RR-1)} \qquad \text{eq. 13}$$

If we substitute here that $RR = \frac{P_{v,i}}{P_{v,u}}$ and simplify, we see that $RR_{adj} = P_c$. In other words, adjusting RR to be between zero and unity makes it equal to our criterium for causal certainty, but our formulation is more practical for defining thresholds.

3.2 Results sensitivity analysis

3.2.1 Parameter sampling method

Using annually varying parameters leads to a non-monotonic development of population abundance, with years in which population abundance declines alternated by years in which population abundance increases (Figure 3). Within the simulations for Northern gannet, modelled population abundances increase with no impact and decrease for 2.5% survival reduction.

Table 3*: Summary statistics at t = 30 for the two methods of parameter sampling for Northern gannet.*

method	Variance	standard deviation	mean	median	Skewness
annual	0.029	0.171	1.297	1.288	0.314
initial	1.243	1.115	1.582	1.312	1.793

With the original ALI method, standard deviation of population abundance increases exponentially through time (Figure 3 and Figure 4). In contrast, with annual parameter sampling the standard deviation of population abundance increases approximately linearly with time. The distribution of population abundance becomes positively skewed with increasing time for the initial sampling method, while skewness remains around zero using annual parameter sampling. Skewness represents the degree of asymmetry of a distribution, with positive skewness being associated with distributions that have a long right tail (mean exceeds the median). The median population abundances are remarkably similar between the two methods, while mean population abundance increases slightly faster with time for initial parameter sampling method. Plotting the distribution of population abundance at t = 30 reveals the large difference in distribution of population abundance between the two methods (Figure 5 & Table 3).

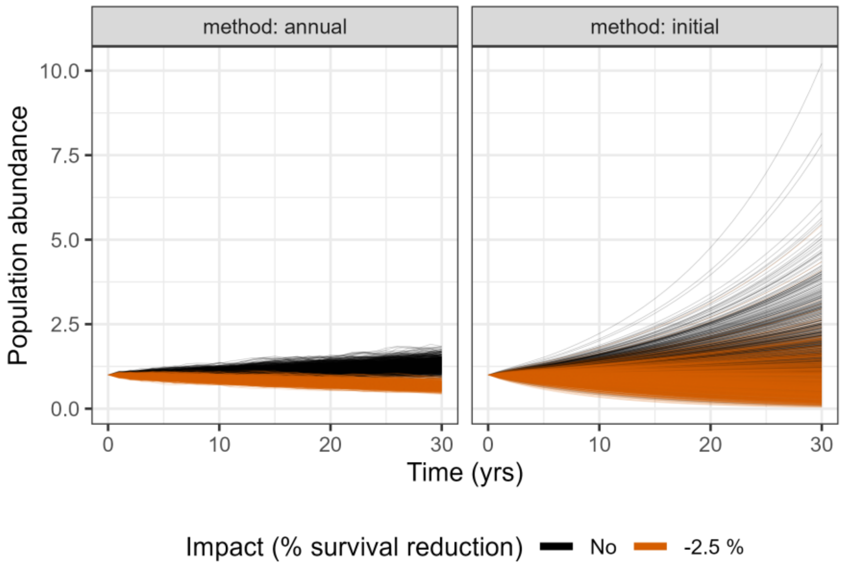

Figure 3: Population trajectories with annual and initial sampled parameters for two impact levels. All simulations start at N(0) = 1. Per impact level, 1000 replicate simulations are shown.

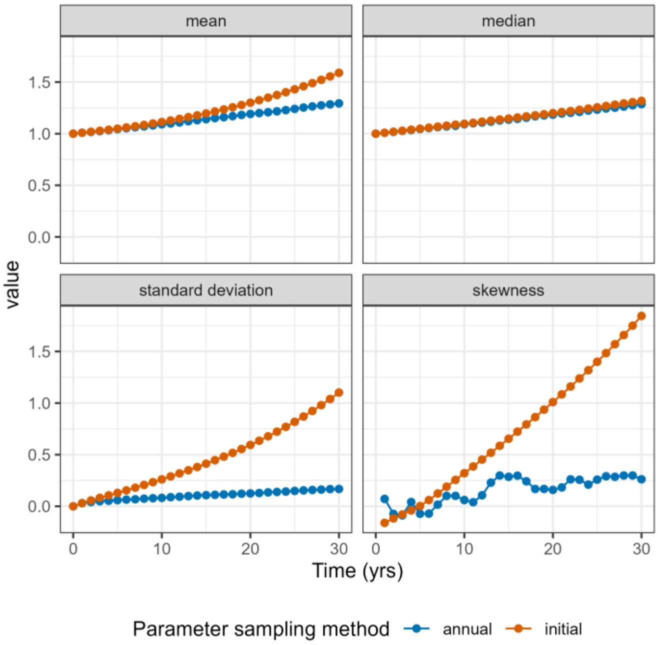

Figure 4: *Statistics of modelled population abundance for the two different methods considered. Note that only median population abundance are approximately equal (the lines are on top of one another), while variation in population abundance (standard deviation) and mean and skewness are larger for the original method that uses initial parameter sampling. All statistics are based on 100,000 replicate simulations.*

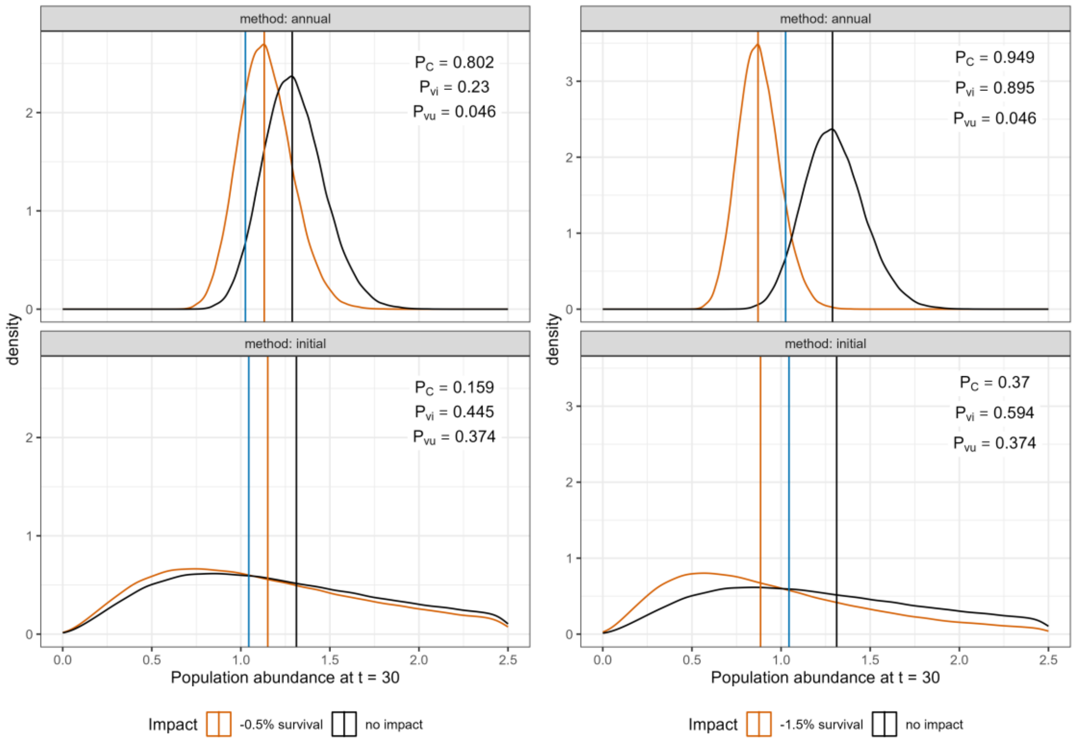

Figure 5: *Distributions of population abundance at t = 30 years for the two different methods of parameter sampling and two impacts. The black and orange lines represent distributions of unimpacted and impacted population abundance, using an impact (% survival reduction) of 0.5% (left panels) and 1.5% (right panels). Vertical lines indicate their median values. The blue line represents the threshold population abundance (N_T) corresponding to an X' threshold of 0.3 (30% lower population abundance at T_{ref} years). $P_{v,u}$ and $P_{v,i}$ correspond to the fractions of, respectively, unimpacted (black) and impacted (orange) population projections below this threshold. The median impacted population abundance is above the threshold population abundance (N_T) with an impact of 0.5% for both annual and initial parameter sampling. With a 1.5% survival reduction, the median impacted population abundance is below N_T*

3.2.2 Effect of parameter sampling method on ALI parameters

Using simulations with a survival reduction of 0.5%, we assessed the effect of the parameter sampling method on ALI parameters $P_{v,u}$, $P_{v,i}$ and P_C (Figure 5; left panels). The probability of an X threshold violation is much larger with initial parameter sampling (Figure 5). This is mainly due to larger variance and positive skewness. With X' = 0.3 in T_{ref} years, the probability of a threshold violation without impact ($P_{v,u}$) is 8.1 times lower (0.374 versus 0.046) with annual parameter sampling than with initial parameter sampling. The probability of a threshold violation with an impact level of 0.995 ($P_{v,i}$) is only 1.9 times lower (0.445 versus 0.23) for the annual parameter sampling and X' = 0.3. As a consequence, the causal certainty, P_C, of an impact is 5 times higher (0.802 versus 0.159) with annual parameter sampling (Figure 5). Using any causality threshold P_T in between 0.159 and 0.802 would lead to an ALI violation in case of annual parameter sampling ($P_T > P_C$), but not for initial parameter sampling.

So although the probability of a threshold violation with an impact level of 0.995 declines with annual parameter sampling, the difference in the probability of a false positive outcome (violation in unimpacted case) between the two methods is even larger. This leads to a higher level of causal certainty (P_C). For a survival reduction of 1.5%, the probability of a threshold violation with impact is 1.5 times larger (0.895 versus 0.594) and the resulting P_C is 2.6 times larger (0.949 versus 0.37) with annual parameter sampling (Figure 5; right panels). In this case, using any causality threshold P_T in between 0.37 and 0.949 leads to an ALI violation with annual parameter sampling, but not with initial parameter sampling.

The effect of the parameter sampling method on the probability of a threshold violation with impact ($P_{v,i}$) depends on the strength of the impact (Figure 6). For small impact levels (fraction of baseline survival > 0.99), the probability of a threshold violation ($P_{v,i}$) is lower for the annual sampling method. This happens if the median value of the impacted distribution is above the population threshold value (as in Figure 5; left panels). If the median impacted population abundance is lower than the threshold value (as in Figure 5; right panels), the probability of a violation ($P_{v,i}$) exceeds 0.5. Now, less variation in population abundance (in case of 'annual' as opposed to 'initial' parameter sampling) will lead to a larger impact. In case the median impacted population abundance is exactly at the threshold population abundance, the probability of a violation is necessarily 0.5 and the degree of variation in population abundance is irrelevant. Consequently, both methods have a probability of violation of 0.5 at the same impact level. This is where the blue and orange lines in the bottom panel of Figure 6 cross.

The causality level of the impact, which represents the fraction of $P_{v,i}$ that can be attributed to the impact, as opposed to uncertainty or biological variation, is larger when using the 'annual' parameter sampling method. This result is irrespective of the impact level used (Figure 6, top panel).

3.2.3 Effect of standard deviation

Varying the standard deviation of model parameters leads to qualitatively the same results as using different parameter sampling methods (Figure 7). Overall, more variation (higher SD multiplier), results in higher values of $P_{v,u}$. The change in $P_{v,i}$ is again more complex. If the impact level is 1% or larger and $P_{v,i} > 0.5$, a larger standard deviation leads to lower values of $P_{v,i}$, while the opposite is true if the impact level is lower than 1% and $P_{v,i} < 0.5$. Nonetheless, a lower SD multiplier always leads to less variation in population abundance at t = 30, and a higher causality level (P_C).

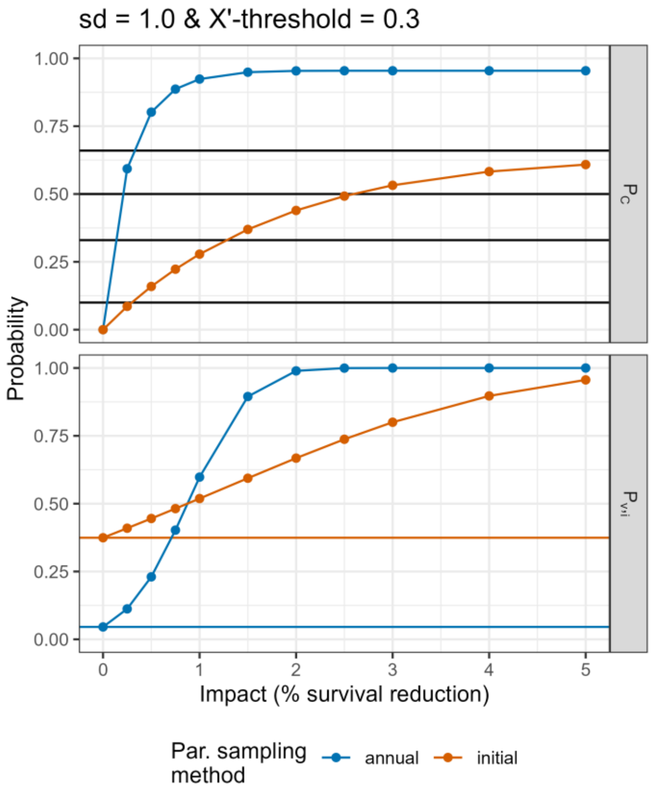

***Figure 6**: Probability of threshold violation ($P_{v,i}$; bottom panel) and associated causality level (P_C; top panel) as a function of impact levels for the two methods of parameter sampling (annual vs initial; in different colours). Impact level expressed as percentage survival reduction compared to baseline, i.e. higher values are associated with more severe impacts. The value of $P_{v,i}$ at an impact level of 0 corresponds to the probability of violation without impact ($P_{v,u}$) and are shown as horizontal lines in the bottom panel. Horizontal lines in top panel relate to different P_T threshold values (0.1, 0.3, 0.5, 0.66; Table 2). X' threshold level equals 0.3 (30% acceptable decline) and SD multiplier equals 1.0, which means the SD is equal to its normal level.*

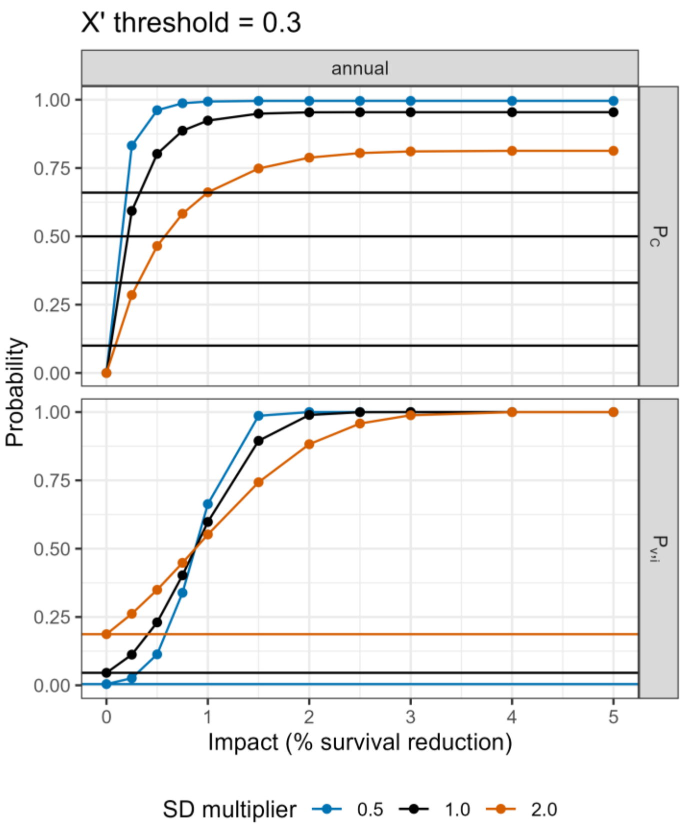

***Figure 7:** Probability of threshold violation ($P_{v,i}$; bottom panel) and associated causality level (P_C; top panel) as a function of impact for the three SD multiplier (in colours). Only results for the annual parameter sampling method are shown. Results for 'initial' parameter sampling method are qualitatively similar. Horizontal lines as in Figure 6. An X threshold of 0.3 was used.*

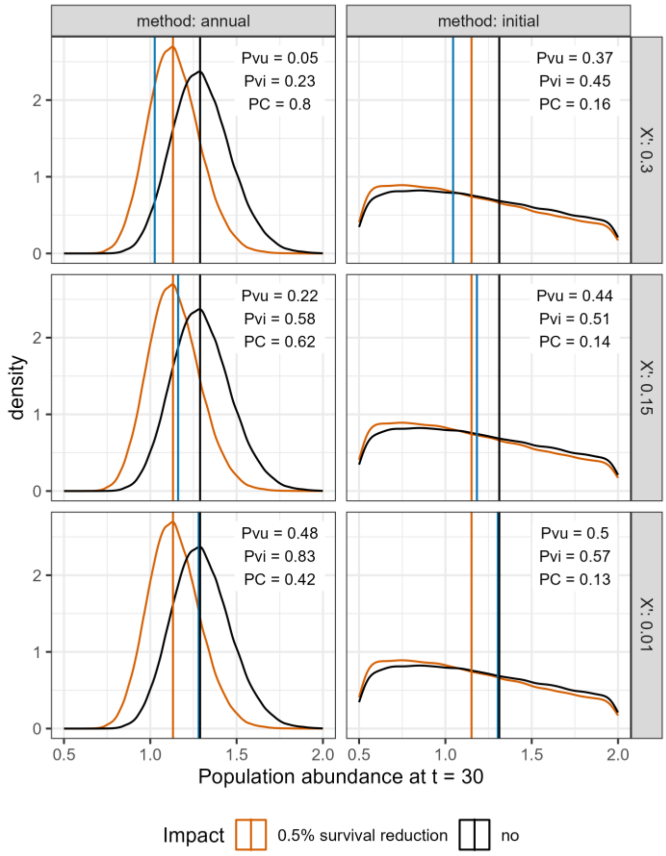

Figure 8: *Illustration of the effect of different X′ thresholds (row panels) for 0.5% survival reduction. The X′ threshold value determines the population abundance thresholds N_T (blue vertical lines). $P_{v,u}$ and $P_{v,i}$ indicate the fraction of, respectively, unimpacted (black) and impacted (orange) projections below the blue line. Note that although $P_{v,i}$ increases with a more strict threshold (X′ = 0.01), the causality level decreases because the associated increase in $P_{v,u}$ is larger. The distributions are not affected by the value of X′, only by the parameter sampling method. Vertical orange and black lines indicate median values of distributions with associated colour. An SD multiplier of 1 was used.*

3.2.4 Effect of X threshold value

The X′ threshold value determines the acceptable population decline over a period of three times the generation time, or 10 years, whichever is largest. This period is called the reference time (T_{ref}). For the northern gannet, the generation time is determined at 15.63 years, and the reference time is hence 46.9 years. An X′ threshold value of 0.01 (1% decline) means that population abundance with impact, T_{ref} years after the onset of impact, can at most be 1% lower than the median value of unimpacted population abundance. To explore the effect of X′ on the outcome of the ALI, we use X′ threshold values of 0.3, 0.15 and 0.01 (Table 2) and for each value we calculate the associated probabilities $P_{v,u}$, $P_{v,i}$ and P_C across a range of impact values. A lower acceptable population decline (lower X′ threshold) means a more strict ALI definition.

A more strict X′ threshold leads to more threshold violations without impact ($P_{v,u}$; Figure 8). If the X′ threshold is very strict (1% acceptable decline), the population threshold value is very close to the median unimpacted population abundance and, consequently, the probability of a threshold violation without impact is close to 0.5 (Figure 8). The probability of a threshold violation with additional mortality resulting from an impact is also larger for more strict X′ thresholds ($P_{v,i}$; in Figure 8). However, the change in the false positive probability ($P_{v,u}$) with stricter X′ thresholds is larger than the associated change in violation probability with impact ($P_{v,i}$). As a result, the causality level is lower at more strict X′ thresholds. This decrease in causality level with more strict X′ thresholds is independent of the simulated impact level (Figure 9).

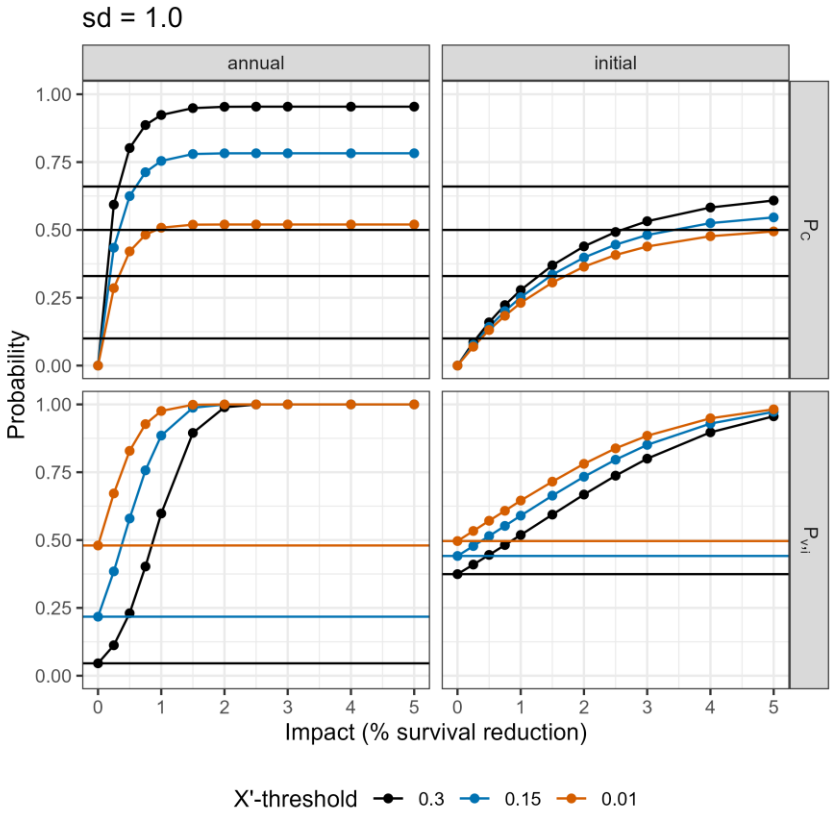

Figure 9: *Probability of threshold violation ($P_{v,i}$; bottom panel) and associated causality level (PC; top panel) as a function of impact levels for the three X' thresholds (in colours) and two methods of parameter sampling (annual vs initial; different column panels). Horizontal lines as in Figure 6. An SD multiplier of 1.0 was used.*

Because the ALI is violated if the causality level exceeds a certain P_T threshold value, choosing a more strict X' threshold (less acceptable population decline) can in some cases prevent a ALI violation. For example, using annual parameter sampling (Figure 8), an impact level of 0.995 and a P_T threshold of 50% (0.5), the ALI is violated if one would choose an X threshold of 0.3 or 0.15, but no ALI violation would occur at a more strict X' threshold of 0.01 (1% acceptable decline). In other words, for a given P_T threshold, choosing a stricter X' threshold makes an ALI violation less likely and so allows for a higher impact to be judged as acceptable. This dependency is problematic given the use of the framework, as it implies that under a stricter X' threshold, a violation of the ALI is less likely for a given mortality impact from OWFs (or a larger impact is required before the Y threshold is violated).

3.2.5 Variation of causalities through time

The different mortality sequences affected the trajectory of the population through time, but did not affect the distribution of the population at t = 30 (Figure 10 & Table 4). Because the ALI calculation is based on the distribution of population abundance at t = 30, it does not matter which mortality sequences is applied in terms of the probability of an ALI violation. It is, however, crucial to apply the correct mean mortality level across the entire period, as higher or lower mortality levels will still change the probability of an ALI violation. In addition, the results might depend on the model used, and age/size classes subject to additional mortality.

Table 4: *Statistic related to the distribution of population abundance at t = 30 for different mortality sequences. See Figure 2 for an overview of the different sequences used.*

statistic	mean	median	sd	variance
constant	0.623	0.618	0.084	0.007
linear	0.623	0.618	0.084	0.007
piramide	0.620	0.615	0.083	0.007
secondhalf	0.619	0.614	0.083	0.007

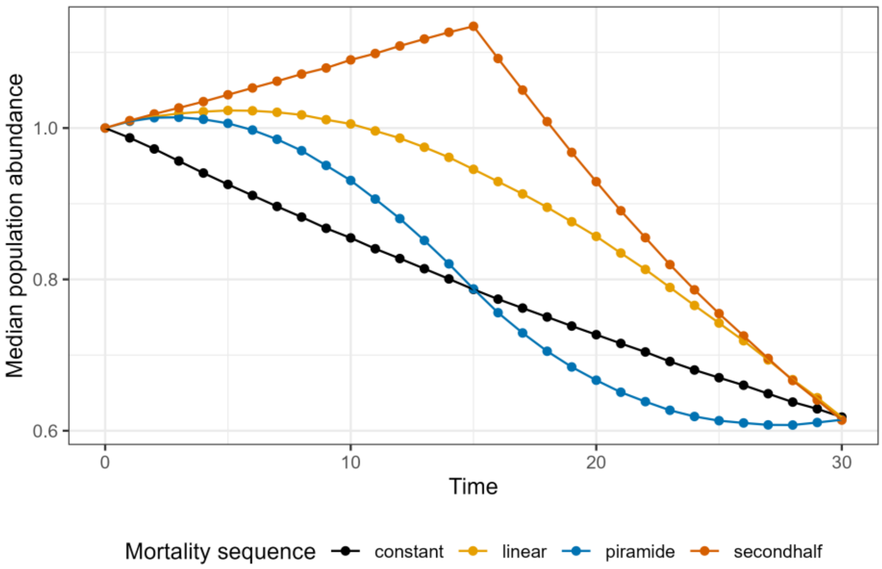

Figure 10: *Median population abundance of four different mortality sequences with identical mean values across the entire 30 year period. The population trajectory depends on the mortality sequence applied, but the distribution of population abundance at the end point is very similar between sequences. See Figure 2 for an overview of the different sequences used.*

4 Conclusions and recommendations

4.1 Method and definition P_C

The causality level used within the ALI framework (P_C) is derived from conditional probabilities and relates to other impact measures (risk ratio). It is also used in epidemiology, where it is called the 'attributable fraction among the exposed' (Levin, 1953). This quantity describes the proportion of incidents (threshold violations in the ALI framework) that are attributable to a risk factor (here, additional mortality from OWFs). We therefore conclude that the method and definition of P_C in itself is correct. However, this does not imply that this causality measure is free of issues in the current framework, which are discussed in the next section.

4.2 Sensitivity analysis

We have performed a sensitivity analysis to explore the effect of the following elements on the outcome of the ALI:

- Initial versus annual parameter sampling
- Standard deviation of parameter distributions
- *X* threshold values
- Effect of time-varying mortality impact

4.2.1 Parameter sampling method

With initial parameter sampling the variation in population abundance increases exponentially over time. In contrast, annual parameter sampling generates variation in population abundance that increases linearly over time. The predicted variation in population abundance at t = 30 is therefore considerably less (approx. 43 times lower variance; Table 3) when using annual parameter sampling. It should be noted that both methods are based on the same knowledge, as the parameters and their distribution are identical between methods (Figure 1 & Table 1). It is solely the way the population trajectories are generated that is causing the observed differences.

Based on our findings, we conclude that the method of annual parameter sampling should be preferred over initial parameter sampling, because of the following reasons:

1. Data show that seabird survival and productivity rates (breeding success) can vary substantially from year-to-year (Horswill et al., 2022; Horswill & Robinson, 2015; Mavor et al., 2008; Miller et al., 2019; Sæther et al., 2013). This will lead to alternations between 'good' and 'bad' years. This is pattern is better reflected by annual parameter sampling than by initial parameter sampling.
2. The current method was chosen based on the precautionary principle (Soudijn et al., 2022; Van Kooten et al., 2019). However, our analysis shows that this precautionary principle does not hold under the current ALI methodology, because the probability of ALI violation is lower with more variation in population abundance. From the perspective of the current ALI, the most precautionary approach is to use annual parameter sampling.

Using annual parameter sampling has a substantial effect on the outcome of the ALI assessment. The reduced variation in population abundance decreases the probability of threshold violation in absence of an impact ($P_{v,u}$). The probability of threshold violation with impact ($P_{v,i}$) changes depending on the magnitude of the impact. Reduced variation in population abundance decreases $P_{v,i}$ with low impact, and increases $P_{v,i}$ with high impact. Despite this, for all modelled impacts, reduced variation in population abundance resulting from annual parameter sampling increases the causality measure P_C.

Less variation in population abundance increases the probability that a threshold violation is caused by an impact-related mortality, rather than by existing biological variation, uncertainty or stochasticity.

4.2.2 Standard deviation of parameter estimates

The effect of changing the standard deviation of parameters of the population model (Table 1) has the same qualitative effect as changing the parameter sampling method. Decreasing the standard deviation decreases variation in population abundance. As a consequence, the probability that a threshold violation is caused by the impact, instead of by variation or uncertainty increases. In other words, the detectability of an impact increases when there is less variation in population abundance.

4.2.3 X’ threshold level

A more strict X’ threshold (a lower acceptable population decline) leads to less certainty about the cause of a threshold violation (lower P_C values). The decline in P_C values for stricter X’ thresholds is caused by the larger probability of threshold violation in absence of an impact. This high false positive probability reduces the chance that a threshold violation is caused by an impact. The relationship between the X’ threshold and P_C can lead to a situation in which higher impact levels are tolerated (lead to causal certainty that is within the P_T threshold of the ALI) when choosing a more strict X threshold. This effect is undesirable; a stricter threshold for acceptable population decline should not lead to more room for additional mortality. On the contrary, it should leave less room for additional mortality caused by the impact.

4.2.4 Time varying mortality sequences

Using a time-varying mortality sequence does not change the outcome of the ALI calculation compared to a time-invariant mortality rate with the same mean value. This is because a pattern of time-varying mortality leads to an identical distribution of population abundance after 30 years. This distribution is what determines the outcome of the ALI.

4.3 The effect of variation and uncertainty within the ALI framework

Uncertainty about the predicted trajectory of populations derives from parameter uncertainty as well as biological variation. The latter manifests itself through changes in performance of individuals (survival & breeding success), which leads to fluctuations in population growth rates. In the current analysis we only consider environmental stochasticity: the variation in individual performance caused by (unpredictable) variation in a species’ environment. This analysis revealed that more uncertainty about the predicted population trajectory, expressed as higher variation in population abundance after 30 years, leads to more threshold violations at a given impact, which decreases the causality level P_C. Because the ALI violation is determined by comparing the P_C to a predetermined causality threshold, a higher level of impact mortality can be tolerated if there is more uncertainty about the predicted population trajectory. In conclusion, in the current ALI framework uncertainty decreases the probability of an ALI violation. This is undesirable from a management perspective and contrary to the precautionary principle, which dictates that more uncertainty requires a more cautious approach. The opposite is true in the current ALI framework, where uncertainty may mask the effect of an impact and allows for higher mortality values to be deemed acceptable.

4.4 Recommendations

We conclude that the current framework to assess acceptable levels of impact from OWFs has undesirable properties: the adopted *X* threshold influences the causal certainty of a violation in such a way that a stricter X threshold increases the probability of a threshold violation in the unimpacted scenario, which makes it is less likely that a threshold violation is the result of an impact. Therefore, when using the same P_T threshold, a stricter X' threshold allows for larger impacts before Y is violated. This dependency problem can in principle be accounted for through a framework for the choice of X and Y thresholds that accounts for their interdependency (see Potiek et al., 2023). However, this will result in a further complication of an already complex methodology. The choice of appropriate X' and Y values would require a deep understanding of the method by its users (policymakers), and even then would remain prone to accidental misuse. We therefore recommend to revise the current framework. A revised framework should avoid the use of a causality measure that attempts to correct for uncertainty (such as P_C). Instead, a framework should be developed in which increased uncertainty leads to a more cautious approach, in the sense that the maximum allowable impact level decreases with increasing uncertainty. Avoiding the concept of causality will also simplify the method considerably, which will contribute to its correct application. The revised ALI methodology should furthermore use annual parameter sampling, instead of initial parameter sampling, for reasons explained above. Standard deviations of parameter estimates should remain at their default values.

Such a revised framework should not have consequences for the original formulation of the ALI, which is expressed as "*The probability of a population decline of X% or more, 30 years after the impact, cannot exceed Y"*. This formulation does not refer to the causality measure; it only refers to the probability of a population decline of *X*% or more, irrespective of its cause. It should be noted, however, that the ALI expression above is multi-interpretable with respect to the quantity over which the decline is calculated. Currently, it suggests a comparison of population abundance between two different points in time, namely $t = 30$ and $t = 0$. However, the ALI compares the population abundance at $t = 30$ between two different scenarios, one without impact from OWFs, versus one with impact. We therefore recommend that the ALI formulation should be revised to properly reflect the latter case. Such a reformulation could be as simple as adding the clause *'as a consequence of the impact'*, but the exact formulation requires careful consideration and consultation with policymakers.

Until such a revised framework is available and approved, the current ALI methodology is used. Given the results of the sensitivity analysis, the species-specific policy decisions of X and P_T should be reconsidered. Within a separate document, we present a guide for policy decisions for the current ALI methodology (Potiek et al., 2023). Here, we urge to choose a more strict threshold for P_T whenever a strict threshold for X is chosen.

5 Quality Assurance

Wageningen Marine Research utilises an ISO 9001:2015 certified quality management system. The organisation has been certified since 27 February 2001. The certification was issued by DNV.

References

Caswell, H. (2001). *Matrix Population Models: Construction Analysis and Interpretation* (2nd ed.). Sinauer Associates.

Horswill, C., Miller, J. A. O., & Wood, M. J. (2022). Impact assessments of wind farms on seabird populations that overlook existing drivers of demographic change should be treated with caution. *Conservation Science and Practice*, *4*(4), e12644. https://doi.org/10.1111/csp2.12644

Horswill, C., & Robinson, R. A. (2015). *Review of Seabird Demographic Rates and Density Dependence* (JNCC Report No. 552; Issue JNCC Report No. 552, p. 126).

Levin, M. L. (1953). The occurrence of lung cancer in man. *Acta Unio Int Contra Cancrum*, *9*(3), 531–541.

Mavor, R. A., Heubeck, M., Schmitt, S., & Parsons, M. (2008). *Seabird numbers and breeding success in Britain and Ireland, 2006* (31; UK Nature Conservation, Issue 31, p. 113). Joint Nature Conservation Committe.

Miller, J. A. O., Furness, R. W., Trinder, M., & Matthiopoulos, J. (2019). The sensitivity of seabird populations to density-dependence, environmental stochasticity and anthropogenic mortality. *Journal of Applied Ecology*, *56*(9), Article 9. https://doi.org/10.1111/1365-2664.13448

Potiek, A., Hin, V., Kraal, J. J., & Van Kooten, T. (2023). *Framework for methodology Acceptable Levels of Impact: Guide for policy decisions.* (Report 23-466). Waardenburg Ecology.

Potiek, A., IJntema, G. J., van Kooten, T., Leopold, M. F., & Collier, M. P. (2022). *Acceptable Levels of Impact from offshore wind farms on the Dutch Continental Shelf for 21 bird: A novel approach for defining acceptable levels of additional mortality from turbine collisions and avoidance-induced habitat loss*. Bureau Waardenburg.

Potiek, A., Leemans, J. J., Middelveld, R. P., & Gyimesi, A. (2021). *Cumulative impact assessment of collisions with existing and planned offshore wind turbines in the southern North Sea. Analysis of additional mortality using collision rate modelling and impact assessment based on population modelling for the KEC 4.0.* (21–205). Bureau Waardenburg.

Rijksoverheid. (2021). *Aanvullend ontwerp Programma Noordzee 2022-2027*. https://open.overheid.nl/documenten/ronl-72c2fa9b-963c-4b28-a058-ecdf9254c043/pdf

Sæther, B.-E., Coulson, T., Grøtan, V., Engen, S., Altwegg, R., Armitage, K. B., Barbraud, C., Becker, P. H., Blumstein, D. T., Dobson, F. S., Festa-Bianchet, M., Gaillard, J.-M., Jenkins, A., Jones, C., Nicoll, M. A. C., Norris, K., Oli, M. K., Ozgul, A., & Weimerskirch, H. (2013). How Life History Influences Population Dynamics in Fluctuating Environments. *The American Naturalist*, *182*(6), Article 6. https://doi.org/10.1086/673497

Soudijn, F., van Donk, S., Leopold, M. F., van der Wal, J. T., & Hin, V. (2022). *Cumulative population-level effects of habitat loss on seabirds 'Kader Ecologie en Cumulatie 4.0'* (C007/22). Wageningen Marine Research. https://doi.org/10.18174/553775

Van Kooten, T., Soudijn, F., Tulp, I., Chen, C., Benden, D., & Leopold, M. (2019). *The consequences of seabird habitat loss from offshore wind turbines, version 2: Displacement and population level effects in 5 selected species* (C063/19; Issue C063/19). Wageningen Marine Research. https://doi.org/10.18174/496173

Justification

Report C088/23
Project Number: 4316100314

The scientific quality of this report has been peer reviewed by a colleague scientist and a member of the Management Team of Wageningen Marine Research

Approved: Dr. F.H. Soudijn
Researcher

Signature:

Date: 14-12-2023

Approved: Dr. A.M. Mouissie
Business Manager Projecten Midden-Noord

Signature:

Date: 14-12-2023

Wageningen Marine Research
T +31 (0)317 48 7000
E: marine-research@wur.nl
www.wur.eu/marine-research

Visitors' address
- Ankerpark 27 1781 AG Den Helder
- Korringaweg 7, 4401 NT Yerseke
- Haringkade 1, 1976 CP IJmuiden

With knowledge, independent scientific research and advice, **Wageningen Marine Research** substantially contributes to more sustainable and more careful management, use and protection of natural riches in marine, coastal and freshwater areas.

Wageningen Marine Research is part of Wageningen University & Research. Wageningen University & Research is the collaboration between Wageningen University and the Wageningen Research Foundation and its mission is: 'To explore the potential for improving the quality of life'