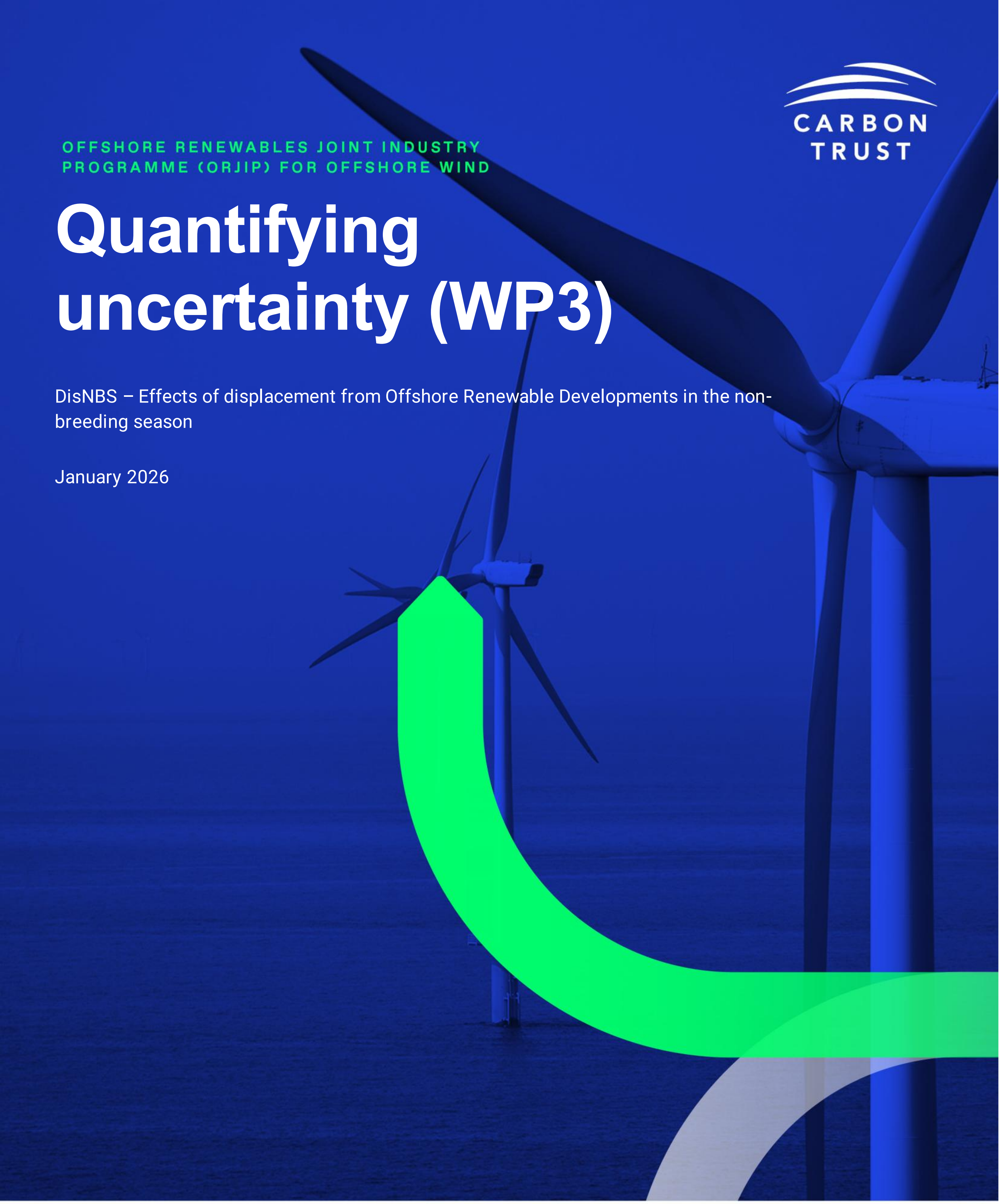


OFFSHORE RENEWABLES JOINT INDUSTRY
PROGRAMME (ORJIP) FOR OFFSHORE WIND

Quantifying uncertainty (WP3)

DisNBS – Effects of displacement from Offshore Renewable Developments in the non-breeding season

January 2026

CARBON TRUST

orjip
working to accelerate
offshore consenting

CARBON TRUST
Crown Estate Scotland
Oighreachd a' Chrùin Alba
Offshore Wind Evidence + Change Programme
THE CROWN ESTATE
EDF renewables
equinor
Ørsted
OW OCEAN WINDS
RWE
Scottish Government
Riaghaltas na h-Alba
sse Renewables
TotalEnergies

Contents

ORJIP Offshore Wind

The Offshore Renewables Joint Industry Programme (ORJIP) for Offshore Wind is a collaborative initiative that aims to:

- Fund research to improve our understanding of the effects of offshore wind on the marine environment.
- Reduce the risk of not getting, or delaying consent for, offshore wind developments.
- Reduce the risk of getting consent with conditions that reduce viability of the project.

The programme pools resources from the private sector and public sector bodies to fund projects that provide empirical data to support consenting authorities in evaluating the environmental risk of offshore wind. Projects are prioritised and informed by the ORJIP Advisory Network which includes key stakeholders, including statutory nature conservation bodies, academics, non-governmental organisations and others.

The current stage is a collaboration between the Carbon Trust, EDF Energy Renewables Limited, Ocean Winds UK Limited, Equinor ASA, Ørsted Power (UK) Limited, RWE Offshore Wind GmbH, Shell Global Solutions International B.V., SSE Renewables Services (UK) Limited, TotalEnergies OneTech, Crown Estate Scotland, Scottish Government (acting through the Offshore Wind Directorate and the Marine Directorate) and The Crown Estate Commissioners.

For further information regarding the ORJIP Offshore Wind programme, please refer to the Carbon Trust website, or contact Ivan Savitsky (ivan.savitsky@carbontrust.com) and Žilvinas Valantiejus (zilvinas.valantiejus@carbontrust.com).

Acknowledgements

The authors acknowledge Research Computing at the James Hutton Institute for providing computational resources and technical support for the “UK’s Crop Diversity Bioinformatics HPC” (BBSRC grants BB/S019669/1 and BB/X019683/1), use of which has contributed to the results reported within this paper. We would also like to thank David Nutter from the BioSS IT department for his advice and guidance on downloading and compiling packages for use in R as well as setting, scheduling and running tasks on the HPC.

This document was produced on behalf of ORJIP Offshore Wind by BioSS, DMP Statistical Solutions UK Ltd, The University of Liverpool and UKCEH The report was authored by Nicosha De Souza (BioSS), Carl Donovan (DMP statistics), Jonathan Green (University of Liverpool), Kate Searle (UKCEH), Ken Newman (BioSS), Esther Jones (BioSS), Francis Daunt (UKCEH), Adam Butler (BioSS).

The project was advised by the ORJIP Offshore Wind Steering Group and the DisNBS Project Expert Panel. We would like to thank the following organisations for their advice and support of the project via participation on the Project Expert Panel:

- Scottish Government
- NatureScot
- Natural England
- The Royal Society for the Protection of Birds (RSPB)

The Offshore Wind Evidence and Change programme

This project forms part of the Offshore Wind Evidence and Change programme, led by The Crown Estate in partnership with the Department for Energy Security and Net Zero and Department for Environment, Food & Rural Affairs. The Offshore Wind Evidence and Change programme is an ambitious strategic research and data-led programme. Its aim is to facilitate the sustainable and coordinated expansion of offshore wind to help meet the UK's commitments to low carbon energy transition whilst supporting clean, healthy, productive and biologically diverse seas.

Who we are

Our mission is to accelerate the move to a decarbonised future. We have been climate pioneers for more than 20 years, partnering with leading businesses, governments and financial institutions globally. From strategic planning and target setting to activation and communication - we are your expert guide to turn your climate ambition into impact. We are one global network of 400 experts with offices in the UK, the Netherlands, Germany, South Africa, Singapore and Mexico. To date, we have helped set 200+ science-based targets and guided 3,000+ organisations in 70 countries on their route to Net Zero.

Executive summary

This project has developed an individual-based model (IBM) of seabirds for the non-breeding season which allows time-energy budgets, and consequent impact on body mass and survival, to be simulated under baseline (current) and future (with offshore wind, OW) scenarios.

The development of a non-breeding season model of displacement is crucial in allowing the quantification of displacement impacts within the assessment process to be as transparent as possible, and to be based on the best available scientific evidence, and thereby plays an important role in filling a key evidence gap in relation to the assessment process.

The development of a mechanistic model-based approach to quantifying displacement impacts, and the uncertainty associated with this, will also be of direct use in reducing consent risk by providing an improved understanding of the mechanisms underpinning displacement risk and the ability to incrementally incorporate new evidence into assessments as it becomes available.

The IBM has been applied (within WP2) to two case studies of UK seabird populations - common guillemot on the Isle of May, and red-throated diver in the Outer Thames Estuary SPA – using a hypothetical but plausible scenario of North Sea OW development.

This work package has focused on the quantification of uncertainty and variability using the IBM, and begins by outlining the approach taken to treatment of uncertainty within the model.

Uncertainty and sensitivity analyses (UA and SA) are then used to identify key sources of uncertainty within the model, which underpin recommendations around future research and data collection.

Since the model is, as with other IBMs, relatively computationally intensive, a high-performance computing cluster is used to undertake the model runs that underpin these analyses

The uncertainty analysis (UA) focuses on evaluating overall levels of variation in mass at the end of the non-breeding season between agents and between parameter combinations

The sensitivity analysis (SA) focuses on sensitivity of model outputs to variation in several key model inputs, specifically 7 key population-level parameters that relate to initial mass, energy costs of activities, proportion of individuals susceptible to OW displacement effects, and the body mass threshold for adult mortality.

These analyses focus on three key model output metrics of direct relevance to offshore wind assessments: difference in population-level average (mean or median) mass at the end of season between impacted and baseline scenarios, which provides a proxy for OW impacts on productivity, and difference in population-level adult survival between impacted and baseline scenarios.

Results of SA show the mean or median OW impact on mass at the end of the non-breeding season are most sensitive to three input parameters: the energy cost of “active” behaviour, followed by the energy costs of “inactive” and “dive” behaviours.

SA also showed that OW impact on adult survival is most sensitive to two input parameters: mean initial body mass and the mortality threshold.

Investigations of the relationships between outputs (R-squared values for SA models, bootstrap standard errors on key output metrics) and the number of agents per combination suggest that a large number of agents (potentially much larger than 8000) may be needed to obtain stable summary statistics of OW impacts, because of high levels of inter-individual variability (e.g., in initial mass and energy costs of activities) relative to levels of variability resulting from changes in parameter values.

However, the qualitative results of the SA were the same regardless of the SA method used (random forest or regression tree) and regardless of whether 8000 or 4000 agents were considered, suggesting that key qualitative results of the analyses may be robust to the number of agents used.

SA results showed relatively low sensitivity to the proportion of individuals from the colony of interest that are susceptible to displacement, but this is likely to be because the effects of competition on the population of interest are dominated by displacement effects on individuals from other colonies, and the fundamental rebalancing of time spent feeding within the IBM in response to energy intake.

Uncertainty would be reduced through an improved understanding of the energetic costs incurred by non-breeding seabirds, which might be obtained through deployment of appropriate biologging devices and/or biophysical modelling.

Future work would be valuable to (a) better understand and incorporate uncertainty in the population-level values of parameters (e.g., via expert elicitation) and (b) evaluate sensitivity to bird distribution maps and sea-surface temperature maps.

1. Introduction

The development of offshore renewable energy can make a significant contribution to the UK Government's target to have decarbonised the energy system almost completely by 2050. However, the UK Government has a duty to ensure that offshore renewable developments (ORDs) are delivered in a sustainable manner, in accordance with the requirements of the Marine Strategy Framework Directive (EC/2008/56), the Habitats Directive (EC/92/43), the Birds Directive (EC/79/409), the Energy Act (2023) and derived legislation. ORDs have the potential to affect protected seabirds year-round, notably from collisions with turbine blades and through displacement from important habitat (Drewitt & Langston 2006; Masden et al. 2010; Scottish Government 2011). Estimating demographic consequences of displacement and barrier impacts to seabirds outside of the breeding season is one of the least developed components of the UK impact assessment process, responsible for generating considerable uncertainty in year-round impacts to protected populations (Searle et al. 2023). This is largely due to a relative lack of understanding of (a) behaviour and ecology of seabirds outside of the breeding season, (b) links between seabird non-breeding ecology and demographic parameters such as survival or subsequent breeding season performance and (c) possible impacts of displacement on seabird behaviour and ecology.

This project has exploited the increasing availability of non-breeding season data on seabird distribution, displacement rates, activity and energetics, specifically of breeding individuals of known provenance, in order to develop an individual-based model (IBM) for the non-breeding season. The IBM allows time-energy budgets, and consequent impact on seabird body mass and survival to be simulated and compared under baseline (current) and future (with offshore wind farm [OWF]) scenarios. The development of a non-breeding season model of displacement is crucial in allowing the quantification of displacement impacts within the assessment process to be as transparent as possible, and to be based on the best available scientific evidence, and thereby plays an important role in filling a key evidence gap in relation to the assessment process. The development of a mechanistic model-based approach to quantifying displacement and the consequences of displacement, and the uncertainty associated with this, will also be of direct use in reducing consent risk by providing an improved understanding of the mechanisms underpinning displacement impacts, thereby reducing the chance of major future changes in assumed levels of risk and providing a mechanism to incrementally incorporate new evidence into assessments as it becomes available.

Work Package 1 (WP1) focused on identifying the requirements for the tool, through engagement with relevant experts and stakeholders via a workshop. Work Package 2 (WP2) developed the underlying code for the model. applied the model to two UK seabird population demonstration examples, both of substantial relevance to conservation and the effects of offshore wind: (a) common guillemot on the Isle of May (Forth Islands SPA) and (b) red-throated diver in the southern North Sea (Outer Thames Estuary SPA). The case studies represent not only two species with differing biology, but also two species for which differing types and amounts of empirical data, particularly in relation to baseline spatial distributions, are available. The windfarm scenario that is considered within both of these case studies is based on synthetic (hypothetical) windfarm footprints that have been derived by randomly perturbing the shape, location and size of each individual footprint within the 2030 scenario of North Sea windfarm developments in Critchley et al. (2024) – see WP2 documentation for further detail of these perturbations. This work package, Work Package 3 (WP3), focuses on applying uncertainty analysis (UA) and sensitivity analysis (SA) to the model to allow prioritisation of future data collection, technology development and other research that can further reduce levels of uncertainty (leading, in turn, to reductions in consent risk). The detailed plan of work for this work package was developed following the discussions and workshop within WP1, to ensure that decisions taken regarding the quantification of uncertainty are consistent with current biological knowledge and with the workshop outcomes, and was developed in close conjunction with the model developers in WP2. We focus here, in WP3, upon the guillemot case study – as the model is relatively computationally intensive to run, and the sensitivity and uncertainty analyses require relatively large numbers of model runs, it was only feasible to apply these approaches to a single case study. The focus is on quantifying uncertainty and sensitivity within the context of a specific population, in relation to the underlying parameters of the model, but we anticipate that the qualitative results will have broader relevance.

Within this report we summarize the way that uncertainty is quantified and communicated within the IBM and the caveats associated with this (Section 2), the methods and results of the uncertainty and sensitivity analyses (Section 3), conclusions on the extent to which parameters and model inputs contribute uncertainty and sensitivity to the model outputs (Section 4), and recommendations regarding prioritisation of future research and data collection that arise from this (Section 5).

2. Treatment of uncertainty within the IBM

We begin by outlining the way in which uncertainty and variability are handled within the IBM, focusing on uncertainty and variability in inputs (Section 2.1) and the way that uncertainty in IBM outputs is handled (Section 2.2).

2.1. Uncertainty and variability in inputs

The IBM contains a range of input parameters and other inputs (see WP1 report and WP2 documentation) and there is substantial variability and uncertainty associated with many of these inputs. Available information around estimates, and associated uncertainty, varies between species and, in some cases, between populations of the species. The two case studies considered in WP2 focus on a relatively data rich example (guillemot at the Isle of May) and a relatively data poor example (red throated diver in the Outer Thames Estuary). We focus here, within WP3, upon the former case study.

The model contains 20 parameter values (Table 1). These 20 values represent 11 biological parameters, each with an estimate (initial time activity budgets for each of four activities, activity costs for each of four activities, initial mass, mortality threshold, proportion of individuals susceptible to windfarm effects). For nine of the 11 biological parameters – initial time activity budgets and costs per activity and initial mass - a second parameter value (SD) was used to capture the variability between individuals and (for activity costs) over time. Parameter values, are taken from the published literature wherever possible, but there is an absence of existing evidence around mortality threshold, whilst existing evidence around the final biological parameter (proportion of individuals susceptible to windfarm effects during the non-breeding season) is limited – see the WP2 report documentation for further details.

.

Table 1. Parameter values used for guillemot at the Isle of May. † these parameters are part of energetics equations which can be functionally related to sea surface temperature and dive duration – refer WP2 documentation for details.

Parameter		Calculation units	Estimate/parameter	SD
Initial time activity budget per activity	Flight	h/day	0.056	
	Dive	h/day	3.11	
	Active	h/day	10.5	
	Inactive	h/day	10.3	
Energy cost per activity†	Flight	kJ/hour	507.6	237.6
	Dive	kJ/hour	3.71	1.3
	Active	kJ/hour	113	22
	Inactive	kJ/hour	72.2	22
Initial body mass		G	929	56
Mortality threshold		G	800	
Proportion susceptible to displacement		Proportion	0.67	

No published information on population-level *uncertainty*, rather than input variability, was either available or used in the simulations.

Aside from these input parameter values, the model also depends on four other input variables (“forcing variables”): (1) a bird distribution map (which, for guillemot, is monthly and colony-specific); (2) a competition map (also monthly and population-specific for guillemot), (3) a sea-surface temperature map (monthly) and (4) polygons (shapefiles) of offshore windfarm footprints.

The sources of information and functional forms for each of these inputs for the guillemot case study are detailed in the WP2 documentation. For the guillemot case study the bird distribution maps and competition maps are both derived from colony-specific monthly utilisation distributions estimated from tracking data (Buckingham et al., 2022). For the SA here we do not consider uncertainty in any of these four sources of information, for logistical reasons (we are already considering a relatively large number of parameters for the SA, given that there are computational constraints around how many model runs we can do), but these could be considered in future work (Section 5.2).

Overall quantification of uncertainty in model outputs (UA), accounting for variability (e.g. across time and between agents) in parameters and inputs for which quantification is possible, is obtained via repeated simulation from the IBM. Stochasticity in model inputs, and stochastic elements within the model itself that are designed to capture variability, both translate into stochasticity in the key model outputs (which are population-level summaries, such as estimated impacts of windfarms on overall survival rates), allowing quantification of uncertainty in these outputs.

2.2. Uncertainty in outputs

The key output for each individual within the IBM is the mass at the end of the non-breeding season, which is assumed to provide a direct link to annual demography. For example, relationships between adult mass at the end of the non-breeding season and productivity in the breeding season that follows this can be used to translate mass into productivity in situations where relevant empirical evidence exists (there is current work underway on this within other projects) but this translation is not directly considered within the model. Adult mortality is not directly simulated within the model but is assumed to occur if mass at the end of the season falls below a mortality threshold (with the value of this threshold being a model parameter). The model is primarily designed to estimate OWF effects so differences in final adult mass and survival values between paired impacted and baseline (unimpacted) runs form the key outputs from the model (described in detail in Section 3.1.1).

The basic operation of the IBM involves running a single individual (agent) across the entire non-breeding season, with and without windfarm effects, and comparing the values obtained from the impacted and baseline (unimpacted) runs. The model produces outputs for each agent, so uncertainty and variability in these key outputs are quantified by summarizing across individuals. A key characteristic of the DisNBS IBM is that it is designed so that the stochastic elements of simulations can be generated independently for each agent – this is a key practical distinction from SeabORD breeding season model of seabird displacement impacts (Searle et al., 2014, 2018), for example, and means that, unlike in SeabORD, the calculation of uncertainty and variability in outputs takes place outside the model itself.

Variability between individuals can be estimated from the outputs by running the model with n individuals, and then quantifying the variability between the outputs generated for each individual – e.g., by calculating the standard deviation across individuals of OWF effect on final adult mass.

In practice, however, primary practical interest will typically lie in the mean response to OWFs at a population (e.g., SPA or regional) level, and in the uncertainty associated with this population-level effect. This can be captured by (1) running the model with n individuals each of m times, (2) calculating the mean across individuals for each of the m replicates, and (3) using the variation in the resulting replicate-level estimate to capture uncertainty in population-level effects. Note that this uncertainty will arise from two sources: (a) inherent stochastic variation between populations as a result of inter-individual variability and (b) uncertainty in population-level parameter values. The former will be strongly linked to population size: with a sufficiently large population size n we would expect this uncertainty to become effectively zero. The latter depends on the level of uncertainty that we assume in the population-level input parameters and variables: it will be zero unless we allow the values of some of these parameters, or variables, to vary between replicates.

3. Uncertainty and sensitivity analysis

UA and SA can be used to determine the sensitivity of model outputs to input parameters and other model inputs. UA can quantify the level of uncertainty in model outputs that is a result of both epistemic and aleatory uncertainty (Mullins, et al, 2016; Beven & Lamb, 2017) in model inputs - epistemic uncertainty in model inputs includes uncertainty about parameters that are assumed fixed but unknown, while aleatory uncertainty in inputs includes temporally or environmentally stochastic input data. SA assesses the extent to which key model outputs are sensitive to variations in model inputs and parameters. We will use a suite of simulations from the model to perform both UA and SA. We consider sensitivity and uncertainty not only in relation to each parameter individually but also in relation to combinations of multiple parameters, since focusing only on uncertainty or sensitivity to each parameter in isolation may fail, due to dependencies between parameters, to capture important characteristics of the model.

Given that there is published information on variability (e.g., between individuals and over time) for many of the parameters, but little or no information around uncertainty in the population-level estimates of these parameters, we focus here upon:

1. Within UA generating simulations for multiple agents, for the fixed set of population-level parameters (including, where relevant, mean and SD values) given in Table 1 – this will automatically account for variation in activity budgets, initial time budgets, initial mass, and these characteristics are assumed within the model to vary between individuals (and for activity budgets over time) based on the mean and SD population-level values of the parameters. We then evaluate the overall levels of variability – between agents, and between parameter combinations. Note that this differs from the standard approach to UA by only considering aleatory, and not epistemic, uncertainty. This is necessary given the lack of available information on epistemic uncertainty in this context, but means that the UA can only, in this case, have a limited interpretation (e.g. because the ranges assumed for the parameters within the simulations that underpin the UA are somewhat arbitrary, and so do not necessarily represent the real-world uncertainty within these parameters).
2. Within SA–using a sufficiently large number of agents, *n*, that the model outputs are stable (based on the UA), so that the model can effectively be regarded as (at least approximately) deterministic, and then looking at the sensitivity of these outputs to varying a subset of the model's 20 parameter values (Table 1): mean and SD of initial mass, activity costs per activity, initial time budgets per activity, mortality threshold, and proportion of individuals susceptible to windfarm effects.
3. Finally, we look at how the outputs from the UA and SA change as the number of agents per parameter combination (*n*) is increased, and, in particular, look to identify the number of agents beyond which the summary statistics of key outputs (e.g. offshore wind impacts on mass at the end of the non-breeding season) and the results of the SA are stable, in order to evaluate how robust our results are in relation to the number of agents used within the simulations.

3.1. Methods

3.1.1 Model outputs

The key model outputs that we will consider within the uncertainty and sensitivity analyses relate to the impact of windfarms on adult mass at the end of the non-breeding season. We focus, for each individual, on their mass at the end of the non-breeding season (simulation day 270), because (a) this is used in this context to calculate survival (e.g. each agent is assumed to die if final mass is below the mortality threshold, and to survive otherwise) and (b) this is assumed to provide a proxy for productivity within the subsequent breeding season. When aggregating across individuals (agents) for each model run we focus upon three key output metrics, relating to OW impacts:

M1: mean OWF impact on mass at the end of the non-breeding season = mean across agents of [mass of agent at the end of non-breeding season under impacted scenario – mass of agent at the end of the non-breeding season under baseline scenario];

M2: median OWF impact on mass at the end of the non-breeding season = median across agents of [mass of agent at the end of non-breeding season under impacted scenario – mass of agent at the end of the non-breeding season under baseline scenario]. We consider this as a potential alternative to the mean, as the mean may be sensitive to outlying and extreme values;

M3: OWF impact on adult survival = [proportion of agents whose impacted mass at the end of the season is above a mortality threshold, M - proportion of agents whose baseline mass at the end of the season is above a mortality threshold, M]

3.1.2 Population-level parameters

The model contains 20 user-specified population-level parameters (outlined in Table 1) that are relevant to non-breeding seabirds:

- Mean and SD of initial time activity budgets for each of four activities (total of 8 parameters);
- Mean and SD of energy costs for each of four activities (total of 8 parameters);
- Mean and SD of initial adult mass (total of 2 parameters);
- Proportion of individuals susceptible to displacement (1 parameter);
- Mortality threshold (1 parameter: note: only relevant for the calculation of metric M3)

This represents a large number of parameters to consider within an SA, particularly for a model that is relatively computationally intensive to run. The structure of the model suggests that the key output metrics may be relatively insensitive to initial time activity budgets, so we take an iterative approach to the SA (e.g. Frey et al., 2022) and begin by undertaking some test runs to verify if this is indeed the case, by checking whether varying the initial time budgets substantially impacts the mass values at the end of the season. Provisional results suggested little impact on the mean (although some impact on the SD and range). Since these test runs involved imposing relatively large variations in time budgets the results provided a basis to exclude these parameters from the main SA. Further test runs suggested that OWF impacts on final mass were generally less sensitive to parameters associated with standard deviations than those associated with means, so these parameters were also excluded from the SA. The main UA and SA analyses therefore focused on varying the values of 7 parameters:

- Mean energy cost for each of four activities;
- Mean initial adult mass;
- Proportion of individuals susceptible to displacement;
- Mortality threshold

3.1.3. Parameter combinations

We generated combinations of these seven parameters for use in the UA and SA by using Latin hypercube sampling (implemented using the maximinLHSfunction from the “lhs” package in R - Carnell, 2024). In all cases, the ranges of parameter values used in generating the parameter combinations were assumed, in the absence of other information (e.g. the absence of information around uncertainty in the mean) to span from -20% to +20% of the current parameter values, for all parameters except the proportion of birds susceptible to displacement effects (for which the range will be 0 to 1, the range within which this proportion must logically lie). The parameter ranges used for each parameter are shown in Table 2.

Table 2. Parameter ranges used in the UA and SA for guillemot. Ranges are taken to be -20% to 20% of the default parameter value, for all parameters except “proportion susceptible to displacement”.

Parameter	Lower	Upper

Mean energy cost per activity (providing kJ/hour within simulation)	Flight	406.08	609.12
	Dive	2.968	4.452
	Active	90.4	135.6
	Inactive	57.76	86.64
Mean initial body mass (g)		743.2	1114.8
Mortality threshold (g)		640	960
Proportion susceptible to displacement		0	1

3.1.4. Model runs

Two key choices are the number of parameter combinations (m) and the number of agents per parameter combination (n). We would ideally take the values of both m and n to be as large as possible, to order to produce stable UA and SA results, but the computational burden of running the model will also depend upon the values of both m and n, with the total number of simulated agents being $m * n$. We therefore aim to balance these two considerations by selecting values of m and n that produce relatively stable results, but also allow the model to be run within a realistic period of time. In practice, we generated model runs sequentially: this involved generating an initial batches of runs using relatively small values of n and m, checking if this is sufficient to produce stable UA and SA outputs, and continuing to increase the number of runs until stability is obtained. Our final set of model runs involved m=150 parameter combinations and n=8000 agents per parameter combination (requiring simulation of 150 * 8000 = 1.2 million agents).

Final simulations for each parameter combination involving m=8000 agents required around 2.5 hours to complete, so the total computer time to complete all n =150 parameter combinations was therefore approximately 375 hours. In order to generate the number of model runs necessary to perform an SA we therefore implemented the model with a high performance cluster (HPC), allowing us to simulate multiple model runs in parallel. The simulations were carried out using the computing resources provided by the UK crop diversity HPC (High performance computing) cluster which is utilised by seven UK institutions including BioSS and supports a variety of research fields (Percival-Alwyn et al., 2024). It has the capacity of multiple CPU cores (>5000+), large multi-nodal parallel storage capacity (~8PB) and high memory (~41 TB), making it efficient for operations requiring high computing power. Technical details of the implementation of the simulations on the HPC cluster are given in Appendix A.

An alternative to the use of a HPC to improve computational efficiency would have been to build an emulator of the model (a fast, statistical approximation to the model), and then to use the emulator to generate large numbers of model runs. However, emulation provides approximate, rather than exact, model outputs, so the use of a HPC, where feasible, provides a more direct solution to the computational challenges. The methodology for emulating stochastic models is also less well developed than emulator of deterministic models, creating methodological challenges in applying this approach to IBMs.

3.1.5. Uncertainty analysis

The distinction between uncertainty analysis (UA) and sensitivity analysis (SA) is that uncertainty analysis would usually involve simulating the values of input parameters using information on the uncertainty within estimates of those parameters. In the absence of such information, we base the UA and SA on the same set of model runs, which have been generated (Sections 3.1.3 and 3.1.4) using the same ranges for the input parameters and assuming that each input parameter is uniformly distributed within this range. Within this context, the distinction between UA and SA lies only in the way in which we analyse and interpret the outputs of these model runs: for the UA we focus upon summarizing OWF impacts on adult mass, associated uncertainties, across full set of model runs associated with the full range of 150 parameter combinations, whereas in the SA we investigate the extent to which model outputs to sensitive to the variations in specific parameters.

For the UA we:

a) calculate the windfarm impact on final adult mass for each of the entire set of 1.2 million agents, and calculate basic summary statistics of these values (including the standard deviation across agents);
b) derive the values of mean OWF impact (M1) and median OWF impact (M2) for each of the 150 parameter combinations, and calculate basic summary statistics of these values (including the standard deviation across parameter combinations).

In addition we use a linear model (applied to the entire set of values of OWF impact on final adult mass for all 1.2 million agents, and using “parameter combination” as a categorical explanatory variable) to investigate the extent to which variations in agent-level OWF impacts can be explained by differences between parameter combinations: we report the percentage of variation explained (the R-squared value) and report whether there are significant differences between parameter combinations via the p-value associated with a one-way ANOVA.

3.1.6. Sensitivity analysis

We investigate sensitivity of the three output metrics of interest (M1, M2, M3) in relation to the seven population-level parameters that were varied between the model runs. Note that the final parameter (mortality threshold) only actually appears in the calculation of metric M3, but is included for consistency in all analyses.

A range of fundamentally different sensitivity analysis methods exist, with four categories being derivative-based (e.g., Morris elementary effects), distribution- or variance-based (e.g., Sobol’ sensitivity indices), variogram-based, and regression-based (Pianosi et al., 2016, Razavi et al 2021). We restrict attention here only to SA methods for which model runs can be generated iteratively (i.e. an initial set of model runs can be generated, but this set can be expanded to include additional model runs if needed), and chose in particular two methods that fall within the regression-based categorization: regression trees and random forests, which we explain below. These two methods allow implementation of this sequential approach, which was logistically necessary given the computational intensity of our model and the timelines of this project. This excluded consideration of some SA approaches that are not appropriate to use when model runs are being generated sequentially as the procedures for generating the parameter combinations, the space-filling algorithms, require a priori specification of the number of parameter combinations, *n*.

Statistical regression methods explain variability in a response variable in relation to multiple explanatory variables and are commonly used in the analysis of empirical data. However, they can also be used for sensitivity analysis of a process-based model, such as an IBM, by applying these methods to a set of outputs from, and associated inputs to, that model. In this context the output of interest from the process-based model is treated as the "response variable", and the model inputs (e.g. parameters) are treated as "explanatory variables". A wide range of statistical methods of this form exist, and therefore have potential to be used for SA. Multiple regression is perhaps the simplest such method, but requires strong assumptions around the nature of the effects of the explanatory variables, by assuming either linearity or simple extensions of linearity. Regression trees (Breiman et al., 1983; James et al., 2021) are a flexible, typically nonlinear, alternative, that attempt to explain variation in the response variable in relation to a hierarchical partitioning of values of the explanatory variables into groups based on thresholds. The approach is termed "regression trees" because the outputs of the approach can be displayed visually as a "tree". Random forests are an extension of regression trees (Breiman, 2001; James et al., 2021) in which multiple regression trees, rather than a single tree, are produced, based on bootstrapping (resampling) the model runs and fitting a regression tree to each bootstrap sample. Predicted outputs are then averaged across trees, with the rationale for using random forests being that the accuracy of these averaged predictions is often higher than that of predictions from a single regression tree. The relative importance of individual input parameters can be evaluated, using outputs from their regression trees or random forests, by calculating the reduction in mean squared error that results from including the input parameter in the tree(s), and this provides the basis for using both methods for sensitivity analysis.

These methods are implemented using the R packages "rpart" and "randomForest" and both methods are applied to the same set of runs from the DisNBS model.

3.1.7. Robustness of results to number of agents

Finally, we evaluate the extent the results of the UA and SA are robust in relation to the number of agents (in order to validate the use of 8000 agents per parameter combination), by:

a) using nonparametric bootstrap to calculate the standard error associated with estimating mean OWF impact (M1) and median OWF impact (M2) for numbers of agents per parameter combination n ranging from 500 to 8000 (in steps of 500), in order to see how the standard errors on M1 and M2 vary as the number of agents per parameter combination n is increased. For each value of n a bootstrap standard error is calculated separately for each parameter combination, and the mean of these standard errors across parameter combinations is then calculated.
b) repeating the SA for 4000 rather than 8000 agents, in order to check that the main qualitative SA results obtained using 4000 and 8000 agents are identical.

3.2. Results

3.2.1. Uncertainty analysis

Table 3 and Figure 1 summarize the simulated differences in final (day 270) mass between impacted and baseline scenarios across the full set of 1.2 million simulated agents (from 150 different parameter combinations, with 8000 agents per combination).

It can be seen that there is a very high level of variation between individual agents, and that both positive and negative effects can occur at the individual level, with the difference in mass ranging from -149.47g to 142.54g. The average difference is negative (mean -2.51g, median -2.21g), indicating that mass is, on average, lower in impacted than baseline scenarios, but the magnitude of variation between agents is substantially higher than the magnitude of the mean (SD = 12.56g). Mass values are approximately symmetrically distributed around the mean, but the distribution has heavier tails than a normal distribution (e.g. includes more extreme values than would be expected based on normality).

Variation between parameter combinations in mean and median differences (M1 and M2) is much lower than the variation between individual agents: the SD across parameter combinations is 0.21g for M1 and 0.12g for M2. Values of M1 are consistently negative, ranging from -2.99g to -1.93g, and values of M2 have an even narrower range (-2.48g to -1.91g). Values of M1 and M2 are fairly highly correlated with each other (correlation = 0.79), but at this level of correlation the relative ranking of different parameter combinations is not always consistent between M1 and M2 – the differences between M1 and M2 (mean and median values across agents) presumably arise because there are a reasonable number of individual agents with extreme differences between impacted and baseline (which are included in the calculation of M1 but not M2).

A linear model of agent-level difference in mass (impacted-baseline), in which "parameter combination" is a categorical variable, showed that (a) there are very highly statistically significant differences between parameter combinations ($p < 10^{-15}$) but (b) "parameter combination" only explains a very small percentage of variation in agent-level differences in mass (R-squared value of 0.03% - i.e. well below 1%). The fact that we can detect significant differences between parameter combinations when they explain such a small percentage of variation explained is because the very large number of agents simulated (1.2 million) allow us to detect subtle differences in mean values between different parameter combinations.

Type	Min	Mean	Median	Max	SD
Agent	-149.47	-2.51	-2.21	142.54	12.56
M1	-2.99	-2.51	-2.51	-1.93	0.21
M2	-2.48	-2.22	-2.22	-1.91	0.12

Table 3. Summary statistics of simulated differences in final (day 270) mass between impacted and baseline. The first row shows summaries across the 1.2 million individual agents (within the 150 parameter combinations and 8000 agents per parameter combination). The second and third rows are based on calculating mean and median values per parameter combination (M1 and M2) and then summarising those values across parameter combinations.

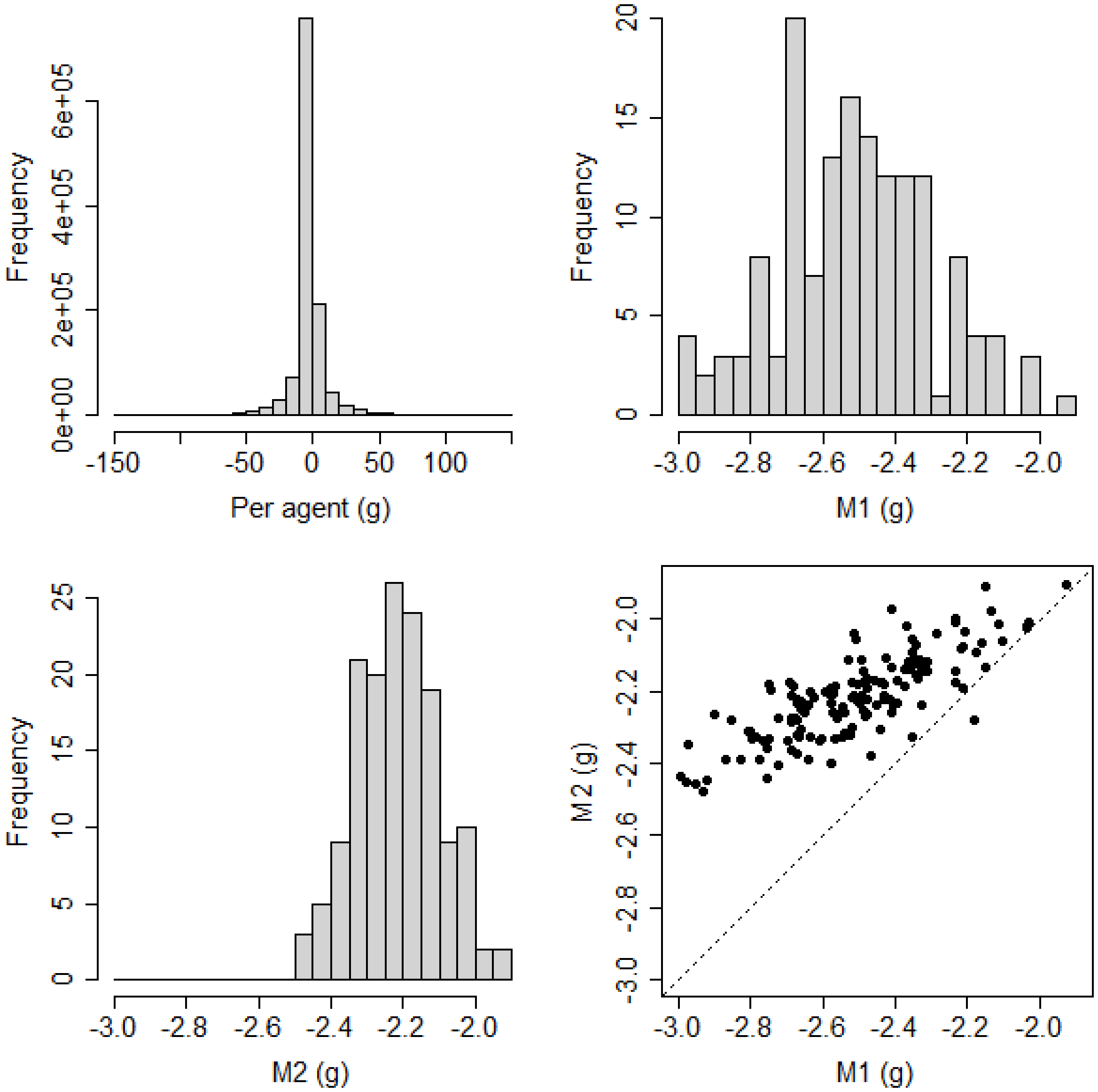


Figure 1. Histograms of differences between impacted and baseline final mass (Day 270) for each of the 1.2 million simulated agents (top left), after aggregating up to the mean value for each of the 150 parameter combinations, M1 (top right) and after aggregating up to median value per parameter combination, M2 (bottom left). A scatterplot of M1 against M2 is also shown (bottom right).

3.2.2. Sensitivity analysis

Mean of windfarm minus baseline body mass difference, M1

Figure 2 summarises the parameters that determine the outcome of the regression tree analysis for the mean of body mass difference (M1 windfarm minus baseline) on day 270. The regression tree on the left of the figure is a depiction of the branching based on increasing importance of the selected variables, whilst the right-hand part of the figure summarises the relative importance of parameters. The final nodes (leaves) in the regression tree (left-hand plot) include the assigned values (M1) for parameter combinations that fall into each node as well as the percentage of parameter combinations in that node. The most important input distribution parameters, as shown in the bar plot on the right of figure were, in order of importance, mean energy costs of Active (3.18), Inactive (1.06), Dive (0.76) and

Flight (0.46) behaviours, OWF Impact (i.e. probability of binomial distribution) (0.29), Mortality (0.27) and Body Mass (0.24). Both the regression tree and the random forest indicate that the four activity means, as a group, have more influence on y1 than the other three input parameters. As the assigned values in the regression tree indicate, the greater the active mean value the more negative the value of M1 (more body mass loss due to ORDs).

Figure 3 summarises the outcome of the random forest analysis for M1. The two figures both interpret the importance of the variable but by using different methods - The order of the variables as per their significance is median of Active (IncNodePurity = 2.44), Inactive (1.01), Dive (0.80), Flight (0.65), OWF impact (Binary Distribution probability) (0.57), Mortality (0.50) and Body Mass (0.47) as shown by the plot on the right. Results for the top four parameters are the same where %IncMSE or IncNodePurity is used, but the order of the 5th and 6th most important parameters differs between the two methods. The variance explained by the model is 47.4%.

Median of windfarm minus baseline body mass difference, M2

Figure 4 summarises the parameters that determine the outcome of the regression tree analysis for median of the body mass difference on day 270 (M2). The order of the branching shown in the regression tree on the left corresponds to the importance of the input parameters. The most important input (mean) distribution parameters in the order of importance were median of Active (0.98), Dive (0.59), Inactive (0.49), Mortality (0.12), Flight (0.11), Body Mass (0.12) and OWF Impact (0.1) as depicted in the bar plot on the right.

Similarly, Figure 5 summarises the outcome of the random forests analysis for the median of the difference in body mass on day 270 (M2). The order of the variables as per their significance is median of Active (0.69), Dive (0.43), Inactive (0.33), Flight (0.13), Body Mass (0.12), OWF impact (Binary Distribution probability) (0.12) and Mortality (0.11) and as shown by the plot on the right. The variance explained by the random forest model is 71.11%. Note that within the random forest analyses more of the variance in the median than the mean is explained – this may reflect the influence of outlying or extreme values that add noise into the values of the mean but not the median.

Windfarm minus baseline difference in survival, M3

Figure 6 summarises the parameters that determine the outcome of the random forest analysis for the difference in the proportion of agents out of the 8000 that are equal or greater than the mortality threshold on day 270 (M3) for every 150 combination values. The regression tree on the left serially represents the importance of the seven parameters. The most important input distribution parameters in the order of importance were mean of Body Mass (2.71e-03), Mortality (1.25e-03), OWF Impact (i.e. probability of binomial distribution) (4.00e-04), Flight (3.35e-04), Active (3.05e-04), Inactive (2.22e-04) and Dive (4.30e-05) as depicted in the bar plot on the right. The order of these is almost reversed as compared to the regression trees for Mean Mass (Figure 2) and Median Body Mass (Figure 4). In particular the 4 parameters related to activity types dominate change in body mass but have much less effect on change in survival, when compared to the body mass and mortality parameters.

Similarly, Figure 7 summarises the outcome of the random forests analysis for the difference in the proportion of agents out of the 8000 that are equal or greater than the mortality threshold on day 270 for every parameter combination. The order of the variables as per their significance is median of Body Mass (1.56e-03), Mortality (9.51e-04), OWF Impact (i.e. probability of binomial distribution) (5.55e-04), Flight (5.44e-04), Dive (4.85e-04), Active (4.68e-04) and Inactive (4.42e-04) as shown by the plot on the right. The variance explained by this random forest model is 34%.

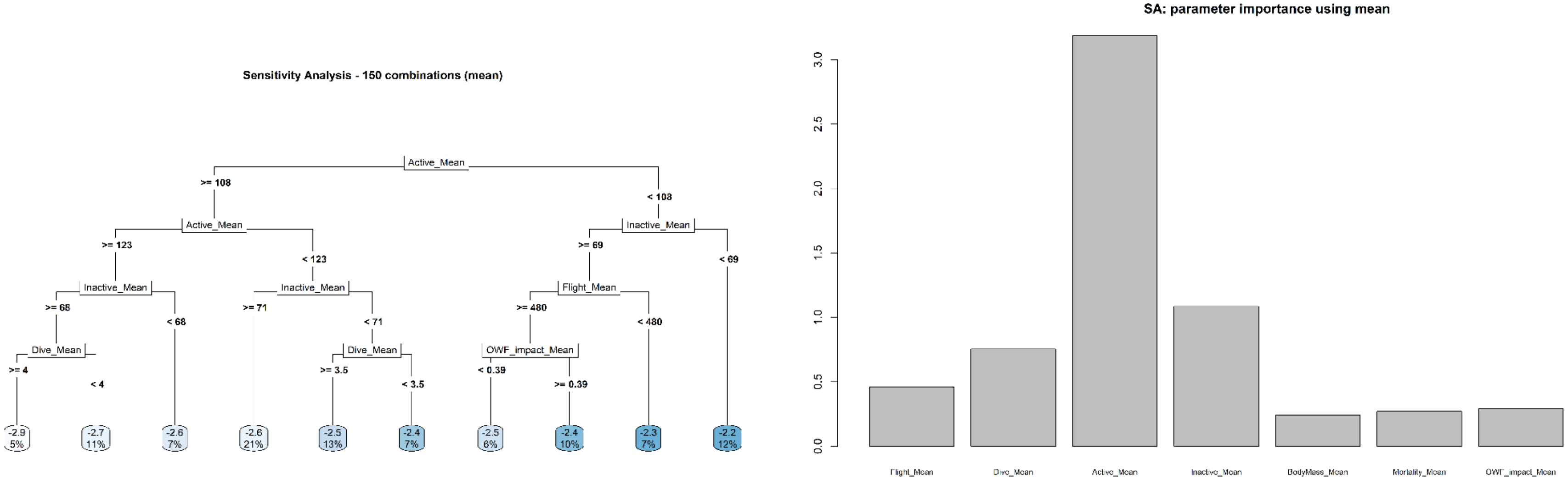


Figure 2. Variable importance within the DisNBS model based on applying regression trees to OW impact on mean mass at the end of the non-breeding season (M1). The regression tree is shown on the left-hand side and relative variable importance on the right-hand side.

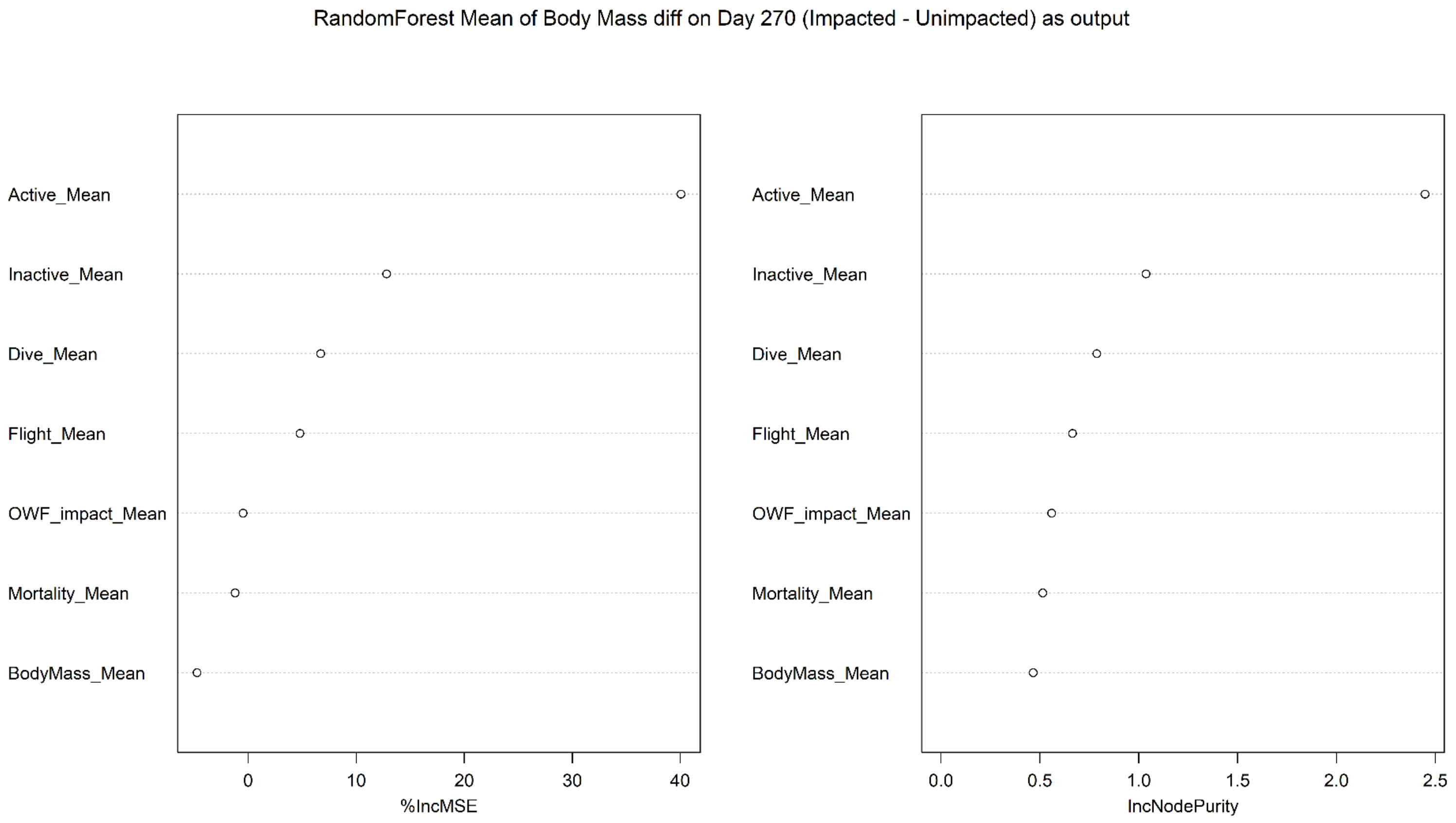


Figure 3. Variable importance within the DisNBS model based on applying random forests to OW impact on mean mass at the end of the non-breeding season (M1), as summarized using two different methods: %IncMSE (Percentage Increase in Mean Squared Error) and IncNodePurity (Increase in Node Purity).

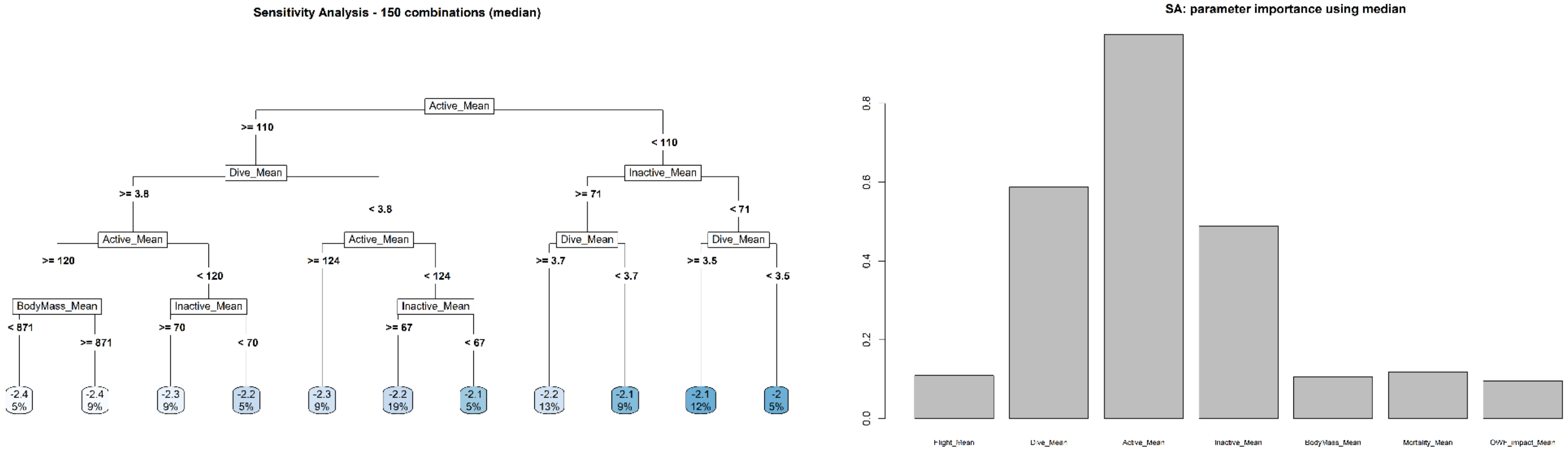


Figure 4. Variable importance within the DisNBS model based on applying regression trees to OW impact on median mass difference at the end of the non-breeding season (M2). The regression tree is shown on the left hand side and relative variable importance on the right hand side.

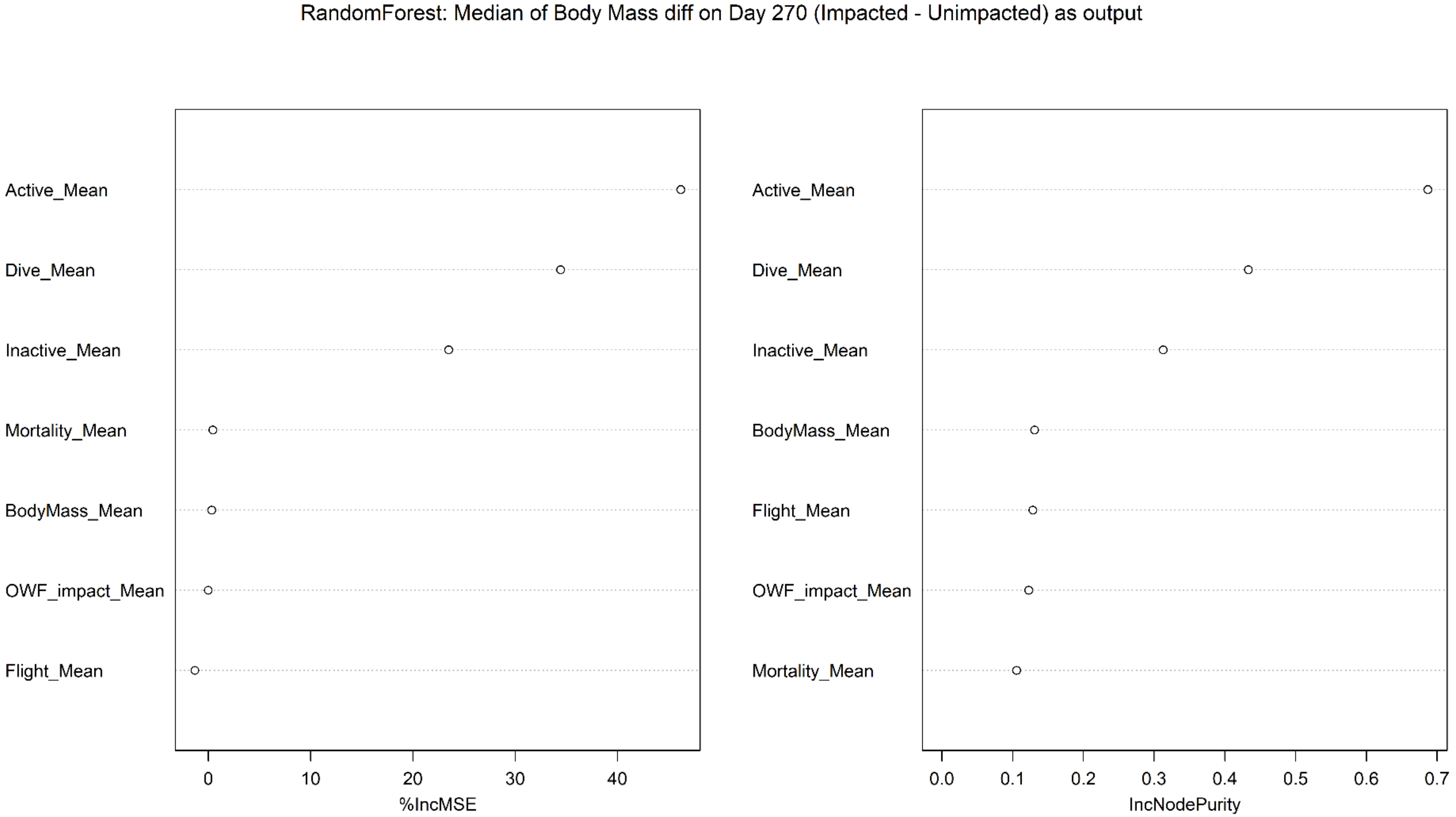


Figure 5. Variable importance within the DisNBS model based on applying random forests to OW impact on median body mass difference at the end of the non-breeding season (M2), as summarized using two different methods: %IncMSE (Percentage Increase in Mean Squared Error) and IncNodePurity (Increase in Node Purity

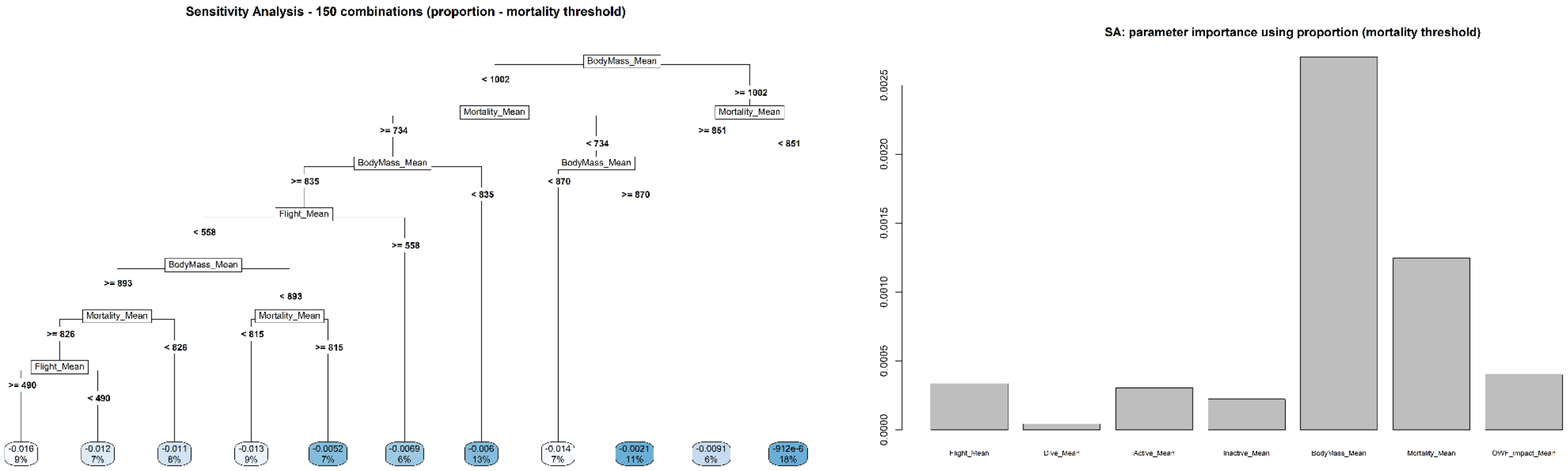


Figure 6. Variable importance within the DisNBS model based on applying regression trees to OW impact on difference in survival at the end of the non-breeding season (M3). The regression tree is shown on the left hand side and relative variable importance on the right hand side.

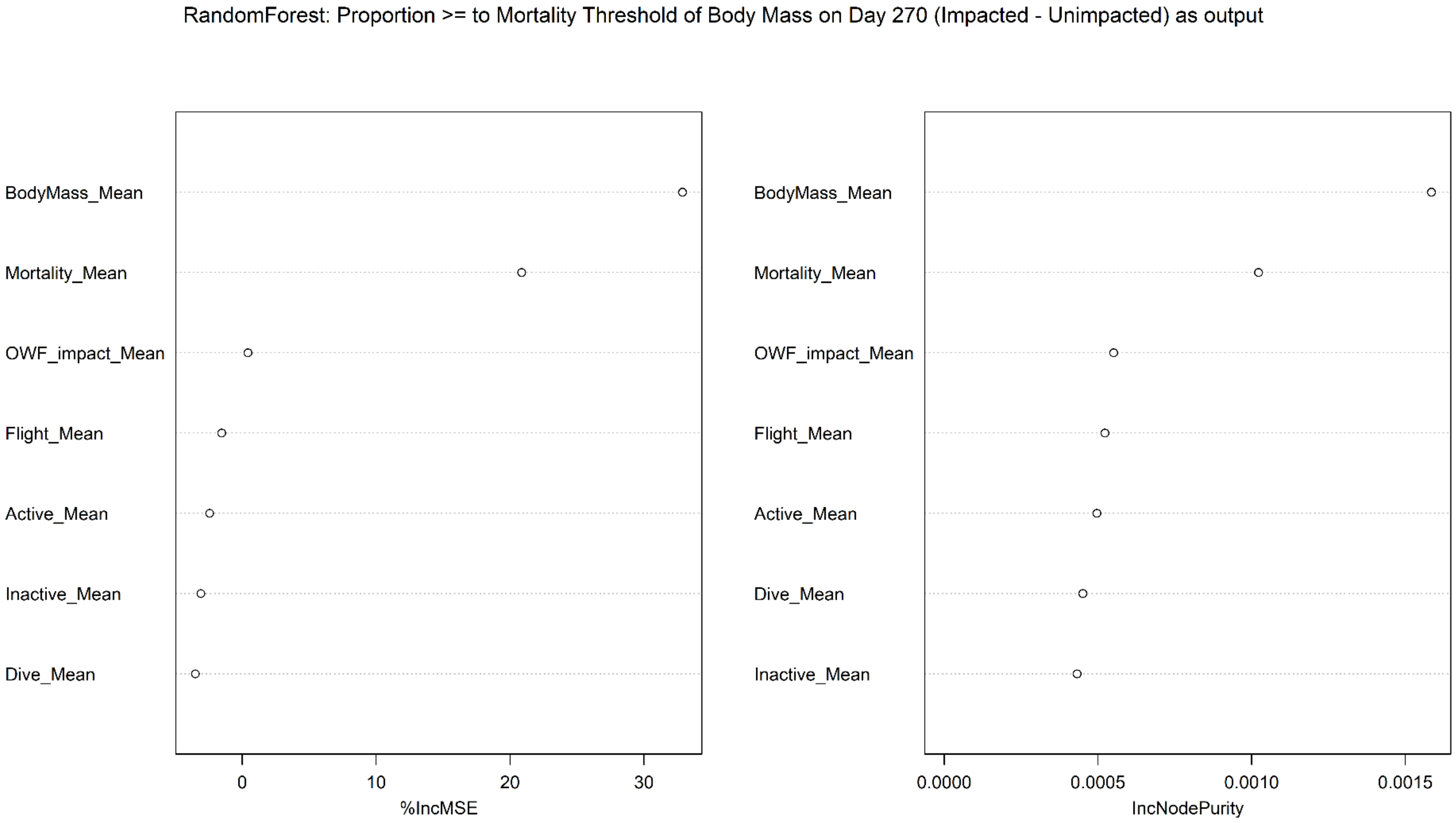


Figure 7. Variable importance within the DisNBS model based on applying random forests to OW impact on difference in survival at the end of the non-breeding season (M3), as summarized using two different methods: %IncMSE (Percentage Increase in Mean Squared Error) and IncNodePurity (Increase in Node Purity).

Table 4. Standardized variable importance scores of input parameters based on analysis of each of the three output metrics (M1, M2, M3) using two SA methods (regression trees and random forests). Scores are summarized to sum to one for each analysis.

Importance scores, standardised to sum to one						
Parameters	M1 (Mean)		M2 (Median)		M3 (Proportion Mortality)	
	Regression Tree	Random Forest	Regression Tree	Random Forest	Regression Tree	Random Forest
Active	0.51	0.38	0.39	0.36	0.06	0.09
Inactive	0.17	0.16	0.20	0.17	0.04	0.09
Dive	0.12	0.12	0.24	0.22	0.01	0.10
Flight	0.07	0.10	0.04	0.07	0.06	0.11
OWF Impact	0.05	0.09	0.04	0.06	0.08	0.11
Mortality	0.04	0.08	0.05	0.06	0.24	0.19
Body Mass	0.04	0.07	0.05	0.06	0.51	0.31

We summarize the overall rankings of parameters in Table 4. SA results consistently (whichever SA method was used) indicated highest sensitivity of the OW impact on mean adult mass at the end of the non-breeding season (M1) to the energy cost of "active" behaviour, followed by the energy cost of "inactive" behaviour and then the energy cost of "diving", with sensitivities to the remaining four parameters (mean initial mass, proportion of individuals susceptible to OW effects, energy cost of flying, mortality threshold) being substantially lower and similar to each other. SA results consisting of OW impact on *median* mass at the end of the non-breeding season (M2) were broadly similar, except that sensitivity to the energy cost of "diving" was now higher than sensitivity to the energy cost of "inactive" behaviour. In contrast, OW impact on mortality/survival at the end of the non-breeding season (M3) was, regardless of SA method, most sensitive to initial body mass, followed by the mortality threshold.

3.2.3. Robustness of results to number of agents

A key decision within our analysis was the number of agents per parameter combination, with the selected value (8000) chosen so as to balance precision in aggregate metrics (M1, M2, M3) against computational effort. In this section we investigate the evidence around the appropriateness of this value.

Ranking of parameter importance: In Appendix B we show the SA results for 4000 rather than 8000 agents, in order to evaluate whether the results are, by 8000 agents, relatively insensitive to the number of agents – the results indicate that the overall ranking of parameters is similar (and, crucially, that the ranking is identical for the three parameters with highest sensitivity for impact on mean/median mass and identical for the two parameters with highest sensitivity for impact on survival), and the qualitative conclusions are therefore the same, whether 4000 or 8000 agents is used, giving reassurance that the results obtained with 8000 agents are likely to be stable (i.e. that we would have obtained similar results by using an even higher number of agents per parameter combination).

R-squared values: R-squared values show the percentage of variance explained, which provides an indication of the adequacy of models that underpin the sensitivity analysis. Within this context, lack of fit could either arise from (a) unexplained variation that arises as a result of stochasticity (e.g. because we are using a stochastic model with a finite number of agents, but treating the model outputs as though the model were deterministic) and (b) lack of flexibility in the SA methods to capture the impact of parameters on the output metrics. If lack of fit is arising primarily from (a) then the R-squared value should increase as the number of agents increases, but this is not necessarily the case for lack of fit arising from (b). In Table 4 we compare the R-squared values associated with random forests models for each of the three output metrics, using either 8000 agents or 4000 agents.

The results with 8000 agents indicate that:

- R-squared values for OW impacts on median mass remain substantially higher than those for OW impacts on mean mass, presumably reflecting that the fact that median mass is less influenced by outlying or extreme values than mean mass
- R-squared values for OW impacts on mean mass are, in turn, substantially higher than those for OW impacts on survival. This may reflect the fact that the thresholding involved in calculating survival (by classing each individual in a binary way based on mass: survived or died) introduces additional stochasticity into the model outputs.

The R-squared values that arise when using 4000, rather than 8000, agents are slightly lower for OW impact on median mass and OW impact on survival, but much lower for OW impact on mean mass, suggesting that the results obtained when looking at OW impacts on mean mass with 4000 agents may be particularly sensitive to the effects of outlying or extreme values.

Table 4. R-squared values for random forest models of each of the three output metrics, using either 4000 or 8000 agents.

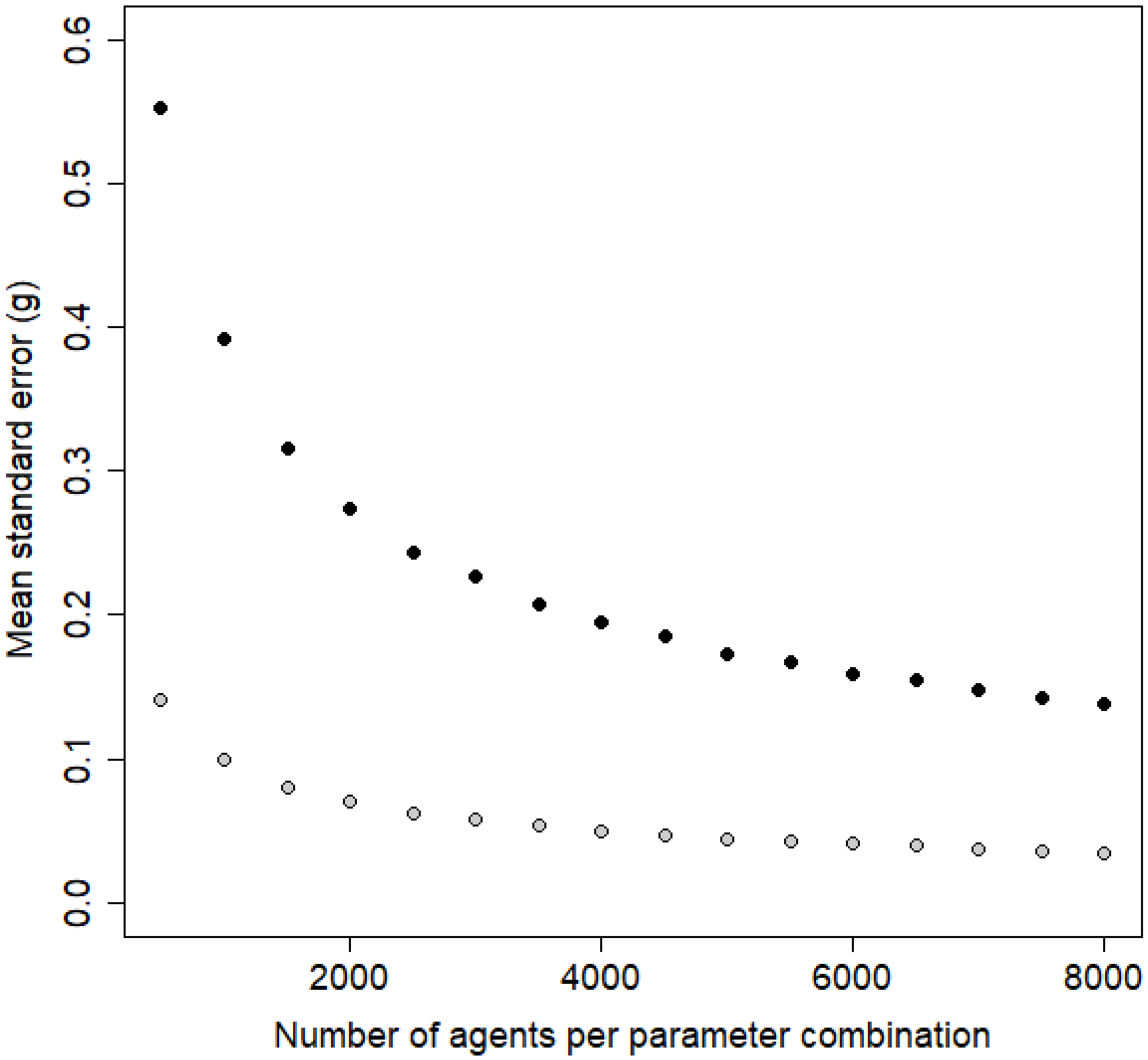
R-squared values	8000 agents	4000 agents
OW impact on mean mass	47.4%	28.7%
OW impact on median mass	71.1%	68.0%
OW impact on survival	34.0%	32.0%

Bootstrap standard errors: In Figure 8, we show how the bootstrap standard errors around M1 and M2 changed as the number of agents per parameter combination changes. With 8000 agents the standard errors are X and X respectively. The standard errors for 8000 agents are of very small magnitude relative to the variation between individual agents, but remain moderately large relative to the variation in mean values of M1 and M2 between parameter combinations.

However, it is clear from the figure that the reduction in standard error from adding additional agents is relatively slow by the time 8000 agents are being simulated, so a very large increase in the number of agents above this would be needed in order to yield substantially smaller standard errors.

Conclusions: The bootstrap sample errors and R-squared values indicate that a very large number of agents (much larger than the maximum value of 8000 considered here) may be needed in order to yield stable estimates for differences between parameter combinations, because the inter-agent variability in outputs is much larger than the variability in mean outputs between parameter combinations. However, the key qualitative results of the SA, in relation to the relative importance of different parameters, emerge strongly from the SA, and appear, based on the limited investigation that was feasible here, to be robust to the number of agents.

FIGURE 8. Bootstrap standard errors (mean values, averaged across parameter combinations) of M1 (black) and M2 (grey), for differing numbers of agents per parameter combination from 500 to 8000 in steps of 500.



4. Conclusions

Published evidence around uncertainty/variability in parameter values was in relation to inter-individual variability, rather than population-level uncertainty, so the UA focused on summarizing (using the same model runs as the SA) overall levels of variation across agents and across parameter combinations. Differences in final (Day 270) mass between impacted and baseline scenarios showed substantial variation between agents. Aggregated outputs (mean and median differences in final mass averaged across agents, M1 and M2) showed much lower levels of variation between parameter combinations, even though parameter values were considered to lie within a relatively wide range (-20% to 20% of case study parameter values for all parameters except the proportion of birds susceptible to windfarm effects which was assumed to be anywhere between 0 and 1). Differences (windfarm minus baseline) in final mass for individual agents ranged from below -140g to above +140g. Mean differences (M1) were consistently negative and ranged, across the 150 parameter combinations considered, from -3g to -1.9g, whilst median differences (M2) had an even narrower range.

Exploratory sensitivity analysis runs, using smaller numbers of agents than those used for the final SA, suggested that outputs had relatively low levels of sensitivity to parameters relating to initial time budget, so these parameters were not considered in the final SA. This is also expected at a theoretical level given the IBM design – the agents are dynamic in their daily activity budgets based on energy requirements i.e. the activity budget can rapidly change based on need. For the final SA seven parameters were therefore considered (mean energy costs of four activities, proportion of individuals sensitive to OWF effects, mean initial mass, and mortality threshold). 150 parameter combinations with 8000 agents per combination (a total of 1.2 million simulated agents), were used within the SA, with the model runs implemented using a high-performance computing cluster. Two sensitivity analysis methods – random forests and regression trees – were considered, and the analyses were run using either 4000 or 8000 agents (to ensure that the results were not sensitive to the number of agents).SA results consistently (whichever SA method was used) indicated highest sensitivity of the OW impact on mean adult mass (M1) and median adult mass (M2) at the end of the non-breeding season to the energy cost of “active” behaviour, followed by the energy costs of “inactive” and “diving” behaviours (with higher sensitivity to “inactive” than “diving” for mean adult mass, and the opposite for median adult mass). Overall these findings are logical since for a model with a bioenergetic calculation at its heart, then increasing the energy requirements of the agents will necessarily lead to greater demand for food from the supporting environment. If this requirement cannot be met then mass loss ensues. While the cost of “active” behaviour is lower than that of “flight”, guillemots spend so little time in flight during the non-breeding season (Buckingham et al., 2023), that the activity using, and so requiring, the greatest energy input is “active” behaviour. This is followed by the other behaviours in broadly descending order of largest daily energy use and input.

The sensitivity of the DisNBS model to activity energetic costs has implications for extension of the model to red-throated divers and/or other species that are less data rich. If the model is sensitive to understanding rates of energy expenditure, then this puts pressure on the accuracy of estimates of these rates. While laboratory estimates for these exist for guillemots (Croll & McLaren, 1993), there are no equivalent values for red-throated divers, with rates at present estimated from allometric equations (Dunn et al., 2023. While uncertainty in these estimates could be calculated, this does highlight the value of empirical data for input into DisNBS and similar models.

The finding that the OW impact on mean or median adult mass is not particularly sensitive to the proportion of individuals susceptible to OW effects may appear counter-intuitive but is explained by two IBM fundamentals. Firstly, and mostly importantly, the parameter relating to "proportion of individuals susceptible to OW effects" relates only to the colony of interest (meaning all individuals from other colonies are displaced from footprints within simulations, thereby increasing densities in surrounding areas). Conspecifics are not simulated in detail within the IBM so not controlled at a granular level, being limited to responsive to OWF or not (conceptually proportions of 1 or 0). Given the practical impact of OWF on the target population is largely due to density dependency, which includes conspecifics, the repulsion of the target population from the small area of OWF has a relatively small contribution to density dependence/competition. As a result, the impact of the OWFs upon levels of competition suffered by the focal colony within the model may be fairly insensitive to this parameter -- because it relates specifically to only the level of displacement experienced by the focal colony, and therefore does not affect the potential increase in competition arising from birds from other colonies that were subjected to a 100% displacement rate (and the number of birds arising from non-focal colonies will be far larger than the number of birds from the focal colony). Secondly, the IBM has been designed (WP1) for the agents to behave in response to energy requirements – put simply, poorer feeding conditions leads to more feeding activity, subject to some constraints e.g. day-length. Consequently, small increases in competition are relatively easily compensated for with more feeding and less inactivity given that initially the birds are relatively inactive (Table 1). An IBM without this responsive behaviour would be unduly naïve – these IBM design features and others are detailed in the WP2 documentation.

SA results (whichever SA method was used) indicated highest sensitivity of the OW impact on adult survival to initial adult mass (M3), followed by mortality threshold, with substantially lower sensitivity to the remaining five parameters. This is also logical since mortality is directly calculated, for both baseline and impacted scenarios, by determining whether final mass (which will be dependent on initial mass) is below the mortality threshold. Again, this has consequences for less data-rich species where there is less reliable data on key inputs such as initial body mass, while mortality threshold may be little more than an informed guess.

A key choice within our analysis was the number of agents simulated per parameter combination, which represents a balance between the precision (and hence stability) of aggregate metrics of interest (such as M1, M2 and M3) and computational effort. We investigated robustness of our results to the choice of this value in a number of different ways. Bootstrap sample errors and R-squared values of SA models indicate that a very large number of agents (much larger than the maximum value of 8000 considered here) may be needed in order to yield stable estimates for detailed differences between parameter combinations, because the inter-agent variability in outputs is much larger than the variability in mean outputs between parameter combinations. The reason that a relatively large number of agents is needed in order to produce stable results in key outputs of interest is because these outputs focus on relatively subtle differences between windfarm and baseline conditions – these differences are typically much smaller than the assumed levels of baseline inter-individual variation, which means that a substantial number of agents is required in order to estimate these differences with a reasonable level of precision. However, the key qualitative results of the SA, in relation to the relative importance of different parameters, emerge strongly from the SA, and appear, based on the limited investigation that was feasible here, to be robust to the number of agents, suggesting that the key findings of the SA obtained here would still be likely to hold if a larger number of agents were used.

5. Recommendations for future work

5.1 Further research to address evidence gaps highlighted by SA

The SA indicated that key DisNBS model outputs relevant to OW effects were sensitive to estimates of energy costs of activities (active, inactive, diving), initial mass and mortality threshold. The outputs considered related to direct OW effects on mortality and to OW effects on mass change (which provide a proxy for OW effects on productivity).

Energetics equations: The IBM is at its heart an energetics model, and the energetics equations associated with behavioural states and the environmental conditions are clearly key components. Published equations were used here which were simple in form, combined independently within days, and with crude measures of uncertainty. The simplicity of the equations is likely justified, but the uncertainty warrants further consideration. For example, without constraints imposed, the energetics components in isolation, and particularly combined, permit nonsensical values with non-negligible probability. An integrated model with uncertainty would provide more stable results. Further, a key design element of the IBM is the rebalancing of states in response to the current energy intake e.g. more foraging to compensate for low energy intake. This is not well understood and educated guesses are implemented here e.g. rebalancing is a function of the previous day’s success.

Understanding more about the energy costs of free-ranging non-breeding seabirds is clearly critical to reducing uncertainty in model predictions and hence reducing consenting risk. This is particularly important for activities that constitute a large proportion of daily energy use and intake, through the combination of high rates and/or high time investment. While common guillemots are relatively well studied, the current estimates of energetic costs come from studies conducted in the laboratory on captive birds over 30 years ago (Croll & McLaren, 1993). Furthermore, the assignment of time to behaviour while using more up to date tools, is still subject to considerable uncertainty as data from biologgers is interpreted without validation. Improved biologging devices are becoming available all the time and can be used to improve time budgets as well as to measure key environmental parameters experienced by the birds relevant to energy cost estimation (e.g. SST). Data from the current Aukestra project will provide additional information for guillemots and razorbills, but data gaps will remain for other species (such as red throated diver). Other approaches can be used to estimate energetic costs directly in free-ranging birds, such as the use of accelerometry or in particular heart rate, which is well suited for the non-breeding season (Green, 2011). Detailed biophysical modelling might also allow for improved estimates of energetic costs for non-breeding seabirds, and such approaches have been applied to guillemots (e.g. Clairbaux et al., 2021). Future work could investigate the combined effects of environmental conditions on energy costs and time budgets on non-breeding seabirds via this approach. As noted above, this will be particularly important for species where there is not any equivalent laboratory data (e.g. red-throated diver) for which a combination of allometric data for Basal Metabolic Rate (BMR; again primarily based on old, captive and laboratory studies) and assumptions about BMR-multipliers for activity are currently used to estimate activity costs (Dunn et al., 2023). Laboratory studies of energetic costs could also provide valuable information for unstudied species, and improve allometric models.

Mortality thresholds and initial body mass: Understanding mortality thresholds is more challenging since gold-standard empirical data would likely involve forced starvation of healthy birds, an experiment unlikely to pass ethical review or regulations. That said, current understanding of mortality thresholds comes from incidental interpretation rather than directed data collection. It may be that more comprehensive weighing of birds captured at sea during non-breeding periods, found stranded on land, accidentally entangled dead or alive in fishing gear or found recently dead on land could all improve understanding in this area. It was clear from the development of the model and the process to estimate starting parameters that data and understanding of this key aspect of seabird biology were extremely limited. It is also unclear, in the absence of evidence empirical, whether the mortality threshold should be specified in relation to absolute mass or relative mass change. Initial body mass would be far easier to quantify with existing methods, with a lack of data rather than a lack of methodologies being the limitation here.

OWF redistribution: The SA results did not indicate high sensitivity to the proportion of individuals susceptible to displacement, but this result should be interpreted carefully, since this parameter relates to individuals from the colony of interest. The low sensitivity may arise from two reasons previously indicated (section 4): because OW effects on competition are dominated by OW effects on individuals from other colonies, which are not determined by this parameter, and because feeding behaviour is dynamic to match energy intake, which dampens counterfactual differences by design. In addition, the redistribution function here was a simple exclusion as it is not known in greater detail, although the IBM is very general in how these can be parameterised. By extension, parameterisation for the various conspecific populations could be very detailed, and potentially influential, but not able to be implemented here. Future research around the frequency, scale and form of displacement is therefore still likely to be key in informing the use and future development of the DisNBS model.

5.2 Further work on uncertainty and sensitivity analysis

The uncertainty and sensitivity analyses that we have conducted here have necessarily been relatively limited, due to the tight timelines of the project and the inherently computationally intensive nature of the individual-based models. A number of key strands of further work, to extend and refine the sensitivity and uncertainty analysis, would, however, be valuable.

Application of alternative sensitivity analysis approaches, such as Sobol indices, that have potential to provide outputs that more informative and interpretable than those derived using regression trees and random forests. These approaches depend upon being able to specify the number of model runs that are required in advance, and so were not viable for this study, but may be feasible to use in future work. Sensitivity analysis methods that are designed to be applied directly to stochastic model outputs (Baker et al., 2022) would also be worth investigating, since this would avoid the need to, as in this study, average across agents in order to provide outputs that can be regarded as approximately deterministic.

Quantifying the impacts of uncertainty in population-level parameters, and including these within the uncertainty analysis – the focus here has been on sensitivity to, and inter-individual variability in, the model parameters, but it is also important to consider population-level uncertainty within these parameters. We have not attempted this here because of the lack of available published information on levels of uncertainty (rather than variability) for these parameters. Expert elicitation would provide a formal mechanism to quantify the uncertainty in these parameters from expert judgement where this is possible based on the current evidence base, but it is important to note that it is not a substitute for empirical data collection and will only be appropriate for those parameters for which relevant evidence exists.

Consider uncertainty and sensitivity in relation to spatial and spatio-temporal forcing variables that are inputs to the model, such as bird distribution maps, competition maps and sea surface temperature maps. The bird distribution and competition maps for guillemot that have been used here are derived from the GLS-based colony-specific distribution maps developed by Buckingham (2022). These were extended within the ORJIP AppSaS project (Butler et al., 2024) to account for the uncertainty associated with converting each GLS light level record into an estimated location (a key source of uncertainty within GLS tracking data) via a simulated-based approach, and developed into an R tool (the ANBS tool; Jones et al., 2024). By using these simulated outputs we could account for uncertainty within two of the key model inputs (bird distribution, competition maps). The Aukestra project, which is currently underway, will produce refined versions of these monthly distribution maps, and associated uncertainties. Uncertainty in the maps produced by Aukestra is anticipated to be lower than in the current maps, because they will use additional tracking data and a wider range of auxiliary data, and will exploit developments in quantitative methodology Information on uncertainty in SST is also available, and it would be possible to account for this uncertainty within the model, although there may be a computational cost associated with this. There may also be advantages in using higher resolution data on SST within the model. Using the same source of SST data for GLS location estimates in the underlying utilisation distribution maps and in the IBM would be helpful in ensuring that differences in bias (from different SST sources) was accounted for

Quantifying sensitivity and uncertainty with the context of species and populations other than those considered here: this is dependent upon parameterizing the model for these species and populations, as shown in the WP2 documentation and case studies, where guillemot for Isle of May and red throated divers in the North Sea are parameterised. The IBM offers a framework and toolset for the relatively rapid development of IBMs for general species and scenarios, where a protracted bespoke programming exercise would typically be required. However, this does not obviate the need for sensible parameter inputs and the detailed specification of certain functions - energetics equations being a fundamental example, albeit not necessarily complex in form. Many of sensitivities shown here are likely to transfer to other species given they relate to these fundamental energetics equations. Others, such as the OWF impact however, would be expected to differ given the dependency here on conspecifics. The red throated diver study is a case in point, where the WP2 case study has no conspecifics and a proportionally larger OWF displacement of the target population. The compensatory rebalancing of behaviours between counterfactuals is required to be larger, and tends to be less effective, as a consequence.

References

Baker, E. Barbillon, P., Fadikar, A., Gramacy, R.B., Herbei, R., Higdon, D., Huang, J., Johnson, L.R., Ma, P., Mondal, A., Pires, B., Sacks, J. & Sokolov, V. (2022) Analyzing Stochastic Computer Models: A Review with Opportunities," Statistical Science, Statist. Sci. 37(1), 64-89.

Beven, K. & Lamb, R., 2017. The uncertainty cascade in model fusion. In: *Integrated environmental modelling to solve real world problems: methods, vision and challenges.* Riddick, A. T., Kessler, H. & Giles, J. R. A. (eds.). London: Geological Society of London, p. 255-266 12 p. (Special Publications; vol. 408).

Breiman, L. 2001. Random forests. Machine learning 45, 5–32.

Breiman, L., Friedman, J., Olshen, R., and Stone, C. 1983. Classification and regression trees. Wadsworth.

Buckingham, L., 2022. Population variation in seabird non-breeding season ecology: implications for vulnerability assessments. Doctoral dissertation., University of Liverpool.

Buckingham, L., Daunt, F., Bogdanova, M.I., Furness, R.W., Bennett, S., Duckworth, J., Dunn, R.E., Wanless, S., Harris, M.P., Jardine, D.C., Newell, M.A., Ward, R.M., Weston, E.D. and Green, J.A. 2023, Energetic synchrony throughout the non-breeding season in common guillemots from four colonies. J Avian Biol, 2023: e03018. https://doi.org/10.1111/jav.03018

Butler, A., Buckingham, L., Thaxter, C., Jones, E., Whyte, Cook, A., Bogdanova, M., Ewing, D., Searle, K. and Daunt, F. 2024. AppSaS – Apportioning seabirds seen-at-sea - WP3 - Method evaluation. Offshore Renewables Joint Industry Programme (ORJIP) for Offshore Wind. https://ctprodstorageaccountp.blob.core.windows.net/prod-drupal-files/2024-03/ORJIP_AppSaS_WP3_Final.pdf

Carnell R (2024). *lhs: Latin Hypercube Samples*. R package version 1.2.0, https://CRAN.R-project.org/package=lhs.

Clairbaux, M., Mathewson, P., Porter, W., Fort, J., Strøm, H., Moe, B. et al. (2021) North Atlantic winter cyclones starve seabirds. Current Biology 31(17), 3964-3971.e3 doi: https://doi.org/10.1016/j.cub.2021.06.059

Critchley, E. J., & Buckingham, L. (2024). Offshore wind farms in the North Sea - a dataset of developments estimated to be operational by 2030 [Data set]. Zenodo. https://doi.org/10.5281/zenodo.10478448

Croll, D.A. McLaren, E. (1993) Diving metabolism and thermoregulation in common and thick-billed murres. Journal of Comparative Physiology 163, 160-166.

Drewitt, A. L., & Langston, R. H. W. 2006 Assessing the impacts of wind projects on birds. Ibis 148: S29–42 Dunn, RE, J Duckworth, JA. Green. 2023. A framework to unlock marine bird energetics. J Exp Biol 15; 226 (24): jeb246754. doi: https://doi.org/10.1242/jeb.246754

Dunn, R.E., Duckworth, J. & Green, J. A. (2023) A framework to unlock marine bird energetics. J Exp Biol 226 (24): jeb246754. doi: https://doi.org/10.1242/jeb.246754

Frey, C. & Summet, P.R. (2002) Identification and review of sensitivity analysis methods. Risk analysis, 22(3), 553-578. , Green, J.A. (2011) The heart rate method for estimating metabolic rate: Review and recommendations. Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology, 158(3), 287-304. doi: https://doi.org/10.1016/j.cbpa.2010.09.011.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2021). An introduction to statistical learning: With applications in R, 2nd Edition. Springer Nature.

Jones, E., Whyte, K., Mobbs, D., Searle, K., Butler, A., Couto, A., Cook, A., Thaxter, C., Buckingham, L., Bogdanova, M. and Daunt, F. 2024. AppSaS – Apportioning seabirds seen-at-sea - Apportioning in the non-breeding season tool (ANBS) - User Guide to the ANBS tool. Offshore Renewables Joint Industry Programme (ORJIP) for Offshore Wind. https://ctprodstorageaccountp.blob.core.windows.net/prod-drupal-files/2024-03/ORJIP_AppSaS_ANBS-User-Guide_2024_Final.pdf

Masden, E.A., Haydon, D.T., Fox, A.D. and Furness, R.W. 2010. Barriers to movement: Modelling energetic costs of avoiding marine wind farms among breeding seabirds. Marine Pollution Bulletin 60, 1085–1091.

Mullins, J., Ling, Y., Mahadevan, S., Sun, L. and Strachan, A. 2016. Separation of aleatory and epistemic uncertainty in probabilistic model validation, Reliability Engineering and System Safety, 147, 49-59.

Searle, K., Mobbs, D., Butler, A., Bogdanova, M., Freeman, S., Wanless, S. & Daunt, F. 2014. Population consequences of displacement from proposed offshore wind energy developments for seabirds breeding at Scottish SPAs (CR/2012/03). Report to Scottish Government.

Searle, K. R., Mobbs, D. C., Butler, A., Furness, R. W., Trinder, M. N., & Daunt, F. 2018. Finding out the fate of displaced birds. Scottish Marine and Freshwater Science, 9(8), 1-161. https://doi.org/10.7489/12118-1

Scottish Government. 2011. Habitats Regulations Appraisal of Draft Plan for Offshore Wind Energy in Scottish Territorial Waters:Appropriate Assessment Information Review (2011). http://www.scotland.gov.uk/Publications/2011/03/04165857/15

Appendix A. High-performance computing

The sensitivity analysis for simulations using the newly built R package “roamR” was carried out using the UK crop diversity high performance cluster. The package and its dependencies were installed and compiled using conda and mamba designed for use in a linux operating system. For some of the R package such as “spaths” the configuration file (Makeconf) within the R library folder needed to be modified with additional information regarding the version of C++. Once the packages were compiled the R script was tested using an interactive job (job here refers to a task in Slurm) option to rule out any warnings and errors. Once the script ran without any issue the simulations were deployed using batch files. One batch file contained a loop that sequentially called every parameter combination and the other batch file that was designed to simulate a fixed number of birds using the new R package “RoamR” as an array job. An array job is designed to run the same script multiple times, each time is called a single task. Every array job has a unique ID and generates an index number corresponding to the number of times the array job is called. In our case it was written to run the same R script 80 times for a single parameter combination,10 tasks at one time depending on the availability of the resources, each time simulating 100 agents.

The resources are made available via nodes that are specific to the type of job also referred as partition or queue within the batch script. Depending on the type of partition the job was allocated to nodes. The job types include short, medium, long, himem (high memory), hicpu and gpu mostly referring to the runtime. For example, a short partition allows allocation of jobs without waiting for long hours. If unspecified the default choice is medium.

In our case most of our batch script was run on nodes that were specific for a short job. The short partition job was allocated between 6 such nodes with a fixed number of CPU cores and memory named black, cordelia, buffy, carrow, greyblack and grindelwald, depending on their availability for usage among all HPC users. The allocating of jobs and resources was done using Slurm making sure all processors within clusters are used.

Appendix B. SA results with 4000 agents

Within Section 3.4 and Figures 1-6 we present SA results using 8000 agents. Here, we present the corresponding results based on 4000 agents.

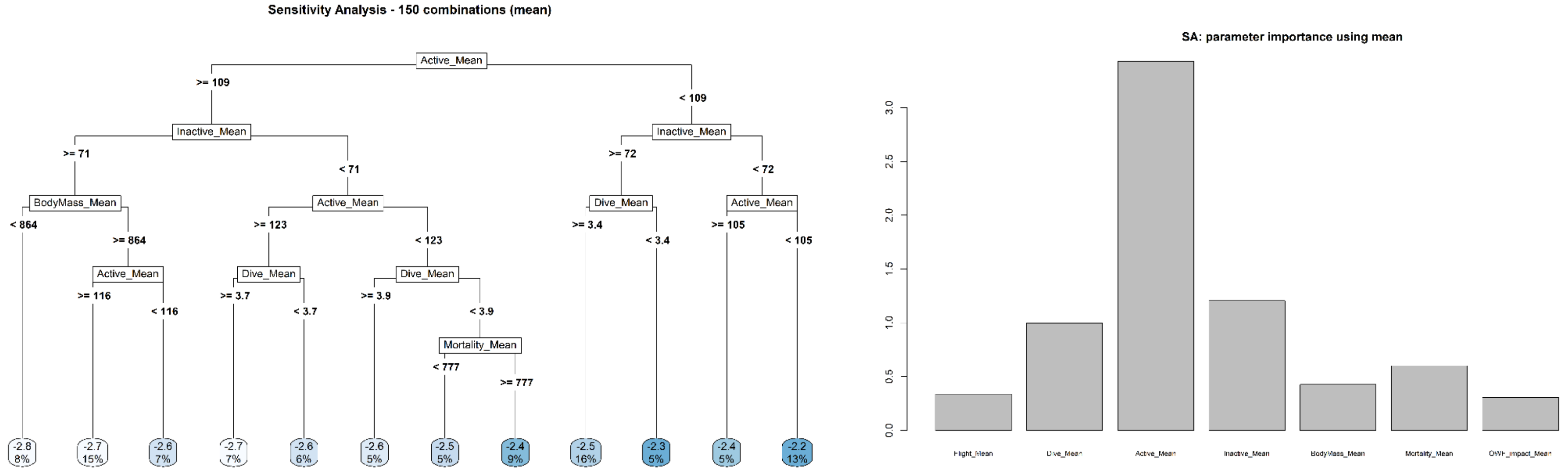


Figure B1. As Figure 2 but based on running SA using 4000 rather than 8000 agents.

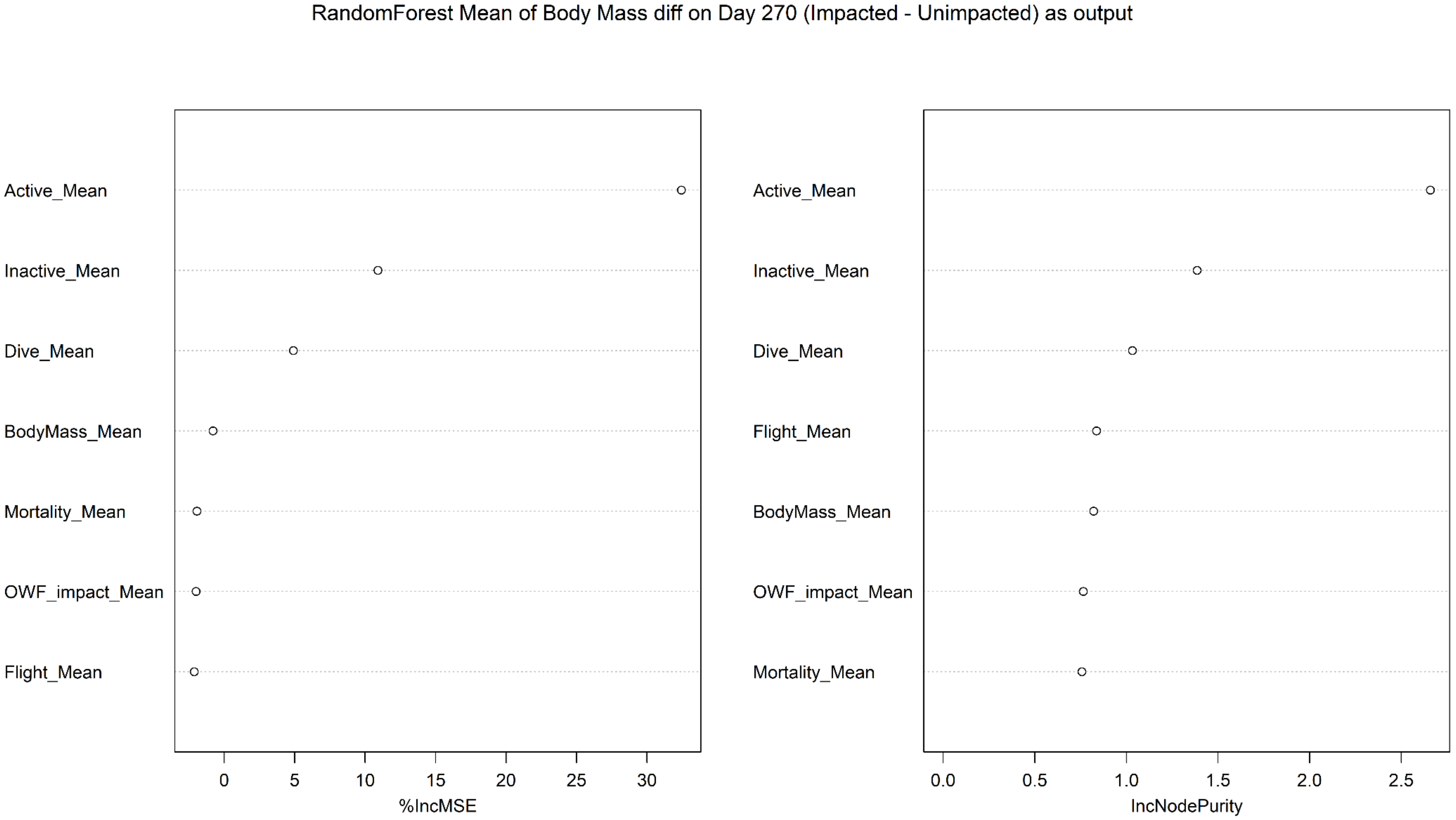
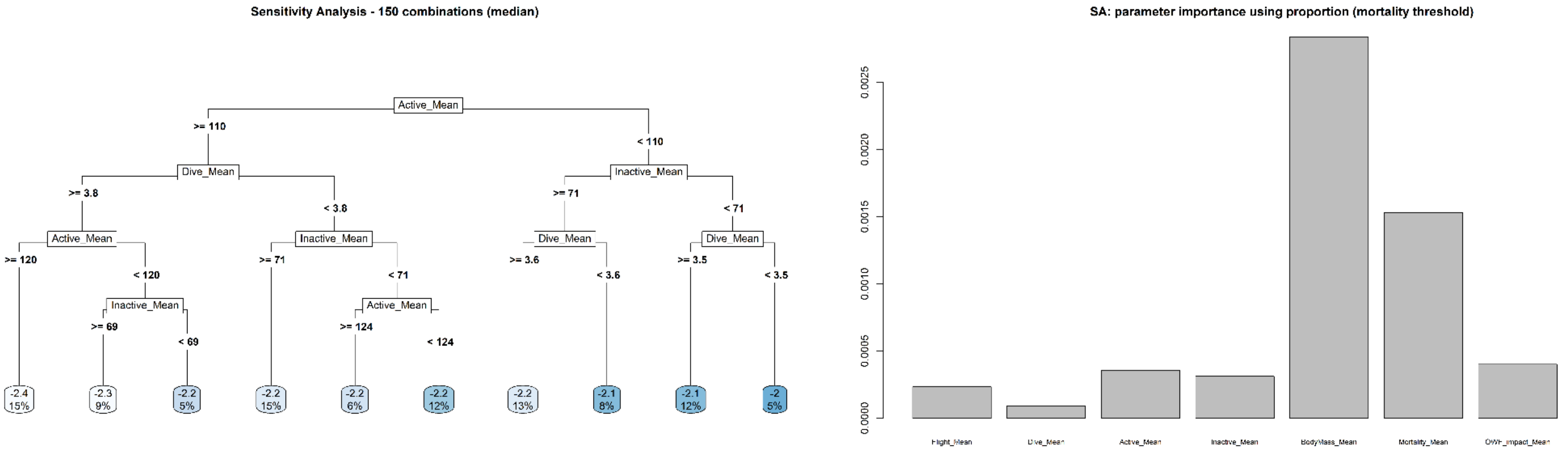


Figure B2. As Figure 3 but based on running SA using 4000 rather than 8000 agents.


Sensitivity Analysis - 150 combinations (median)

SA: parameter importance using proportion (mortality threshold)

Figure B3. As Figure 4 but based on running SA using 4000 rather than 8000 agents.

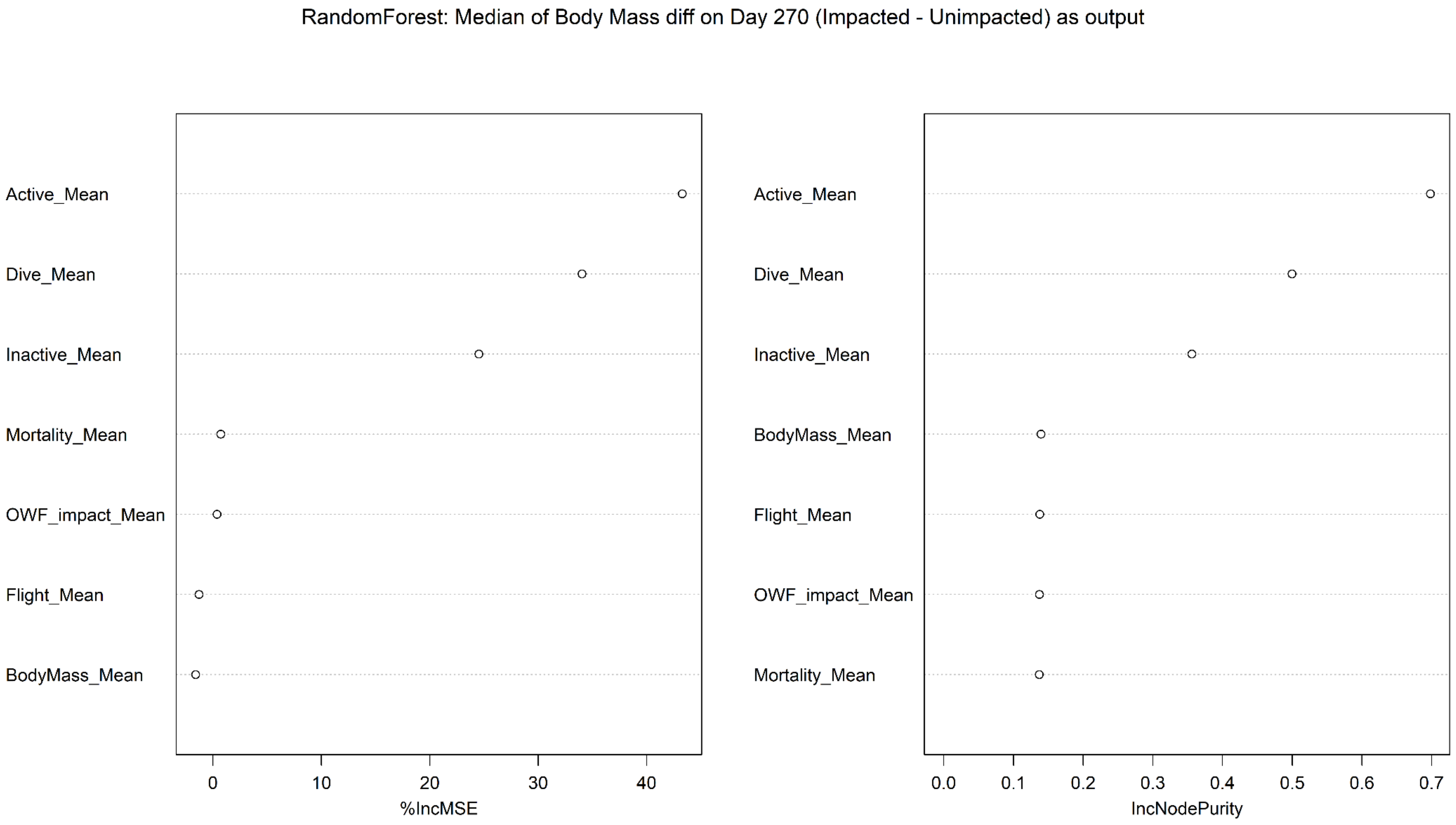
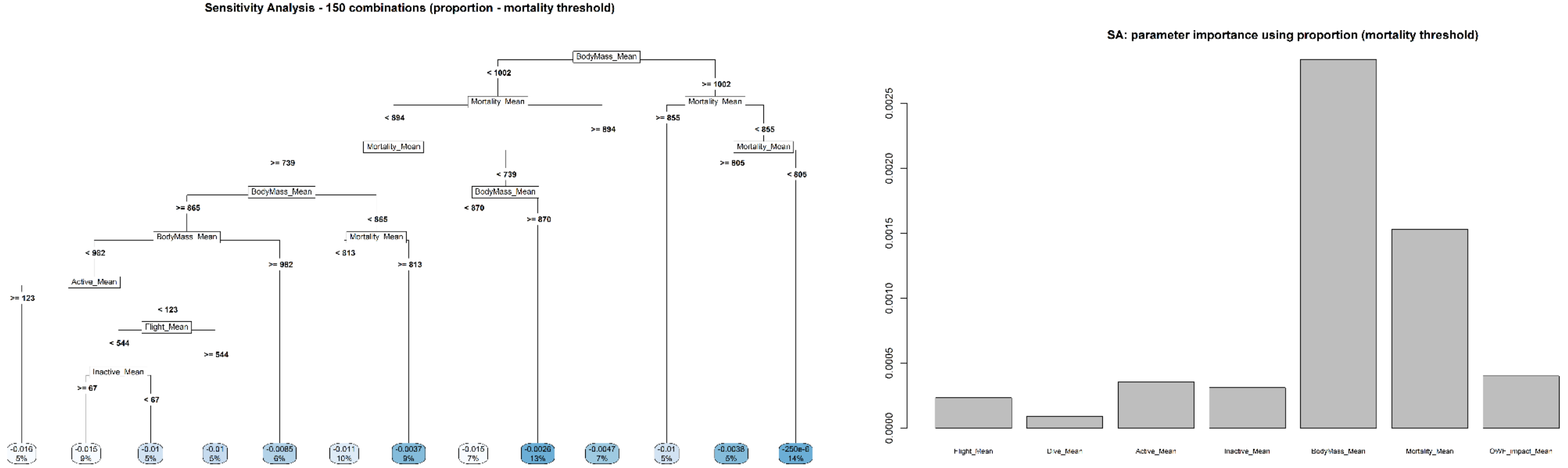


Figure B4. As Figure 5 but based on running SA using 4000 rather than 8000 agents.



Sensitivity Analysis - 150 combinations (proportion - mortality threshold)

SA: parameter importance using proportion (mortality threshold)

Figure B5. As Figure 6 but based on running SA using 4000 rather than 8000 agents.

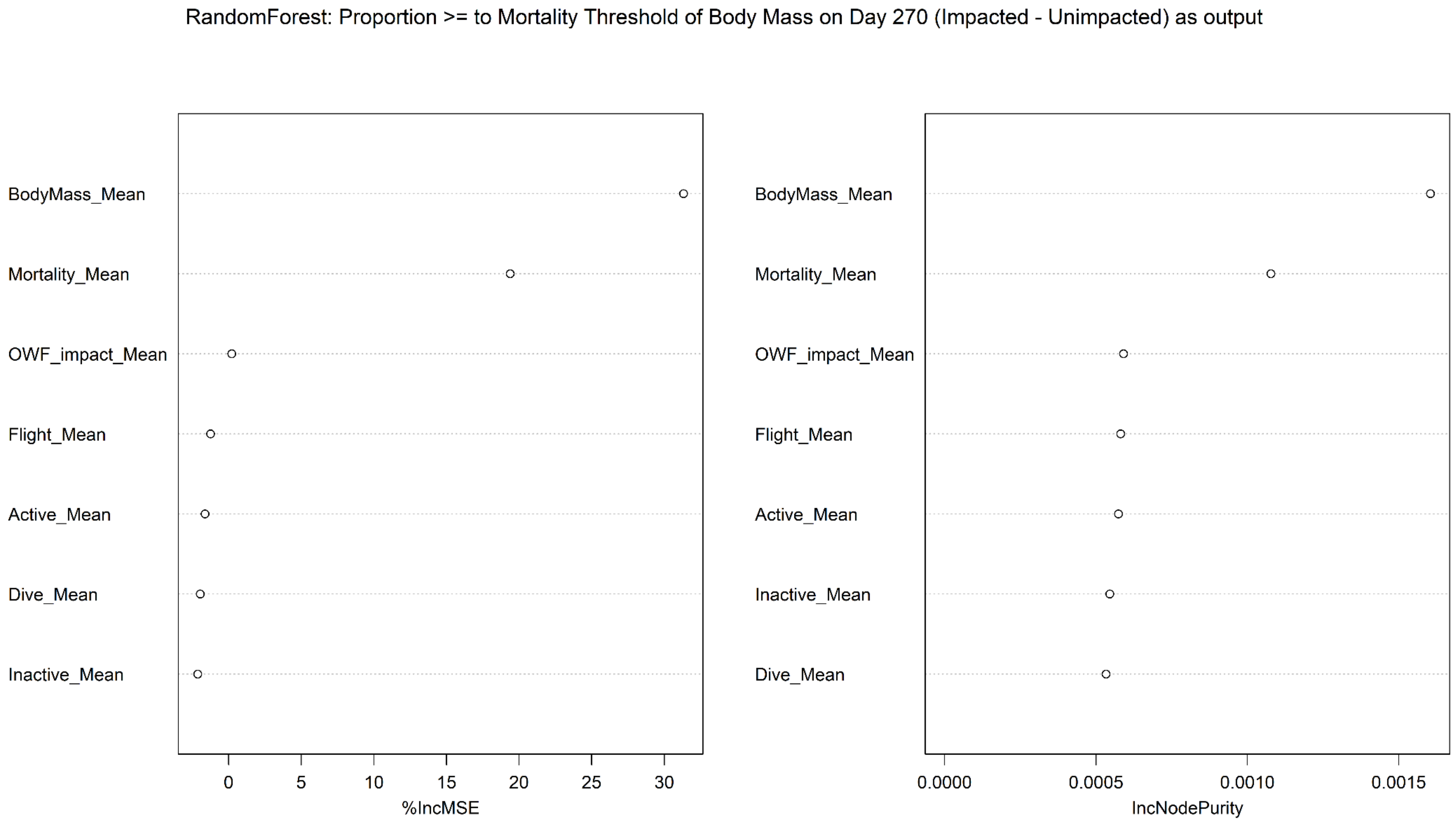

RandomForest: Proportion >= to Mortality Threshold of Body Mass on Day 270 (Impacted - Unimpacted) as output

Figure B6. As Figure 7 but based on running SA using 4000 rather than 8000 agents.

carbontrust.com

+44 (0) 20 7170 7000

The Carbon Trust is a company limited by guarantee and registered in England and Wales under Company number 4190230 with its Registered Office at: Level 5, Arbor, 255 Blackfriars Road, London, SE1 9AX.

Published in the UK: 2026