OFFSHORE RENEWABLES JOINT INDUSTRY PROGRAMME (ORJIP) FOR OFFSHORE WIND

Quantification of mortality rates associated with displacement (QuMR)

Appendix – C

May 2025

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

This document was produced on behalf of ORJIP Offshore Wind by UK Centre for Ecology & Hydrology and BioSS. The report was authored by Dr. Kate R Searle. Professor Francis Daunt, Dr. Christopher John Pollock, and Dr Adam Butler.

The project was advised by the ORJIP Offshore Wind Steering Group and the QuMR 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:

- Joint Nature Conservation Committee (JNCC)
- Natural England
- Natural Resources Wales
- NatureScot
- Scottish Government's Marine Directorate

This report was sponsored by the ORJIP Offshore Wind programme. For the avoidance of doubt, this report expresses the independent views of the authors.

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

Contents

List of Tables

List of Figures

Abbreviations

Term	Description
ABC	Approximate Bayesian Computation
AIC	Akaike Information Criterion
ANOVA	ANALYSIS OF VARIANCE
BACI	Before After Control Impact
BDMPS	Biologically Defined Minimum Population Scales
BTO	British Trust for Ornithology
CEH	Centre for Ecology & Hydrology
DISNBS	Displacement in the non-breeding season
EE	Expert Elicitation
EIA	Environmental Impact Assessment
GLM	Generalized Linear Model
GIS	Geographic Information System
GPS	Global Positioning System
GU	Common Guillemot
KI	Black-legged Kittiwake
NERC	Natural Environment Research Council
ORD	Offshore renewable development
ORJIP	Offshore Renewables Joint Industry Programme
OWEZ	Offshore Windpark Egmond aan Zee
OWEC	Offshore Wind Evidence and Change
OWF	Offshore wind farm
PAWP	Prinses Amalia Wind Park
PBR	Potential Biological Removal
PINS	Planning Inspectorate
PVA	Population Viability Analyses
RSPB	Royal Society for the Protection of Birds
SEANSE	Strategic Environmental Assessment North Seas Energy
SNCB	Statutory Nature Conservation Bodies
TDR	Time Depth Recorders
WP	Work Package

1. Introduction

One way to try to quantify displacement mortality, and sources of variation in mortality, is via a mechanistic model. Simulations from the mechanistic model can be used to estimate the mortality associated with different scenarios, and these results can then be used (WP4) in estimating displacement mortality rates.

We focus here upon using SeabORD, an individual-based model of seabird behaviour, energetics, demography and windfarm interactions, to estimate the displacement mortality rates associated with different colonies under different wind farm scenarios. SeabORD includes two main mechanisms of windfarm interaction: displacement effects (a switch in foraging location as a result of a windfarm) and barrier effects (increased flight distances, and hence energetic costs, to reach foraging locations when birds avoid flying over windfarms). We refer to the combination of both of these as “displacement mortality”. SeabORD can consider either a single windfarm, or the simultaneous effects of multiple windfarms, and considers impacts on both chicks and adults. SeabORD directly simulates the impact of windfarms on chick mortality. For adults, SeabORD simulates the impact on windfarms on the change in body mass over the course of the breeding season and then translates this into impacts on annual adult survival via published mass-survival relationships. SeabORD only considers the impacts of windfarm interactions during the chick rearing period within the breeding season and is parameterized for four species: black-legged kittiwake, common guillemot, razorbill and Atlantic puffin.

The biological parameters within SeabORD are largely fixed based on expert judgement or published values from the literature (Searle et al., 2018). However, the remaining inputs to SeabORD - bird distributions maps, colony sizes, and prey maps – need to be specified separately for each colony. In addition, there is one input parameter for SeabORD – total prey – that cannot meaningfully be derived from expert judgement or the published literature. SeabORD outputs are highly sensitive to the values of this parameter, so it is calibrated separately for each colony by selecting the parameter values for “total prey” that lead baseline adult and chick survival rates (in the absence of a windfarm) to lie within biologically plausible ranges. Because the values of this parameter are calibrated against a range of baseline survival rates, a range of plausible parameter values are identified for each colony. Parameters from across this range are then used when simulating windfarm impacts via SeabORD, allowing this key source of uncertainty to be accounted for within the SeabORD outputs.

SeabORD simulates changes in adult and chick mortality (and survival) rates, and in a range of other quantities: adult mass loss during the chick rearing period is one of the key outputs, because the adult mortality and survival rates are derived from this using published relationships. In principle, we could use SeabORD to provide model-based estimates of displacement mortality for a wide range of colonies and windfarm scenarios, and (via the work in WP4) thereby build up information on typical mortality rates, and on the variability in these rates between colonies and windfarms. In practice, SeabORD is a computationally intensive model to run, so there would be substantial practical challenges in doing this. The recoding of SeabORD into R as part of the Marine Scotland CEF project, which we exploit here, has involved improvements to computational efficiency, but realistically complicated runs of SeabORD continue to require large amounts of computer time. In addition, the calibration process within SeabORD has always required human input, and this is required each time the model is run for a new colony for which it has not previously been calibrated. Work within the CEF project to try to automate the calibration process revealed additional challenges in doing this – the work aiming to find simple proxies of SeabORD outputs that could be used to predict whether SeabORD would produce broadly plausible baseline demographic rates for a particular total prey level, but the results suggest that the obvious choices for such proxies did not perform well in predicting whether SeabORD would produce plausible baseline demography. For the moment, SeabORD therefore continues to require a manual calibration step for each colony.

Emulation provides a framework for using statistical models to approximate mechanistic models. It uses a "training set" of mechanistic model inputs and outputs to build a general model for the relationship between the mechanistic model inputs and outputs, and, as such, provides an approximation to the mechanistic model. Emulation is typically designed to approximate computationally intensive models using much less computationally intensive models, so that the emulator can then be used, predictively, as a rapid but approximate substitute for the mechanistic model. As SeabORD is a computationally intensive model it is an obvious candidate to be emulated. The types of models used for emulation are generally similar to the models that could be used to model relationships in empirical data – for example, multiple regression, mixed models, Gaussian processes, random forests and neutral networks. Within the context of emulation, the "response variables" are the mechanistic model outputs (or a subset of these), and the "explanatory variables" are the mechanistic model inputs (or a subset of these).

Emulation is not a replacement for mechanistic modelling: the emulator is designed to provide a more rapid alternative to the mechanistic model, but at the cost of some loss of accuracy caused by using the emulator to approximate the mechanistic model. This accuracy can be increased by using a large training set (i.e., more mechanistic model runs), but since the rationale for using the emulator is to reduce computational effort, there is a trade-off: more mechanistic model runs will lead the emulator to produce a more accurate approximation to the mechanistic model but will take more computational time. In the extreme, using an extremely large set of mechanistic model runs would allow us to build an emulator, but would also negate the main rationale for using the emulator (computational savings). Using a very small number of mechanistic model runs as a training set would represent a major computational saving, but would risk producing an emulator that was a poor approximation to the mechanistic model.

Aside from computational savings, another rationale for using the emulator is that it allows us to examine the properties of the mechanistic model. In particular, it allows to examine the extent to which variations in the outputs from the mechanistic model can be explained by simple relationships between the model inputs and outputs. This is useful in determining the key features of the mechanistic model that are influencing the behaviour of the model and may be useful in identifying parts of the model that could be simplified without loss of accuracy.

Within this work package we focus upon using SeabORD, and an emulator of SeabORD, to produce estimates of displacement mortality for three species – kittiwake, guillemot and razorbill – under to a range of SPAs and windfarm scenarios, and to identify sources and levels of variation in these mortality rates within each species. For each species we focus upon attempting to build an emulator of SeabORD wind farm impacts that can be applied to all UK SPAs, and, for each SPA, to all wind farms whose footprints are in the CEF Data Store and that lie within the foraging range of the SPAs. This represents an exceptionally large number of scenarios, so we build the emulator using a much smaller training set of SeabORD model runs for each species. For pragmatic purposes, we focus upon three SPAs per species and use colony-specific bird distribution maps derived from the maps of Wakefield et al. (2017) and upscaled to SPA level within the CEF. Since the vast majority of possible windfarm scenarios involve extremely low levels of interaction between windfarm footprint and SPAs, and practical interest lies in situations in which there is a fairly substantial interaction with windfarms, we develop the training set using windfarm scenarios whose interaction with an SPA (as defined using "totalpinords") exceeds a minimum threshold, and thereby consider the effects of between 5 and 11 windfarm scenarios per SPA per species.

We use the SeabORD training runs to build an emulator for each species, in which we relate the SeabORD impacts on adult and chick mortality, and adult mass change, to a range of metrics that summarise the characteristics of the windfarm footprint(s), SPA and spatial interaction between SPA and windfarm(s). We use this emulator to identify the percentage of variation in impacts that can be explained by these characteristics, and to identify the key characteristics that influence the simulated impacts for each species. The aim of the emulator is to predict the levels of displacement mortality that we would expect

SeabORD to produce for a much wider set of SPAs and windfarm scenarios than those used in developing the emulator.

The results of the work must be interpreted cautiously, in large part because of the relatively small training set of SeabORD runs that we were able to use to build the emulator for each species, so we conclude by outlining the key limitations and caveats underpinning the work, and describing the potential for future work in this area (including work that is already planned to take place within WP4 of the ECOWINGS project). We finally conclude by examining the wider implications of the work.

2. Methods

2.1. Selection of species and bird distribution maps

We focus here upon three species: kittiwake, guillemot and razorbill. SeabORD is also available for a fourth species, puffin, but this was not considered here because multi-colony GPS-based maps are not available for this species, and initial explorations using distance decay maps suggested that these were not plausible for all colonies.

For each of these species the colony-specific bird density maps used as inputs to SeabORD for each SPA are derived by upscaling the maps produced by Wakefield et al. (2017) to the level of SPAs.

Wakefield et al. (2017) used multi-colony GPS tracking data to estimate general relationships between spatial distribution and a range of explanatory variables (relating to environmental conditions and levels of competition) for each species and then used the models of these relationships to predict distributions for untracked colonies.

SeabORD is typically applied to SPAs, but the maps of Wakefield et al. (2017) relate to Seabird 2000 subsites (or, strictly, 1km segments within each subsite, but the aggregation from segments up to subsites is straightforward). A method is therefore needed to convert these maps into SPA-based maps. We use a dataset within the CEF Data Store (https://cef-librarybook.datalabs.ceh.ac.uk/) that provides a simple geographical overlap of Seabird 2000 subsites with SPA polygons to estimate the proportion of each Seabird 2000 subsite that lies in each SPA (Data Store dataset #609). The total proportion of birds to apportion to each SPA is then calculated to be the sum of the product (e.g., multiplication) of the proportion of birds apportioned to each Seabird 2000 subsite and the proportion of that subsite that is within the SPA. In the special case that the proportion of subsite in SPA were equal to one for a single Seabird 2000 subsite, and to zero for all other subsites, then the proportion of birds apportioned to the SPA would therefore be equal to that for the Seabird 2000 subsite that had a proportion of one within the SPA. The approach used to derive these overlaps (i.e., to produce dataset #609) is provisional, in the absence of an alternative generic way of aligning Seabird 2000 colony definitions with SPA boundaries but has been flagged as a dataset that requires feedback from stakeholders and may be revised within subsequent iterations of the CEF. Table 2 lists the Seabird 2000 subsites that are calculated within this dataset to relate to each of the SPAs under consideration.

2.2. Selection of SPA-windfarm scenarios

We extract bird distribution maps all of the SPAs included in the CEF Data Store (except Marine SPAs), and extract footprint shapefiles for all of the windfarms for which footprint shapefiles and other information are available in the CEF Data Store. For each species at each SPA, we then identify the set of windfarm scenarios to consider as follows:

- **Stage 1.** we use the bird distribution map and ORD footprint shapefiles to calculate, for each footprint that lies within the foraging range of each SPA, the proportion of the bird distribution that lies within the footprint ("pineachord").

- **Stage 2.** for every possible combination of windfarms that contains five or less windfarms, we then calculate the (a) number of windfarms within the scenario ("nords") and (b) the total proportion of the bird distribution that lies within a footprint ("totalpinords"), which we calculate by summing across the values of "pineachord" for the footprints included within the scenario.

- **Stage 3.** We now retain those scenarios that contain a single windfarm (i.e., for which "nords" is equal to one) if "totalpinords" is greater than 0.001, and those scenarios that contain multiple windfarms (i.e., for which "nords" is greater than one) if "totalpinords" is greater than 0.02.

- **Stage 4.** For each number of windfarms in multi-windfarm scenarios (e.g., each value on "nords" that is greater than one) we select one of the scenarios from step (3) at random. All scenarios with a single windfarm from step 3 (i.e., scenarios for which "nords" is equal to one) are retained.

- **Stage 5.** We exclude any SPAs for which no multiple windfarm scenarios met the criteria at Stage 3.

Considering a maximum of five windfarms at step (2) is necessary for computational reasons but does impose an important caveat on how the results of the work can be used, by ruling out the potential to apply the results to situations involving larger numbers of impacts, or higher levels of overall baseline OWF usage than those captured by the resulting scenarios. The rules at stages (3) and (4) are designed to extract a set of single windfarm scenarios that include non-negligible levels of interaction between SPAs and footprints, and to extract a relatively small number of multi-windfarm scenarios that involve relatively high levels of interaction between SPAs and footprints. The thresholds used are inevitably somewhat arbitrary. The reason for treating single-windfarm and multi-windfarm scenarios differently is that there are an extremely large number of the latter type of scenario, and that multi-windfarm scenarios are computationally more computationally expensive to run through SeabORD than single windfarm scenarios. From a practical perspective, the impacts of single windfarms may be of interest even if they are relatively small, if they contribute to large in-combination effects, so there is also an application-driven rationale for setting different thresholds for consideration of single windfarm and multi-windfarm scenarios.

This selection process resulted in identifying three to four SPAs with a range of sets of offshore windfarms for each of the three species (Table 1).

Table 1. Summary of scenarios considered for wind farm runs of SeabORD for each colony for each species. For each species at each colony the number of scenarios that (a) contain a single windfarm, or (b) contain multiple windfarms. Total number of scenarios per type (single/multi), colony and species are also shown, in grey. The column for the number of scenarios containing multiple windfarms also indicates, in brackets, the numbers of windfarms represented by these scenarios.

Species	Colony	Number of scenarios		
		Single windfarm	Multi windfarm	Total
Kittiwake	UK9002271 (Fowlsheugh)	5	4 (2-5)	**9**
	UK9004171 (Forth Islands)	5	2 (2-5)	**6**
	UK9004271 (St Abb's Head to Fast Castle)	5	4 (2-5)	**8**
	UK9006101 (Flamborough and Filey Coast)	6	4 (2-5)	**10**
	Total	**21**	**14**	**35**
Guillemot	UK9002271 (Fowlsheugh)	7	4 (2-5)	**11**
	UK9004171 (Forth Islands)	6	4 (2-5)	**10**
	UK9004271 (St Abb's Head to Fast Castle)	5	4 (2-5)	**9**
	Total	**18**	**12**	**30**
Razorbill	UK9002271 (Fowlsheugh)	6	2 (4-5)	**8**
	UK9004171 (Forth Islands)	6	4 (2-5)	**10**
	UK9004271 (St Abb's Head to Fast Castle)	5	2 (4-5)	**7**
	Total	**17**	**8**	**25**
TOTAL across all species		**56**	**34**	**90**

Table 2. Alignment of Seabird 2000 subsites with the SPAs considered here, according to the provisional CEF Data Store dataset #609 that contains a simple geographical match of the Seabird 2000 subsite start and end locations to SPA polygons. For each SPA all Seabird 2000 subsites are listed for which the line between the start and end location contains any part of the SPA; for such subsites the proportion in SPA, according to a simple geographical overlap, is listed.

SPA	Seabird 2000 subsite	Prop in SPA
UK9002271 (Fowlsheugh)	Catterline Bay	0.167
	Crawton Bay	0.304
	Fowlsheugh 2	1
	Fowlsheugh 3	1
	Fowlsheugh 4	0.833
	Thornyhive Bay	0.402
	Trelong Bay	0.137
	Tremuda/Old Hall Bay	0.431
	Trollochy Cove	1
UK9004171 (Forth Islands)	Bass Rock	1
	Craigleith	1
	Eyebroughy	1
	Fidra	1
	Inchmickery	1
	North Berwick Coast 1	0.029
	North Berwick Coast 2	0.029
	The Lamb	1
	Whole Island count	1
UK9004271 (St Abb's Head to Fast Castle)	Broadhaven to Moorburn Point	1
	Fast Castle Head to unnamed Cleugh east of Green Stane	0.637
	Lansey Bank to unnamed Cleugh east of Green Stane	0.990
	Moorburn Point to Fast Castle	1
	Redheugh Cottages to Lansey bank	0.167
	St Abb's Head NNR	0.294
	St Abbs to Eyemouth 1	0.059
UK9006101 (Flamborough and Filey Coast)	Barlett Nab	1
	Bempton Gannetry	1
	Breil Newk	1

	Buckton	1
	Cayton Bay 2	0.559
	Dykes End	1
	East Scar	1
	Filey 1	1
	Filey 2	1
	Filey 3	1
	Flamborough	0.853
	Flamborough 1	1
	Flamborough 2	0.343
	Flamborough 3	0.392
	Grandstand	1
	Jubilee Corner	1
	North Cliff	1
	Speeton	1
	Thornwick	1
	Trig Point	1
	Wandale	1

2.3. Generation of SeabORD runs

The new R version of SeabORD being developed within the CEF project is run for each of the scenarios outlined in Section 2.1. Biological parameters within SeabORD for each species are fixed to the values used in Searle et al. (2018), and a unform prey map is used in all cases. The underlying functionality of the R version of SeabORD is very similar to that for the Matlab version of SeabORD. The R version incorporates a revised implementation of inter-colony competition to improve computational efficiency, but this does change not alter the underlying assumptions that the model makes regarding inter-colony competition (Marine Scotland CEF project report, *in preparation*). The R version also utilises a different grid (with wider spatial extent) and a coarser grid resolution.

For each colony, for each species, SeabORD requires a bird distribution map and colony size to be specified. The prey level parameter also needs to be calibrated for each colony. The bird distribution maps are based on scaling the maps of Wakefield et al. (2017) to SPA level, and the colony sizes on scaling estimates of abundance for Seabird 2000 subsites up to SPA level in the same way (Marine Scotland CEF project report, *in preparation*). Within SeabORD birds are simulated to fly from the colony to cells on a regular grid that are assumed to represent foraging locations. The probability of visiting each grid cell at each time step is dependent on the bird distribution map, with birds assumed to have zero probability of visiting grid cells whose distance from the colony exceeds the species-specific foraging range used in Wakefield et al. (2017).

SeabORD is run under both baseline and windfarm scenarios for each colony. The baseline scenario represented the situation in which no wind farms are present. Windfarm scenario consider impacts of wind farms on the flight paths and foraging locations of birds.

For each species at each colony, baseline-only runs of SeabORD are initially used to calibrate the range of prey levels to use within SeabORD. A regular grid of prey values are initially used, and the results inspected by an ecologist with expertise in calibration of SeabORD; if necessary, SeabORD is then run for additional prey levels, until the range of prey levels that produce plausible baseline outputs has been identified. The calibration is primarily based on looking at baseline mass change and baseline chick survival, but other baseline outputs are also examined and may be used in assist in setting the range of plausible values. The output from the calibration, for each species at each colony, is the range of prey levels that lead to plausible baseline outputs. The prey levels used in running SeabORD for windfarm scenarios are then based on taking a regular sequence of twenty values that span this range.

10% of the population is simulated for calibration runs, to improve computational efficiency, as is standard practice when running SeabORD, but 50% (half) of the population was simulated when considering windfarm scenarios.

Following a previous research project (Searle et al. 2020) the displacement rate is assumed to be 60% for all species, colonies and scenarios, and the proportion of displaced birds that also experience barrier effects was assumed to be 100% in all cases. Note that these rates are not intended to capture current advice or may not necessarily be realistic; we use the same rate for all species to aid interpretation of the comparisons between species.

A “border” of 2km (SNCBs, 2022) and a “buffer” of 5km (Searle et al., 2018) were used for all for windfarm footprints. This means that birds lying either within the footprint or within 2km of the windfarm footprint were assumed to be displaced away from the footprint, and that displaced birds were redistributed into the area that lies between 2km and 7km aware from the footprint. Note that the meaning of the term “buffer” in the context of SeabORD differs from the usage of this term in the context of collision risk modelling or the Displacement Matrix (the “border” for SeabORD is effectively equivalent to the “buffer” used in those methods, whilst they do not explicitly consider an equivalent to the “buffer” used in SeabORD).

2.4. Estimating windfarm impacts from SeabORD outputs

Key SeabORD outputs relate to changes in adult mass loss over the course of the breeding season, to chick mortality over the breeding season, and to year-round adult mortality. The final of these is based on translating changes in adult mass into annual survival, and thereby annual mortality, using published mass-survival relationships. The chick and adult survival rates represent the proportion of individuals (chicks and adults, respectively) in the population that are simulated to survive, and the corresponding mortality rates represent the proportion of individuals that are simulated to die. Chick mortality occurs during the breeding season, and adult mortality may occur during both the breeding and non-breeding season, but in practice la largely restricted to the overwinter period.

The adult mass loss over the course of the breeding season is represented as a proportional change, to enhance interpretability and readily enable comparisons between species, so that the proportional adult mass loss is equal to (Mean initial adult mass - Mean simulated adult mass at the final timepoint within the chick rearing period)/ (Mean initial adult mass). Note that the mean initial adult mass is a species-level biological input parameter to SeabORD and is assumed to be the same for all colonies.

SeabORD evaluates the impacts of windfarms by comparing outputs for scenarios that involve windfarms against a “baseline” scenario in which there are no windfarms, so that:

windfarm effect on adult mortality = (Simulated adult mortality rate with windfarm(s) – Simulated adult mortality rate under baseline) = (Simulated adult survival rate under baseline – Simulated adult survival rate with windfarm(s)) windfarm effect on chick survival = (Simulated chick survival rate with windfarm(s) – Simulated chick survival rate under baseline)

windfarm effect on adult mass change = (Simulated mean proportional adult mass change over the chick rearing period with windfarm(s) - Simulated mean proportional adult mass change over the chick rearing period under baseline)

These represent differences between the windfarm(s) and the baseline, so these values will be positive if the windfarms lead to an increase in mass loss or adult/chick mortality relative to the baseline (i.e., to a decrease in mass change or adult/chick survival relative to the baseline), and negative if they lead to a reduction in mass loss or adult/chick mortality (i.e., to an increase in mass change or adult/chick survival) relative to the baseline.

The impacts of windfarms within SeabORD depend on all of the windfarms that are designed to be present: SeabORD does not attempt to separately estimate the effect of these windfarms, but rather considers their simultaneous cumulative impacts, and automatically accounts for any interactions between windfarms (if there are of a form that can be captured by the mechanisms represented within SeabORD).

The "windfarm effect on adult mortality" and "windfarm effect on chick mortality" values defined above are population-level values, so relate to all individuals within the population, including those that never interact with the windfarm(s) or experience displacement. In WP4 we outline how these outputs can be used to derive estimates of "displacement mortality rates": i.e., mortality rates relevant to the subset of birds that experience displacement.

2.5. Variables used for emulation

A separate emulator was developed for each species for each of the following three response variables: a) windfarm impact on adult mass loss, b) windfarm impact on chick mortality and c) windfarm impact on adult mortality. Each emulator attempted to quantify the relationship between the response variable and a range of potential explanatory variables.

These explanatory variables are designed to represent the inputs to SeabORD. The biological parameters to SeabORD are held fixed across all runs for each species, and prey maps are assumed to be uniform in all cases, so these are not considered here. We instead focus on those inputs that do vary between runs, which are of key practical interest in developing models of SeabORD outputs that could be translated to new colonies and windfarm scenarios. The inputs to SeabORD that do vary between runs, within each species, are (1) prey level, (2) colony size, (3) windfarm footprints, (4) colony location and (5) bird distribution map. The prey level can vary between all runs (because multiple calibrated prey levels are considered for each scenario), whereas colony size, colony location and bird distribution map vary only between colonies, and windfarm footprints only between windfarm scenarios.

The primary explanatory variable that we consider, which we abbreviate as "ptdisp", is the proportion of individuals that are displaced at each timepoint. Since the spatial locations of individuals are assumed within SeabORD to have the same distribution at each time point within the chick rearing period, and are assumed to be independent from one timepoint to the next, this quantity can be derived by simply multiplying the displacement rate by the total proportion of the colony-specific spatial distribution of birds that is contained within any windfarm footprint (or within 2km of the footprint): this is the quantity "totalpinords" that we defined in Section 2.2, and have already used in deciding the set of windfarm scenarios to consider. Since we assume a displacement rate of 0.6 for all species and scenarios, it would effectively to be equivalent to include either "ptdisp" or "totalpinords" as an explanatory variable in the emulator, but we use "ptdisp" because the parameter associated with the effect of this variable then has a direct interpretation in relation to displacement mortality rates (WP4).

We expect “ptdist” (or equivalently “totalpinords”) to be the explanatory variable that is most strongly related to the windfarm effects on adult survival, chick survival and adult mass change, because it provides a direct measure of the rate at which bird-windfarm interactions will occur within SeabORD.

All of the other explanatory variables that we consider are designed to try to capture more subtle features of SeabORD, that lead displacement mortality to vary in relation to SPA and ORD characteristics in ways that are not solely determined by the rate at which displacement is occurring.

The first additional explanatory variable that we consider is

- “prey”: prey level in grid cell (grams per km^2).

and the second is

- “colsize”: colony size (number of birds).

both of which are direct inputs to SeabORD. The next two explanatory variables that we consider represent key characteristics of the windfarm scenario being considered, and are calculated from the footprint shapefiles for each windfarm:

- “nords”: number of wind farms in scenario.
- “totalfparea”: total area of wind farms within the scenario, summed across all windfarm footprints (km^2).

The next three explanatory variables represent characteristics of the spatial relationship between the colony location and the windfarm footprints:

- “meandist2spa”: mean distance by sea from windfarm midpoint(s) to colony (km).
- “maxsweptangle”: mean angle ‘swept’ each wind farm in relation to seabird colony. For a single wind farm, this is the range of angles within which the bird flying in a straight line away from the colony would, at some point, fly across the wind farm footprint. For scenarios with multiple windfarms this is calculated by averaging the angle across windfarms.
- “fpalignment”: average wind farm alignment in relation to bird colony. For a single wind farm, this metric is calculated as ratio of the radius of a circle determine by the maximum swept angle and the mean distance to colony (which can be calculated as (TAN(maxsweptangle/2) x meandist2spa) by trigonometry) to the radius of a circle determined by the windfarm area (which can be defined as SQRT(totalfparea/π) based on the formula for the area of a circle). For scenarios with multiple windfarms this is calculated by averaging the ratio across windfarms.

The metric “maxsweptangle” represents the extent to which the wind farm “blocks” flight paths away from the colony. It is, however, related to wind farm size (since a larger windfarm will, all else be equal, have a larger value of “maxsweptangle”). The alignment metric, “fpalignment”, is constructed in order to try to construct a metric that quantifies the magnitude of this “blockage”, but is independent of the size of the windfarm. Note that the alignment metric, “fpalignment”, is equal to one for a wind farm that is a circle, less than one for a wind farm which appears shorter when viewed from the colony than when viewed from other angles, and greater than one for a wind farm which appears longer when viewed from the colony than when viewed from other angles. Windfarms with values of “fpalignment” greater than one operate as a greater barrier to flight paths from the colony than might be expected from the size of the windfarm alone, whereas windfarms with values of “fpalignment” less than one operate as less of a barrier to flight paths from the colony than might be expected from the size of the windfarm alone.

The final explanatory variable represents a characteristic of the interaction between the windfarm footprints and the bird distribution map that is not captured by "ptdisp":

- "pbeyond": proportion beyond windfarm: this quantifies the proportion of the bird distribution whose distance to colony is lower than the mean distance (by sea) between grid cells within the footprint and the colony. This is designed to provide a crude estimate of the proportion of individuals that will experience barrier effects that can be rapidly calculated – a more accurate calculation of the proportion of individuals experiencing displacement would involve calculating the shortest path distance from the SPA to each cell of the grid, which can be relatively computationally intensive to calculate.

The direct inputs to SeabORD ("preylevel" and "colsize") are natural explanatory variables to include, and the windfarm scenario characteristics ("nords" and "totalfparea") appear obvious variables to represent the key characteristics of the windfarms. By contrast, the remaining metrics are all designed to provide useful summaries of the interactions between windfarms, colonies and bird distribution maps, but there are also many other potential metrics that could have been considered to summarize these interactions based on the same inputs. The selection of the above metrics as potential explanatory variables is therefore not designed to be comprehensive, but rather to ensure that the variables being considered capture all of the windfarm-colony characteristics that are likely to be key to predicting the level of impact that SeabORD will simulated.

All else being equal, we might expect that the average magnitude of windfarm impacts would increase with increases in the proportion of inividuals displaced per timepoint ("ptdisp"), which is in turn directly related to the total proportion of colony-specific spatial distribution of birds that is contained within a windfarm footprint ("totalpinords"). We might also expect positive relationships between impacts and the number of windfarms ("nords"), the total area within windfarms ("totalfparea"), the mean swept angle ("maxsweptangle") and because increases in each of these quantities would, all else being equal, lead to an increased rate of interactions between bird and windfarms. However, which of these variables captures this best is unclear, and each of these characteristics may capture subtly different effects.

Given the strong central place foraging constraint in the chick rearing period we might also expect impacts to reduce with increases in the mean distance to colony ("meandist2spa"). The alignment metric ("fpalignment") is designed to capture the relative importance of barrier effects, given that displacement and barrier effects operate differently within SeabORD, by identifying the extent to which the windfarm operates as a barrier. The swept angle ("maxsweptangle") will also capture this, but the advantage of "fpalignment" is that, unlike "maxsweptangle", it standardizes for the size of the windfarm, and so can be interpreted more readily.

Within R, the footprint areas needed for "totalfparea" are calculated using the "st_area" function from the "sf" package, the angles needed for "maxsweptangle" can be calculated using the "angle.calc" function from the "Morpho" package, and the metrics that rely on spatial matching between footprints and colony locations or bird distributions maps ("meandist2spa", "maxsweptangle", "totalpinords") can be calculated using the "extract" function from the "raster" package.

2.6. Emulation modelling

Emulation involves using a statistical model to describe the relationship between the inputs and outputs of a mechanistic model. We construct emulators separately for each of the three response variables described above (i.e., each of the key metrics of windfarm impacts produced by SeabORD) and do this separately for each species.

A wide range of models or methods can potentially be used for the purposes of emulation, but we focus here upon two approaches that are relatively simple to interpret, and computationally extremely fast to use: multiple regression, and linear mixed models. Within each of these methods we develop emulators separately for each response variable. Exploratory analyses suggest that there are no clear benefits in using bivariate approaches to emulation in this context, although this merits further exploration.

Defensible approaches to emulation within this context need to account for the clear hierarchical structure of the data – the most obvious aspect of this hierarchical structure is the fact that there are twenty SeabORD runs per scenario, with each of these runs having identical inputs for everything apart from the level of prey (and even this only varies between runs within the relatively narrow range generated by the baseline calibration). These runs are used to represent both uncertainty in the (unknown) prey level and inherent stochasticity within the model. The result is that the number of raw SeabORD runs per species (600 for guillemot, 700 for kittiwake, 500 razorbill) is always twenty times higher than the number of scenarios considered (30 for guillemot, 35 for kittiwake, 25 razorbill).

Modelling approaches that fail to account for this hierarchical structure risk over-stating the effective sample size – none of the key explanatory variables of interest vary between the runs within a scenario, for example, so the effective sample size for learning about these effects will effectively be much lower than the number of individual SeabORD runs.

We consider two approaches to constructing emulators in ways that address this issue, and assess the sensitivity of the results to the choice of approach. The first approach involves using multiple regression. Using multiple regression to model the raw SeabORD outputs is problematic, because these outputs have a clear hierarchical structure that cannot readily be captured through a regression model, and this is liable to lead the model to over-state the effective sample size, and hence to under-estimate uncertainty (and thereby, for example, over-state the significance of explanatory variables). We attempt to avoid this pitfall by calculating the mean impact per scenario, and then modelling this mean impact, rather than the outcomes of individual SeabORD runs, via multiple regression. The averaging across runs per scenarios means that the sample sizes for the multiple regression are small (25 to 35 per species), which is, in turn, likely to lead to high levels of uncertainty and relatively low levels of power to detect effects.

We also investigate a second approach, whereby we apply emulators to the raw SeabORD runs. For these emulators we use mixed models, rather than linear regression models, in order to be able to account for the hierarchical structure of the data. The mixed models each account for a single source of variation via the random effects, "scenario", with other sources of variation being dealt with via the fixed effects (see below). We did initially investigate the construction of mixed models that accounted for other sources of variation via random effects, but there were issues of non-convergence that did not appear straightforward to resolve, possibly arising from the imbalanced structure of the data in relation to ORDs and SPAs, so we have instead tried to capture these sources of variation in other ways (e.g., by considering models that include "SPA" as a fixed effect).

2.6.1. Regression models of mean impact per scenario

We use the multiple regression analyses of the aggregated data (mean impact per scenario) as our main analyses, since model selection and goodness-of-fit evaluation are more straightforward to interpret within these models but then compare the final models against equivalent mixed models.

The "null" model represents the simplest model that we are prepared to consider, and all of the other models that we consider generalise this model in different ways. We define the null model in this context (which we call "R1") to be a simple linear regression model of the mean impact per scenario, without intercept, in which the explanatory variable is "ptdisp" (the expected proportion of individuals displaced per timepoint).

This model is a linear regression model in which the response variables, which each relate to a population-level impact, is assumed to be linearly related to the proportion of individuals displaced per timepoint ("ptdisp"). This model contains no intercept, so that windfarm impacts are assumed to be exactly equal to zero when there are zero birds within the footprint: this is a fairly plausible assumption in this context because the structure of SeabORD guarantees that SeabORD must always, when there is no overlap between the bird distribution maps and any windfarm footprint, be no individuals that experience displacement, hence no impact of displacement. It is possible for there to be barrier effects in this situation, although this is unlikely to occur whenever the bird distribution map is dominated by distance to colony effects (as is the case in the Wakefield et al., 2017, maps that are used here to provide the spatial inputs to SeabORD). It is also possible for the omission of an intercept to be problematic as a statistical modelling assumption if the point of intercept involves substantial extrapolation beyond the range of the data, but that is not the case here, as the distribution of values of "ptdisp" are heavily skewed towards values close to zero (as a result of the actual distribution of windfarms and colonies). The intercept is excluded here primarily to aid interpretation, because it allows the parameters of the model to be expressed in relation to displacement mortality rates (WP4).

The null model, R1, is extremely simple: it only contains one parameter (aside from the level of residual variation), which represents the slope between the response variable (windfarm impact on adult mortality, chick mortality of adult mass loss, depending on the analysis) and the proportion of birds in the footprint. In WP4 we will argue that this slope parameter can effectively be regarded as the displacement mortality rate, under one particular definition of that rate – in order to distinguish this from the differing definitions of this rate used in the Displacement Matrix and Expert Elicitation we will term this a "model-based displacement mortality rate". Model R1 represents an assumption that this "model-based displacement mortality rate" (for either adults or chicks, depending on the response variable being considered) is the same under all circumstances – in particular, for all SPAs and all windfarm scenarios. It is for this reason that we consider this to be our "null" model and compare other models against this. The null model for adult mass loss has the same form but obviously does not have an interpretation as a mortality rate.

Model R1 assumes that the same slope can be assumed for all SPAs and windfarm scenarios, and for all combinations of SPA and windfarm scenario, implying the same model-based displacement mortality rate in all of these situations. We compare this against alternative models in which the slope between the response variable (e.g., SeabORD estimates of population-level wind farm effects) and the level of displacement per timestep ("ptdisp") is allowed to vary in different ways, in order to evaluate the extent to which it is plausible to assume a common displacement mortality rate (or equivalent standardised rate of adult mass loss) across all situations. All of the models that we consider can be expressed in the form:

Response variable = Overall slope parameter * ptdisp * (1 + model for variations in slope) + Noise

[Equation 1]

and interest lies in (a) quantifying the absolute value of the overall slope parameter (which translates to an overall model-based displacement mortality rate) and (b) understanding variations in this slope.

The first additional model that we consider, R2, is a regression model that has both linear and quadratic effects of "ptdisp". The null model (R1) is a special case of this model in which the quadratic effect is equal to zero. Quadratic models are generally considered in order to examine the evidence of the existence of non-linear, rather than linear, effects, and that is also true in this case within the structure of the model, which focuses on the response variable. However, our interest is in variations in the slope (model-based displacement mortality rate), and, viewed from that perspective, the quadratic model is instead testing whether there is evidence for the slope (model-based displacement mortality rate) varying with the level of displacement ("ptdisp"). Viewed in terms of Equation 1, model R1 is equivalent to a model in which:

Model for variations in slope = 0 and model R2 equivalent to a model in which Model for variations in slope = Slope of slope * ptdisp

Models R3-R10 consider whether the slope varies in relation to the number of windfarms (nords, R3), the total area within a footprint (totalfparea, R4), the mean distance from windfarm to SPA (meandist2spa, R5), maximum angle swept by a wind farm (maxsweptangle, R6), mean windfarm alignment (fpalignment, R7), proportion beyond windfarm (pbeyond, R8), mean prey level (prey, R9) and colony size (colsize, R10). In each case, the additional variable is included in the model as an interaction with “ptdisp”. Each of these models therefore contains an overall slope in “ptdisp”, and a parameter that represents the interaction between “ptdisp” and the additional variable. This is equivalent to assuming that the slope depends on the additional variable, so that, in the notation of Equation 1,

Model for variations in slope = Slope of slope * additional explanatory variable

We only consider the additional variables in interaction with “ptdisp” because (a) this allows the model to be interpreted in relation to displacement mortality rates, (b) many of the variations are only defined if a windfarm is present (so, e.g., cannot even be defined when “ptdisp” is zero), and (c) because the structure of SeabORD means that the effects of all of these variables must be zero when “ptdisp” is zero, and means that the magnitude of their effect is likely to depend on the level of displacement.

The final regression model that we consider, R11, assumes a separate slope in “ptdisp” for each SPA (effectively equivalent to considering the interaction of “ptdisp” with SPA).

All of the models that we consider contain between 2 and 4 parameters (the residual standard deviation being one of the parameters in each case). We restrict attention only to relatively simple models because the small sample sizes mean it is unlikely to be realistic to fit more complex models – we did investigate some more complex models at an exploratory stage, but the results appeared to indicate that there were difficulties in estimating all of the model parameters, so we restricted attention only to the models listed here.

The models that we have considered can be divided into two types: those which descriptively summarise the variation not explained by the slope with “totalpinords”, but cannot provide a basis for predicting the variations in this slope that would occur at SPAs or windfarms beyond those for which we have data (i.e., SeabORD runs), and those that attempt to explain the variation in this slope in terms of explanatory variables that would be known for other SPAs or windfarms, thereby allowing this model to provide predictive estimates of displacement mortality.

All eleven of the models considered can, at least in principle, provide a basis for predicting rates for new windfarm scenarios for the SPAs that were used in fitting, although the relatively small number of SeabORD runs used to fit the emulators mean that these models, and predictions, should be interpreted cautiously.

The model that assumes a separate slope for SPA (R11) cannot be generalised to SPAs beyond those used in the modelling, because it provides no basis for quantifying the rates at SPAs other that those used in the model. This means that, if this model is found to perform well relative to other models, that the emulator cannot defensibly be applied to SPAs apart from the SPAs used to train it (e.g., included in the SeabORD model runs used to develop the emulator). The model that includes “prey” as an explanatory variable (R9) can also, in effect, not be extrapolated to SPAs beyond those used in the modelling, because we do not know the calibrated prey levels of SeabORD for these colonies (and, indeed, the avoidance of needing to calibrate SeabORD separately for each new colony is one of the key motivations for using the emulator). The remaining models all do, in principle, allow prediction to SPAs other than those used in model fitting, because they describe variations in terms of explanatory variables that are readily available for all SPAs and scenarios. However, as the number of runs, and particularly number of SPAs, is too small to allow a thorough assessment of the generalizability of the results, predictions should again be treated very cautiously. In the case of model R10 (colony size), the small number of SPAs used in the model fitting (3 to 4 per species) makes extrapolation particularly difficult.

We consider both transferrable (to new SPA) and non-transferrable models in order to evaluate the extent to which relationships can be generalized across colonies – if a model that contains "prey" exhibits much better fit than a model that does not, for example, then this would suggest that we should be cautious in generalising from the model without prey, because there are key sources of variation that it is failing to capture. Similarly, the models that contains "SPA" are not generalisable, but can be used to provide information on the generalisability of other models: if models containing "SPA" have greater empirical performance than models containing explanatory variables that can be used for prediction, this indicates that the explanatory variables are not capturing all of the variation between SPA seen in the raw data.

The models being considered, and their potential for transferability, are summarised in Table 3.

Table 3. Summary of regression models compared for the purpose of selecting an emulator. The same set of models are used for all three response variables (impact on adult mass loss, impact on chick mortality, impact on adult mortality) for each of the three species. The final column indicates whether the model is potentially transferrable to SPAs other than those whose SeabORD runs have been used in fitting the emulators; "yes*" indicates a model that could technically be transferred, but where this is not recommend due to the small number of unique values available.

Model	Model structure for response variable (population-level impact)	Model for model-based displacement mortality rate	Transferrable to new SPAs?
R1	β_1*ptdisp + Noise	β_1	Yes
R2	β_1*ptdisp + β_2 * ptdisp^2 + Noise	$\beta_1 + \beta_2 * \text{ptdisp}$	Yes
R3	β_1*ptdisp + β_2 * ptdisp * nords + Noise	β_1 + β_2 * nords	
R4	β_1*ptdisp + β_2 * ptdisp * totalfparea + Noise	β_1 + β_2 * totalfparea	Yes
R5	β_1*ptdisp + β_2 * ptdisp * meandist2spa + Noise	β_1 + β_2 * meandist2spa	Yes
R6	β_1*ptdisp + β_2 * ptdisp * maxsweptangle + Noise	β_1 + β_2 * maxsweptangle	Yes
R7	β_1*ptdisp + β_2 * ptdisp * fpalignment + Noise	β_1 + β_2 * fpalignment	Yes
R8	β_1*ptdisp + β_2 * ptdisp * pbeyond + Noise	β_1 + β_2 * pbeyond	Yes
R9	β_1*ptdisp + β_2 * ptdisp * prey + Noise	β_1+ β_2 * prey	No
R10	β_1*ptdisp + β_2 * ptdisp * colsize + Noise	β_1 + β_2 * colsize	Yes*
R11	β_{SPA}*ptdisp + Noise	β_{SPA}	No

2.6.2. Model comparison

We have considered a range of different possible models in order to investigate the sensitivity of the results to the choice of model, and to evaluate the relative importance of the different explanatory variables. By considering this range of models we aim to ensure that we have identified the key features that are driving explainable variations in the SeabORD model outputs. Consideration of a range of different possible models reflects the fact that is essentially an exploratory, rather than confirmatory, analysis, that aims to identify the factors that influence the relationship between model inputs and outputs within SeabORD and to ensure that no important factors are omitted, rather than to investigate a specific hypothesis. It also reflects the fact that the main objective is effectively prediction rather than explanation – i.e., to produce an emulator that provides an accurate approximation to SeabORD. We aim to interpret the results across models, rather than focusing on a single "best" model, and focus on those results that are consistent across different modelling assumptions, and that do not appear unduly influenced by the caveats that underpin our analyses.

We fit each multiple regression model in R using the "glm" function in the "stats" package.

We extract the R-squared for each model – this quantifies the overall proportion of variance in the response variable that is explained by the explanatory variables and gives an indication of the extent to which the explanatory variables that we can identified are capturing variations in model output. R-squared does not account for model complexity, but provides a general, and readily interpretable, indication as to whether the addition of additional complexity into models is leading to practically substantive increases in the level of variation explained. R-squared values for models without intercept are calculated (by default, within the lm function in R) in relation to the total squared value of the response variable (i.e., in effect, using the variance that would be obtained if the mean were zero, rather than using the sample mean).

In order to select between models, however, we need to use a criterion that does adjust for model complexity. We compare the performance of the different models empirically using AICc (the Akaike Information Criterion, with small sample size adjustment applied) – AICc is a measure of empirical model performance that penalizes model complexity. Lower AICc values indicate better empirical fit, after adjusting for complexity. Only the relative values of AICc between different models matter, as the absolute value has no useful interpretation: the deltaAICc value is calculated by deducting the minimum AICc value (across all models) from each AICc value, so that all AICc models are evaluated relative to the model with best fit (whose deltaAICc value is, by definition, equal to zero). Standard rules of thumb are that models whose deltaAICc value is less than two have comparable empirical performance to the best model, whilst models whose deltaAICc value is greater than ten have very low empirical support relative to the best model (Burnham & Anderson, 2002). Rather than attempting to select the best model for each species and response variable we instead, to aid interpretability, look for the model that provides the best overall empirical performance (as assessed by AICc) when looking across all response variables and species.

2.6.3. Goodness of fit and cross-validation

We explore the performance of the models using standard visual model diagnostics.

We also investigate whether there is any evidence for variation between colonies, by refitting the emulation models separately to training data for each colony. Such models could not be used for prediction to new colonies, so are rather used to evaluate the potential for variations in effects between colonies that might represent failures in the assumptions of the species-wide (i.e., cross colony) emulation model. The colony-specific emulation models have the same structure as the cross-colony models.

We use cross-validation to investigate the out-of-sample predictive performance of the models, focusing on the regression models of mean impact per scenario. We begin by splitting the data into a training set (containing 75% of the model runs) and test set (containing 25% of the model runs) and show the results of building an emulator on the former and attempting to predict the values of the response variable in the latter. One potential difficulty with this approach to cross validation, however, is that it ignores the hierarchical structure of the data and so may yield over-optimistic assessments of performance. We therefore also consider an alternative approach in which we treat 2 of the 3 SPAs as the “training set” and fit an emulator to these which we then use to try to predict the response variable for the remaining SPA.

2.6.4. Linear mixed models

The regression models are applied to mean impacts per scenario. Since this represents a loss of information relative to the raw data, which consist of 10 runs per scenario, we also analyse the raw data using linear mixed models and investigate whether the results are similar to those obtained by applying regression to the aggregate (i.e., mean) values. Linear mixed models are needed for the analysis of the raw data in order to avoid pseudo-replication (e.g., over-stating the effective sample size) by accounting for the structure of the data.

We construct the linear mixed models by assuming a random slope with “scenario” (i.e., unique combination of windfarms, SPA and species). This is a “random coefficients” model. We assume, for comparability with the regression models, that there is no random intercept in the model. Random coefficient models are appropriate, given that we would expect the effect of “ptdisp” to dominate variation in the response variables and so need to consider variation in this effect. In terms of model-based displacement mortality rates, these models assume that there is random unexplained variation in the displacement mortality rates between scenarios. These random coefficients capture the idea that each SPA-windfarm combination may have a separate slope.

We take the null model (R1) from the regression modelling and the overall best fit model from the regression modelling (based on AICc), and re-fit each of these as mixed models. We fit each mixed model using the “lmer” function in the “lme4” package.

Within the initial stages of the work, we also investigated mixed models and regression models that did contain intercept terms (for both fixed effects and random coefficients, in the case of the mixed models). For the regression models the intercepts were estimated to be close to zero, and for the mixed models the estimates of intercept terms either produce estimates of the intercept that were close to zero, or else encountered problems in estimation. Since we effectively know that an intercept does not exist, and since the “no intercept” models have a clearer biological interpretation in terms of displacement mortality rates, we have therefore focused throughout on fitting models without an intercept. We also investigated other generalizations of the mixed models considered here - in particular to include a random effect for windfarm scenario, and to regard SPA as a random rather than fixed effect, but these ran into a range of problems around non-convergence that seem likely to arise from the small sample size and imbalanced structure, so we did not pursue these further.

2.6.5. Prediction

A key aim of the emulator was the development of a model that could be used to predict displacement mortality rates for SPA and windfarm scenarios other than those for which SeabORD has already been run.

Because the emulation models are explicitly formulated in relation to model-based displacement mortality rates, prediction from these models is, from a technical perspective, very straightforward. Model R1, for example, assumes that the same model-based displacement mortality rate applies in all circumstances, and can be estimated by estimating the slope parameter of this model. Within that model, “prediction” is therefore trivial – it simply involves assuming that this same rate applies everywhere. Prediction from R11 is similarly trivial, although in that case the rates, and hence predictions, are SPA-specific. Prediction from models R2-R10 is also straightforward, since each of these models can be expressed as a model for the displacement mortality rate.

Note, however, that although prediction is technically straightforward, prediction beyond the range of values used for model fitting involves extrapolation, which relies on very strong assumptions that may not be plausible. The emulation models we construct are therefore intended only to be used for prediction within the range of values of the explanatory variables used in constructing the emulation – i.e., the SeabORD runs outlined in Section 2.3.

We extract the estimated effects of the explanatory variables on the response variable within each model, together with associated standard errors (which provide a measure of uncertainty) and p-values. We use the p-values to investigate the strength of evidence for effects (i.e., to indicate the chance that the effect could have arisen by chance alone), and the parameter estimates to give an indication of effect sizes (i.e., the extent to which the estimated effects are of a magnitude that is of practical importance).

3. Results

3.1. SeabORD runs

The calibrated prey level ranges associated with each colony for each species are shown in Table 4. Note that the ranges are typically relatively narrow, and that there is considerable variation in ranges between colonies within a species, emphasising the need to calibrate SeabORD separately for each colony.

Table 4. Calibrated range of prey levels used for running SeabORD at each colony. Runs with windfarm impacts were run using 20 prey levels spread uniformly across this range: for guillemot at Forth Islands, for example, there were 20 levels from 336 to 340 in steps of 0.2015.

Species	Colony	Prey range
Kittiwake	UK9002271 (Fowlsheugh)	283-283
	UK9004171 (Forth Islands)	156-158
	UK9004271 (St Abb's Head to Fast Castle)	197-200
	UK9006101 (Flamborough and Filey Coast)	253-256
Guillemot	UK9002271 (Fowlsheugh)	494-510
	UK9004171 (Forth Islands)	336-340
	UK9004271 (St Abb's Head to Fast Castle)	372-382
Razorbill	UK9002271 (Fowlsheugh)	286-302
	UK9004171 (Forth Islands)	298-315
	UK9004271 (St Abb's Head to Fast Castle)	224-233

In Figures 1-3 we plot the levels of displacement per timestep (ptdisp) against the simulated SeabORD impacts for each of the three impacts (chick mortality, adult mortality, adult mass loss) for each species. We do this using both the raw SeabORD runs and using the mean SeabORD impact per scenario. We plot impacts (on adult mortality, chick mortality and adult mass loss) against the proportion of birds that experience displacement per timestep, and at any point during the chick-rearing period, in Figures 1-3. The proportion of birds experiencing displacement per timestep is simply calculated to be the proportion of time spent in the footprint (totalpinords) multiplied by the displacement rate. This figure is designed to illustrate the extent to which the impacts simulated via SeabORD can be explained by this simple metric. There is some apparent relationship with this simple metric in all cases (Figures 1-3), with the relationship being strongest for kittiwake (in part because there is the greatest level of variation in this metric for kittiwake, as the scenarios for kittiwake include higher values for this metric than the scenarios for guillemot and razorbill), but there is also substantial noise (unexplained variation). , There is no obvious visual evidence of non-linearity over the range of values considered - we would ultimately expect this to a non-linear relationship, at very high levels of displacement, but the assumption of linearity appears broadly plausible for the range of displacement levels considered here. However, this may well indicate insufficient evidence from the available runs to demonstrate the form of any non-linearity, rather than evidence against non-linear effects. There is clear visual evidence for differences in relationships between SPAs for kittiwake.

In general, for all three species, predicted impacts on chick mortality (Figure 1) and adult mortality (Figure 2) resulted in increases in mortality when the windfarm(s) was present, but there are a substantial proportion of runs for which both chick and adult mortality was predicted to decrease in relation to the baseline – SeabORD allows for both positive and negative impacts on chicks and adults within simulations, and such positive impacts are often related to stochastic noise around very small or negligible impacts from windfarms. Adult mass change was predicted to alter in the presence of windfarm(s) across all the SeabORD runs for each species, with adults in general, but not always, losing more mass in the presence of windfarm(s) when compared to the baseline (Fig. 3). Some SeabORD runs resulted in no apparent average net loss of additional adult mass in the presence of the windfarm(s) for guillemot and razorbill, or a net gain, at the population level, but this is to expected, given the relatively high levels of inherent variation between runs and the low level of interaction between colony-specific bird distribution maps and ORD footprints for many of the single windfarm scenarios. Overall, the relationship with mass loss showed less residual variation than the relationships with adult and chick mortality, reflecting the additional stochasticity involved in simulating mortality.

There was considerable variation in simulated impacts across the SeabORD runs, although this was substantially lower when considering mean impacts per scenario. Our focus within the emulation work that follows will essentially be on (a) trying to estimate the slopes associated with the graphs shown in Figure 1-3, and (b) trying to quantify and, where possible, explain, the remaining variation seen within these plots in terms of other factors.

Figure 1. Scatterplots of proportion of birds displaced per timestep (ptdisp) against simulated SeabORD impacts on chick mortality for each species, with different colours representing different SPAs. Left hand plots show raw SeabORD impacts, and right-hand plots show mean values per scenario. The dashed black line represents an impact of zero.

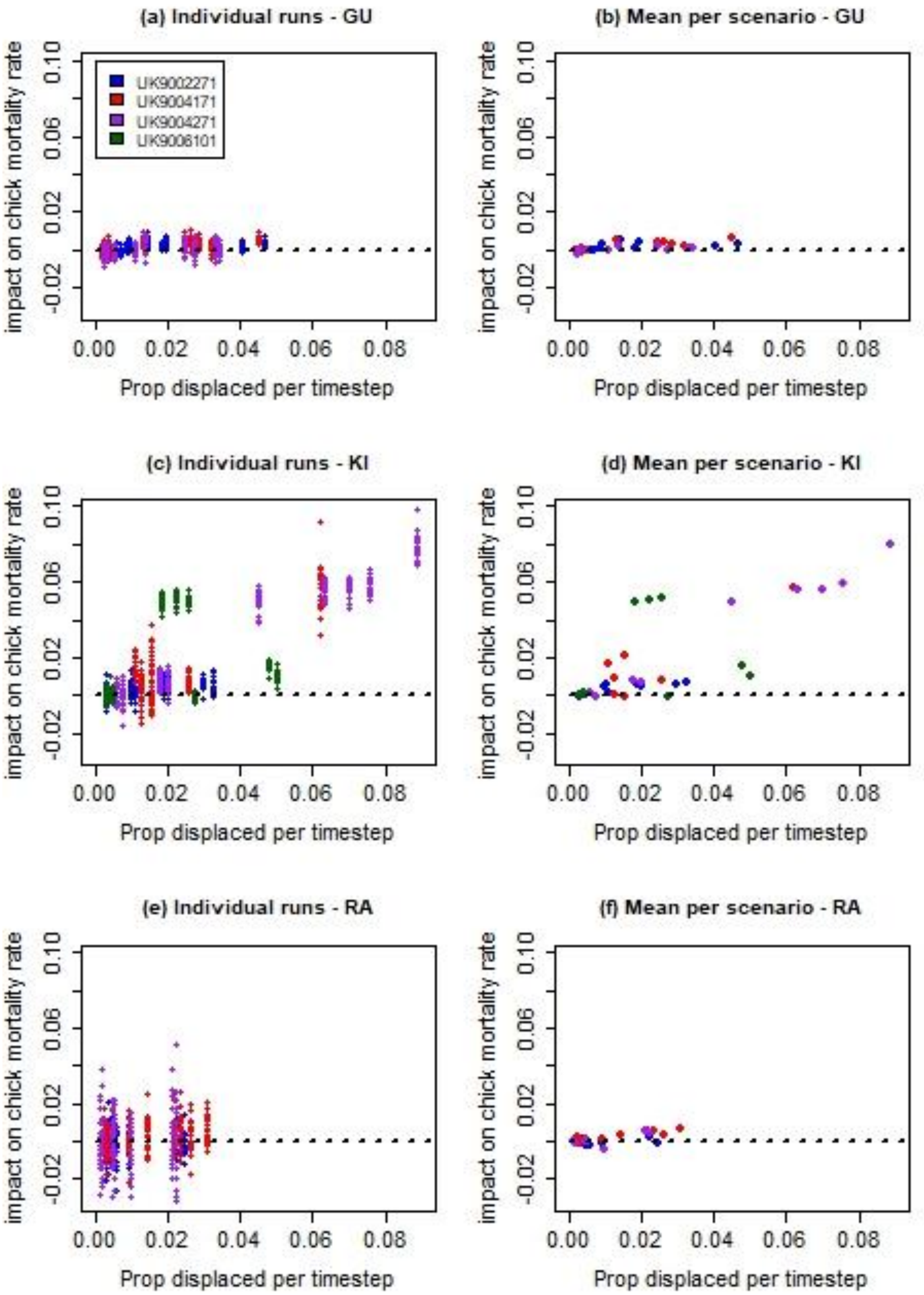


Figure 2. Scatterplots of proportion of birds displaced per timestep (ptdisp) against simulated SeabORD impacts on adult mortality for each species, with different colours representing different SPAs. Left hand plots show raw SeabORD impacts, and right-hand plots show mean values per scenario. The dashed black line represents an impact of zero.

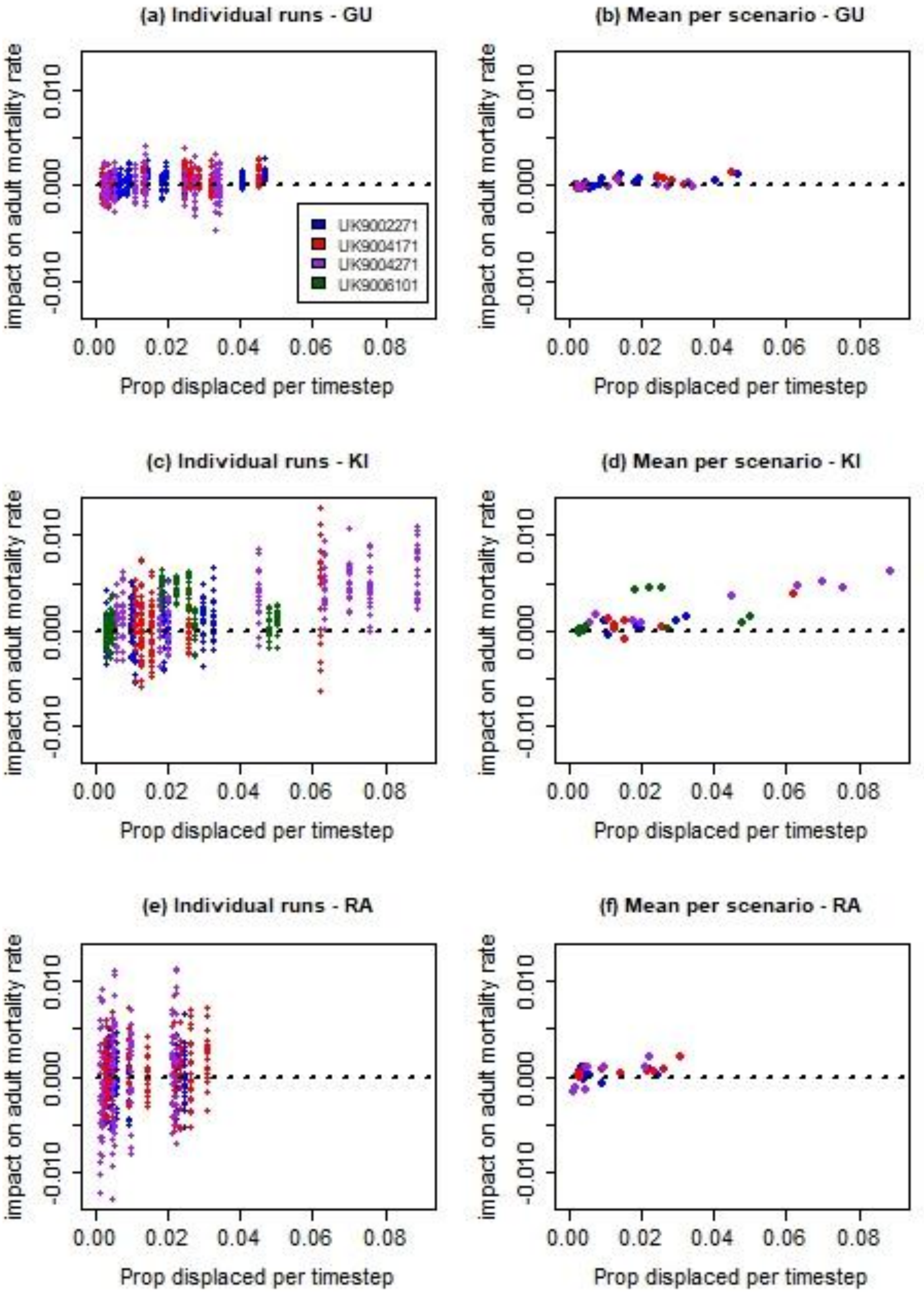
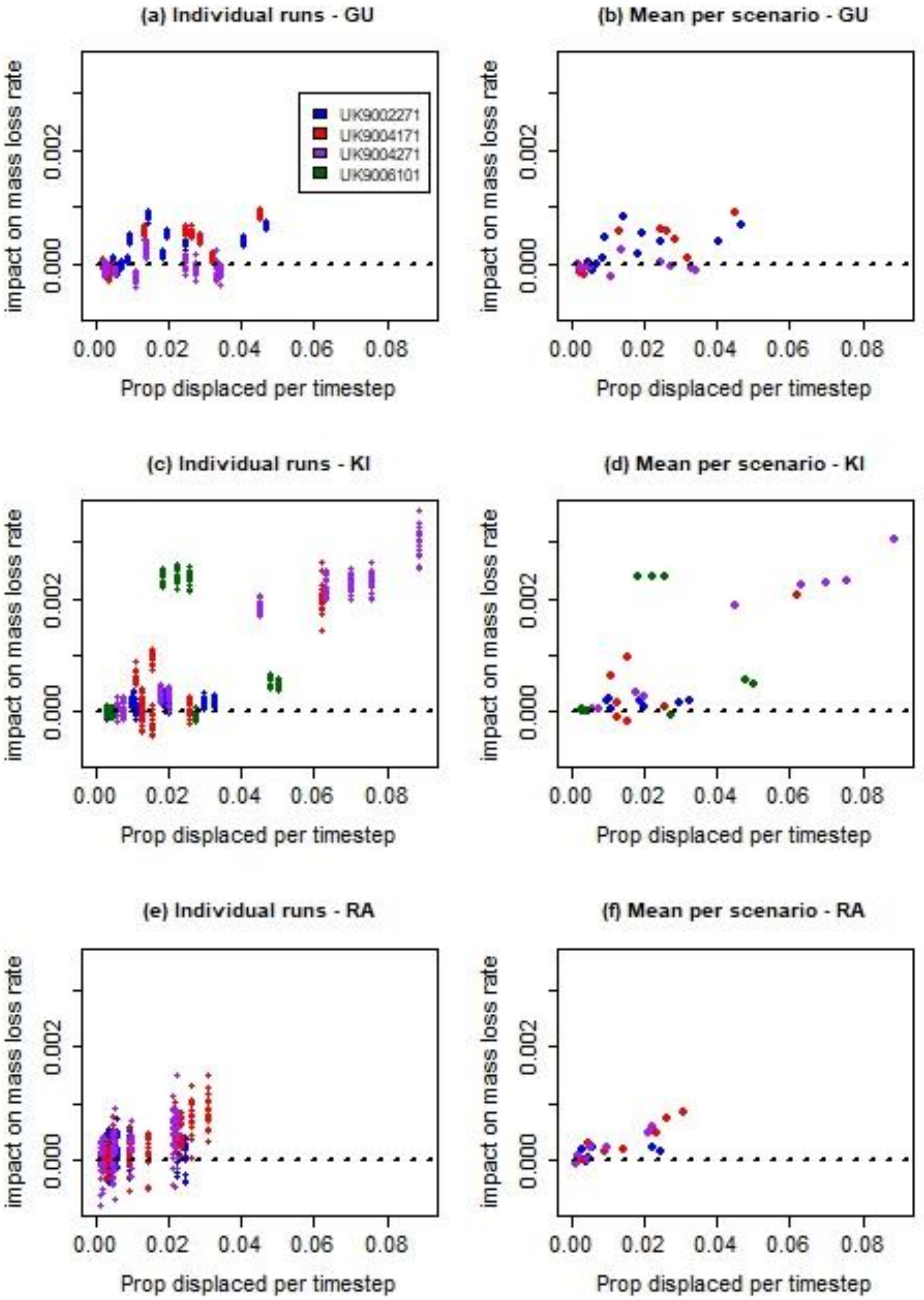


Figure 3. Scatterplots of proportion of birds displaced per timestep (ptdisp) against simulated SeabORD impacts on adult mass loss for each species, with different colours representing different SPAs. Left hand plots show raw SeabORD impacts, and right-hand plots show mean values per scenario. The dashed black line represents an impact of zero.



In Tables 5-7 we summarise the mean population-level impacts per scenario for each species, for each of the three impact variables. There are substantial variations between scenarios – this is to be expected, since the scenarios differ substantially in their levels of displacement. Figures 4-6 show broadly positive relationships between levels of displacement (ptdisp) and impacts, but with high levels of noise relative to the strength of these relationships. We can see from Table 5 that the largest mean population-level impacts seen in the SeabORD outputs represent an increase in the chick mortality rate of 0.69% (guillemot), 8.0% (kittiwake) and 0.7% (razorbill). The corresponding increase in the adult mortality rate are 0.1% (guillemot), 0.6% (kittiwake) and 0.1% (razorbill) (Table 6), and the increase in the adult mass rates are 0.1% (guillemot), 0.3% (kittiwake) and 0.13% (razorbill) (Table 7). Note that these are percentage point increases – e.g., an increase of 5%, relative to a baseline rate of 10%, represents a rate of 10 + 5 = 15%. The differences between species in the maximum levels of impact seen in the SeabORD outputs largely reflects differences in the levels of displacement seen within the scenarios for which SeabORD has been run – i.e., in effect, differences in the species in the maximum level of overlap seen between bird distributions and footprints, for the footprints that were considered within the screening exercise.

Table 5. Summary of simulated SeabORD impacts (mean per scenario) on chick mortality for each species, pooled across SPA and separated by SPA. Summary statistics shown are the minimum, lower quartile (Q1), median, mean, upper quartile (Q3), maximum and standard deviation (SD).

Species	SPA	Min	Q1	Median	Mean	Q3	Max	SD
GU	All	-0.0021	-0.0001	0.0010	0.0015	0.0034	0.0060	0.0021
	UK9002271	-0.0004	0.0003	0.0018	0.0019	0.0034	0.0048	0.0018
	UK9004171	-0.0013	0.0002	0.0024	0.0022	0.0042	0.0060	0.0026
	UK9004271	-0.0021	-0.0004	0.0000	0.0001	0.0005	0.0021	0.0013
KI	All	-0.0010	0.0010	0.0067	0.0183	0.0349	0.0795	0.0237
	UK9002271	-0.0010	0.0016	0.0045	0.0039	0.0063	0.0067	0.0028
	UK9004171	-0.0004	0.0039	0.0089	0.0157	0.0185	0.0565	0.0196
	UK9004271	-0.0001	0.0069	0.0491	0.0350	0.0561	0.0795	0.0307
	UK9006101	-0.0006	0.0008	0.0059	0.0181	0.0412	0.0513	0.0230
RA	All	-0.0048	-0.0006	0.0005	0.0012	0.0031	0.0066	0.0029
	UK9002271	-0.0025	-0.0015	-0.0006	-0.0004	0.0007	0.0021	0.0016
	UK9004171	-0.0006	0.0013	0.0025	0.0029	0.0050	0.0066	0.0025
	UK9004271	-0.0048	-0.0004	0.0000	0.0007	0.0025	0.0052	0.0034

Table 6. Summary of simulated SeabORD impacts (mean per scenario) on adult mortality for each species, pooled across SPA and separated by SPA. Summary statistics shown are the minimum, lower quartile (Q1), median, mean, upper quartile (Q3), maximum and standard deviation (SD).

Species	SPA	Min	Q1	Median	Mean	Q3	Max	SD
GU	All	-0.0006	-0.0004	0.0000	0.0002	0.0007	0.0014	0.0006
	UK9002271	-0.0002	0.0000	0.0005	0.0004	0.0007	0.0011	0.0004
	UK9004171	-0.0004	-0.0002	0.0002	0.0004	0.0009	0.0014	0.0007
	UK9004271	-0.0006	-0.0005	-0.0005	-0.0003	-0.0003	0.0008	0.0004
KI	All	-0.0008	0.0000	0.0005	0.0013	0.0021	0.0055	0.0018
	UK9002271	-0.0008	-0.0002	0.0000	0.0001	0.0004	0.0012	0.0006
	UK9004171	-0.0004	0.0000	0.0009	0.0013	0.0018	0.0047	0.0018
	UK9004271	0.0000	0.0003	0.0023	0.0023	0.0036	0.0055	0.0021
	UK9006101	-0.0003	0.0000	0.0005	0.0015	0.0033	0.0046	0.0020
RA	All	-0.0007	-0.0002	0.0006	0.0007	0.0014	0.0027	0.0010
	UK9002271	-0.0005	-0.0003	0.0005	0.0003	0.0007	0.0010	0.0006
	UK9004171	-0.0005	0.0005	0.0010	0.0010	0.0018	0.0022	0.0009
	UK9004271	-0.0007	-0.0003	0.0004	0.0007	0.0016	0.0027	0.0013

Table 7. Summary of simulated SeabORD impacts (mean per scenario) on adult mass loss for each species, pooled across SPA and separated by SPA. Summary statistics shown are the minimum, lower quartile (Q1), median, mean, upper quartile (Q3), maximum and standard deviation (SD).

Species	SPA	Min	Q1	Median	Mean	Q3	Max	SD
GU	All	-0.00024	-0.00008	0.00007	0.00020	0.00045	0.00091	0.00033
	UK9002271	-0.00011	0.00007	0.00038	0.00032	0.00050	0.00084	0.00030
	UK9004171	-0.00020	-0.00007	0.00026	0.00028	0.00059	0.00091	0.00039
	UK9004271	-0.00024	-0.00008	-0.00008	-0.00005	-0.00004	0.00026	0.00013
KI	All	-0.00017	0.00004	0.00020	0.00073	0.00141	0.00304	0.00099
	UK9002271	0.00002	0.00005	0.00018	0.00013	0.00020	0.00022	0.00008
	UK9004171	-0.00017	0.00001	0.00017	0.00052	0.00079	0.00204	0.00078
	UK9004271	0.00004	0.00027	0.00187	0.00138	0.00227	0.00304	0.00119
	UK9006101	-0.00006	0.00001	0.00027	0.00082	0.00192	0.00240	0.00110
RA	All	-0.00008	0.00003	0.00018	0.00023	0.00031	0.00085	0.00024
	UK9002271	-0.00003	0.00003	0.00015	0.00011	0.00019	0.00022	0.00010
	UK9004171	-0.00001	0.00008	0.00025	0.00033	0.00050	0.00085	0.00030
	UK9004271	-0.00008	0.00005	0.00022	0.00022	0.00035	0.00058	0.00024

In Table 8 we investigate the correlations between the potential explanatory variables within the SeabORD runs used to develop the emulator. We can see that there are two blocks of variables that are highly (positively) correlated with each other -

a) a block of three variables that aim to quantify the overall expected level of interaction between the population and windfarm: ptdisp, nords and totalfparea; and
b) a block of three variables that aim to calculate the relative importance of barrier effects: maxsweptangle, fpalignment, pbeyond

There are three variables (meandist2spa, colsize) that are contained in neither of these blocks.

We note that the relatively high levels of correlation between other pairs of variables, and relatively small sample sizes, mean the results should be interpreted cautiously, because it is likely to be difficult to disentangle the effects of different explanatory variables. The emulation models treat ptdisp as the primary variable of interest in each analysis and only consider the effects of other variables insofar as they modify the effects of ptdisp (e.g., via the inclusion of interactions between these variables and ptdisp), because this reflects the structure of SeabORD, and this approach helps in partitioning out the effects of different variables.

Table 8. Correlations between potential explanatory variables within the species-SPA-windfarm scenarios for which SeabORD has been run. For "prey" the mean prey level per scenarios is used. Correlations greater than 0.8 (of either sign) are shown in dark grey, correlations between 0.6 and 0.8 (of either sign) in light grey.

	ptdisp	nords	totalfparea	meandist2spa	maxsweptangle	fpalignment	pbeyond	prey	colsize
Ptdisp	1.00	0.73	0.65	0.02	0.17	0.10	0.18	-0.16	0.05
Nords	0.73	1.00	0.78	-0.01	0.03	0.03	0.03	0.00	0.00
Totalfparea	0.65	0.78	1.00	0.42	0.13	0.10	0.07	-0.15	0.26
meandist2spa	0.02	-0.01	0.42	1.00	-0.28	-0.16	-0.39	-0.24	0.56
maxsweptangle	0.17	0.03	0.13	-0.28	1.00	0.72	0.92	-0.04	-0.20
fpalignment	0.10	0.03	0.10	-0.16	0.72	1.00	0.52	-0.03	-0.11
Pbeyond	0.18	0.03	0.07	-0.39	0.92	0.52	1.00	-0.04	-0.26
Prey	-0.16	0.00	-0.15	-0.24	-0.04	-0.03	-0.04	1.00	0.29
colsize	0.05	0.00	0.26	0.56	-0.20	-0.11	-0.26	0.29	1.00

3.2. Emulation model selection

Table 9 summarises the performance of each of the potential models considered for each of the three response variables for each of the three species, in terms of the percentage of variance explained. It can be seen that the percentage of variation explained is moderate (over 39.7%) for all models (except models of adult mortality for guillemot), even the null model (R1), reflecting the positive relationship between "ptdisp" and each of the three response variables for all species, but that a large proportion of variance remains unexplained. The most complicated model considered, R11, has higher fit than the simpler models, with values of over 46.6% for all response variables and species, but it is to be expected that the model complex models would have the highest R-squared value, since R-squared does not adjust for model complexity.

Note that Table 9 relates to variation in *mean* impacts per scenario. To put this in context, we also considered the percentage of variance in raw SeabORD outputs that can be explained by differences between scenarios (by considering a one-way ANOVA with "scenario"):

- Chick mortality: Guillemot = 43.8%, Kittiwake = 96.2%, Razorbill = 7.3%
- Adult mortality: Guillemot = 17.5%, Kittiwake = 46.2%, Razorbill = 6.3%
- Adult mass loss: Guillemot: 96.5%, Kittiwake = 98.6%, Razorbill = 48.1%

It can be seen that run-to-run variation within each scenario (which arises from a mixture of inherent natural variability and uncertainty in the total prey level parameter) is higher for mortality than for adult mass loss, and much higher for razorbill than for guillemot and kittiwake (probably, at least in part, due to the colony sizes being lower for razorbill, which will lead, all else being equal, to greater natural variability in population-level characteristics). These results indicate that the R-squared values in Table 9 may be substantially over-optimistic, in terms of the total proportion of variation between individual SeabORD runs that is described by the OWF and SPA characteristics. However, we are essentially interested in this WP in investigating variations in mean SeabORD response per scenario, so in this sense the R-squared values in Table 9 are potentially more directly relevant that values based upon the variance of the raw SeabORD outputs.

Percentage of variation explained ("R-squared") is a measure of within-sample performance, so will not capture any loss of performance that arises from generalising the model to scenarios other than those used to fit the model (we try to capture this in subsequent analyses) and therefore tends towards an over-optimistic assessment of performance. It also imposes no penalty for model complexity, so will never assign lower performance when the complexity of a model is increased (e.g., by inclusion of additional parameters). It is nonetheless a widely used and interpretable metric, and it provides a useful initial assessment of absolute performance, particularly in the context where we are interested in understanding sources of variation, but we use a different metric, the Akaike Information Criterion with small sample size correction (AICc) to select between the empirical performance of models. Unlike R-squared, AICc does include a penalty for model complexity, so aims to find the optimal trade-off between fit to the data and model complexity. Low values indicate better fit, and, as it is only useful as a relative, rather than absolute, measure of fit, deltaAICc values are created by deducting (for each species and response variable) the lowest AICc value from all values. The "best" model has a deltaAICc value of zero by definition. Models with a deltaAICc of less than 2 are typically regarded as having comparable performance to the best model, whilst those with a deltaAICc of more than 10 are typically regarded as having very low empirical performance relative to the best model.

Table 10 shows the AICc values for each response variable and each species. In general, there are multiple models with reasonable levels of empirical support for each impact metric and species, indicating that the data (SeabORD runs, in this context) are inconclusive regarding the best model, and the best performing model (that with zero deltaAICc value) varies between both metrics and species. In all cases, the best performing model includes some form of variation (either between SPAs or in relation to SPA-windfarm metrics), but the form of this variation – i.e., the best supported explanatory variable – varies between species and impact measures. The model with consistently reasonable support (e.g., the maximum number of species and impact measures with deltaAICc less than 10) is – the model that allows separate trends per SPA (model R11).

Since there are only three or four SPAs per species it is difficult to disentangle SPA-windfarm interaction variables (e.g., distance to windfarm) from the effects of other differences between SPAs.

The reasons for differences in the effects of "ptdisp" between SPAs and windfarms are likely to be complex. The key point to note is that the explanatory variables that we considered do not entirely capture the characteristics of the SeabORD inputs (or, at least, those inputs that vary between runs) – they are only partial summaries of the inputs (since the full set of inputs would be very high dimensional). This creates a possibility that the parameters of the emulator could need to be different for different colonies.

Table 9. R-squared values (expressed as a percentage of variation explained) associated with each of the multiple regression models that were considered for the mean value of each response variable under each scenario for each species. "p" denotes the number of parameters in each model. Models are expressed in the notation used in R, whereby "+" represent variables that are considered additively and ":" denotes interactions between variables. All models exclude an intercept, and R-squared values are therefore regarded with respect to squared variation in the response variable. All explanatory variables except "ptdisp" are centered and scaled. Models in standard font denote models that can be used to predict for new SPAs, models in italics denote models that cannot (because they require information that would not be available for an SPA at which SeabORD has not yet been run).

		Impact Chick Mortality			Impact Adult Mortality			Impact Mass Loss		
Model structure	**p**	**GU**	**KI**	**RA**	**GU**	**KI**	**RA**	**GU**	**KI**	**RA**
[R1] ptdisp	**2**	54.6	76.8	49.4	21.0	65.9	39.7	46.1	70.5	83.7
[R2] ptdisp + ptdisp^2	**3**	56.1	77.6	55.8	22.5	66.4	41.5	46.1	70.9	84.4
[R3] ptdisp + ptdisp:nords	**3**	59.6	77.2	49.5	21.2	66.1	43.3	48.8	71.2	83.7
[R4] ptdisp + ptdisp:totalfparea	**3**	65.4	77.9	49.9	32.8	66.6	41.9	60.4	71.3	83.7
[R5] ptdisp + ptdisp:meandist2spa	**3**	74.4	77.8	50.6	45.4	66.5	43.0	70.6	70.8	84.1
[R6] ptdisp + ptdisp:maxsweptangle	**3**	58.3	79.8	63.8	21.3	67.6	51.4	47.2	72.7	86.3
[R7] ptdisp + ptdisp:fpalignment	**3**	63.7	79.2	68.6	28.3	66.8	49.3	52.1	72.5	91.4
[R8] ptdisp + ptdisp:pbeyond	**3**	58.8	79.9	60.6	21.6	68.1	47.9	47.7	72.7	84.7
[R9] ptdisp + ptdisp:prey	**3**	54.9	79.1	49.6	22.7	68.6	39.8	46.9	71.7	83.7
[R10] ptdisp + ptdisp:colsize	**3**	58.0	77.7	52.3	45.1	66.3	42.5	60.0	70.7	86.8
[R11] ptdisp:SPA	**4**	67.1	80.4	64.8	59.6	69.9	46.6	69.2	74.0	92.1

Table 10. Delta-AICc values associated with each of the multiple regression models that were considered for the mean value of each response variable under each scenario for each species. Values shown in dark grey are those with a deltaAICc value of less than 2 (typically regarded as comparable empirical support to the best model) and values shown in light grey have a deltaAICc of between 2 and

10 (indicating some empirical support). "p" denotes the number of parameters in each model. Models are expressed in the notation used in R, whereby "+" represent variables that are considered additively and ":" denotes interactions between variables. All models exclude an intercept. All explanatory variables except "ptdisp" are centered and scaled. Models in standard font denote models that can be used to predict for new SPAs, models in italics denote models that cannot (because they require information that would not be available for an SPA at which SeabORD has not yet been run).

		Impact Chick Mortality			Impact Adult Mortality			Impact Mass Loss		
Model structure	**p**	**GU**	**KI**	**RA**	**GU**	**KI**	**RA**	**GU**	**KI**	**RA**
[R1] ptdisp	**2**	14.7	2.6	9.3	14.9	0.4	2.8	15.7	0.3	13.5
[R2] ptdisp + ptdisp^2	**3**	16.2	3.8	8.5	16.9	2.3	4.6	18.2	2.2	15.0
[R3] ptdisp + ptdisp:nords	**3**	13.7	4.4	11.9	17.3	2.6	3.9	16.7	1.9	16.1
[R4] ptdisp + ptdisp:totalfparea	**3**	9.1	3.3	11.7	12.6	2.1	4.5	8.9	1.7	16.0
[R5] ptdisp + ptdisp:meandist2spa	**3**	0.0	3.6	11.3	6.4	2.3	4.0	0.0	2.3	15.4
[R6] ptdisp + ptdisp:maxsweptangle	**3**	14.7	0.3	3.5	17.3	1.1	0.0	17.6	0.0	11.7
[R7] ptdisp + ptdisp:fpalignment	**3**	10.5	1.3	0.0	14.5	2.0	1.0	14.7	0.2	0.0
[R8] ptdisp + ptdisp:pbeyond	**3**	14.3	0.0	5.6	17.2	0.5	1.7	17.3	0.0	14.5
[R9] ptdisp + ptdisp:prey	**3**	17.0	1.4	11.8	16.8	0.0	5.4	17.8	1.3	16.0
[R10] ptdisp + ptdisp:colsize	**3**	14.9	3.7	10.4	6.5	2.5	4.2	9.3	2.5	10.8
[R11] ptdisp:SPA	**4**	10.2	4.5	5.7	0.0	3.8	5.2	4.1	3.5	0.8

3.3. Emulation predictions for other colonies and scenarios

In Figures 4-6 we show the predicted population-level impacts on chick mortality (Fig 4), adult mortality (Fig 5) and adult mass loss (Fig 6) associated with the "null" model (R1) and the best overall model according to AICc (R11). Both of these models assume a linear relationship between each impact variable and the level of displacement, "ptdisp", so predictions are straightforward to generate these are simply equal to the estimated slope parameter (either overall [R1] or SPA-specific [R11]) multiplied by the value of "ptdisp". The standard error on these predictions can be calculated in the same way (multiplying by "ptdisp").

The value of "ptdisp" can readily be calculated for any windfarm scenario and population of interest, by (a) calculating the proportion of the bird distribution map that is within any footprint or within 2km of any footprint (a standard GIS calculation) and (b) multiplying this by the displacement susceptibility rate (often just called the "displacement rate").

We see from Figures 4-6 we see some visual evidence of relationships between the level of the displacement and the mean SeabORD results per scenario, although there is considerable noise, and, for guillemot and razorbill, the noise dominates the fitted relationships. The results for kittiwake show substantial variations in relationships between SPAs, and, for Flamborough and Filey, show high levels of variation within the SPA.

There is no obvious visual evidence for nonlinearity, but this is unsurprising given the high levels of noise, relative to the strength of the relationships, and does not mean that non-linear effects are not present. We would ultimately expect the assumption of linearity will become implausible once levels of displacement become high, so the linear relationships estimated here should be interpreted with caution.

Given the empirical support for SPA-specific relationships, and the relatively small number of scenarios per SPA, the predicted relationships should be treated with a very high degree of caution, especially for the model (R1) that assumes a common relationship in all circumstances. Emulator R1 can in principle be used to produce predictions for new SPAs and new windfarm scenarios, and emulator R11 for new scenarios within each SPA, but the number of SeabORD runs used to develop these emulators, combined with the high levels of noise (unexplained variation) are currently insufficient for it to be advisable to produce predictions in this way.

Figure 4. Scatterplots of proportion of birds displaced per timestep (ptdisp) against simulated SeabORD impacts on chick mortality for each species, with different colours representing different SPAs, with fitted emulation models shown. Left hand plots show raw SeabORD impacts (points) together with predicted values from the pooled mixed model M1 (thick lines) and SPA-specific mixed model M11 (dotted thick lines). Right hand plots show mean values per scenario with predicted values from the pooled regression model R1 (thick lines) and SPA-specific regression model R11 (dotted thick lines). The thick dashed black line represents an impact of zero.

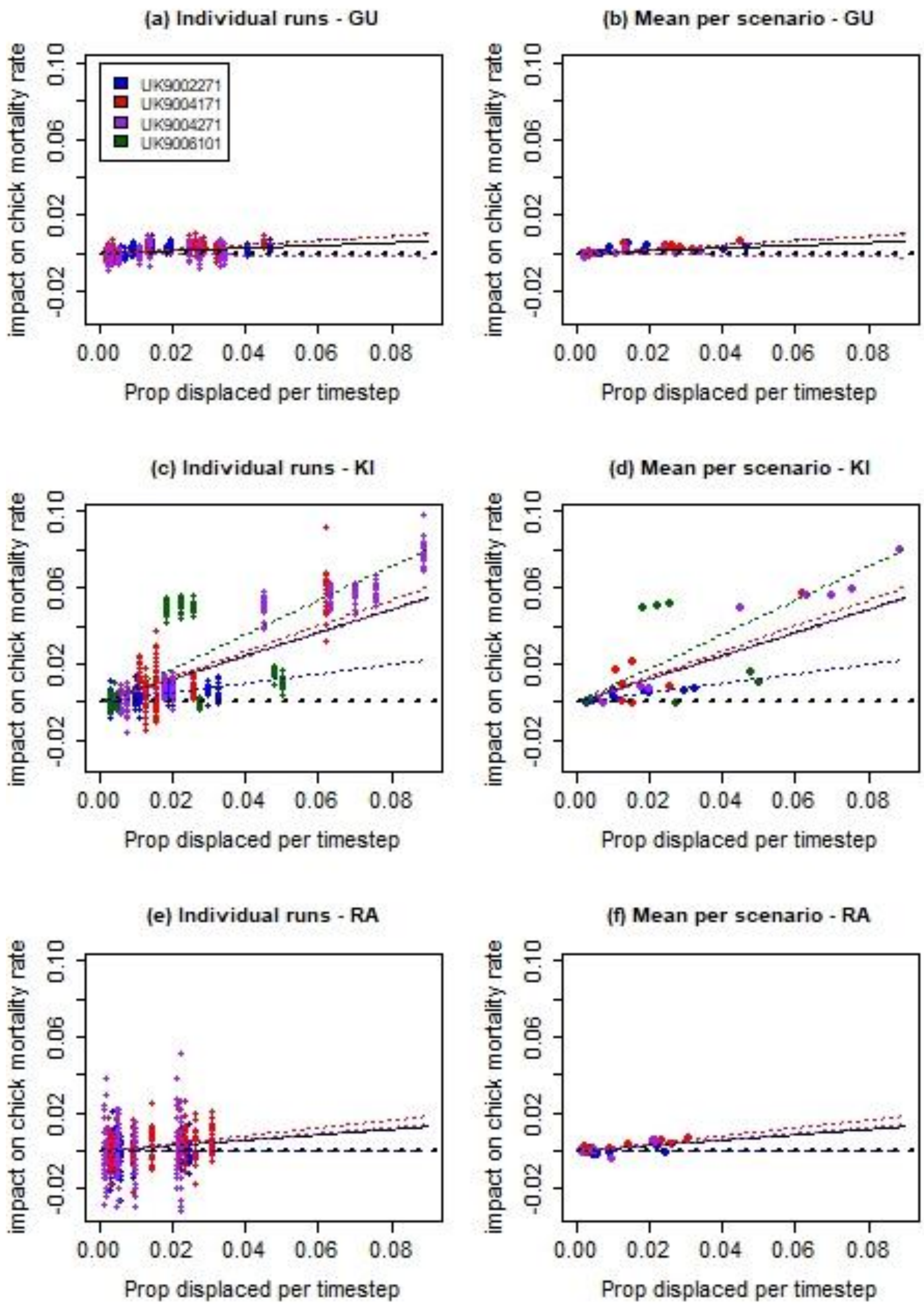


Figure 5 Scatterplots of proportion of birds displaced per timestep (ptdisp) against simulated SeabORD impacts on adult mortality for each species, with different colours representing different SPAs, with fitted emulation models shown. Left hand plots show raw SeabORD impacts (points) together with predicted values from the pooled mixed model M1 (thick lines) and SPA-specific mixed model M11 (dotted thick lines). Right hand plots show mean values per scenario with predicted values from the pooled regression model R1 (thick lines) and SPA-specific regression model R11 (dotted thick lines). The thick dashed black line represents an impact of zero.

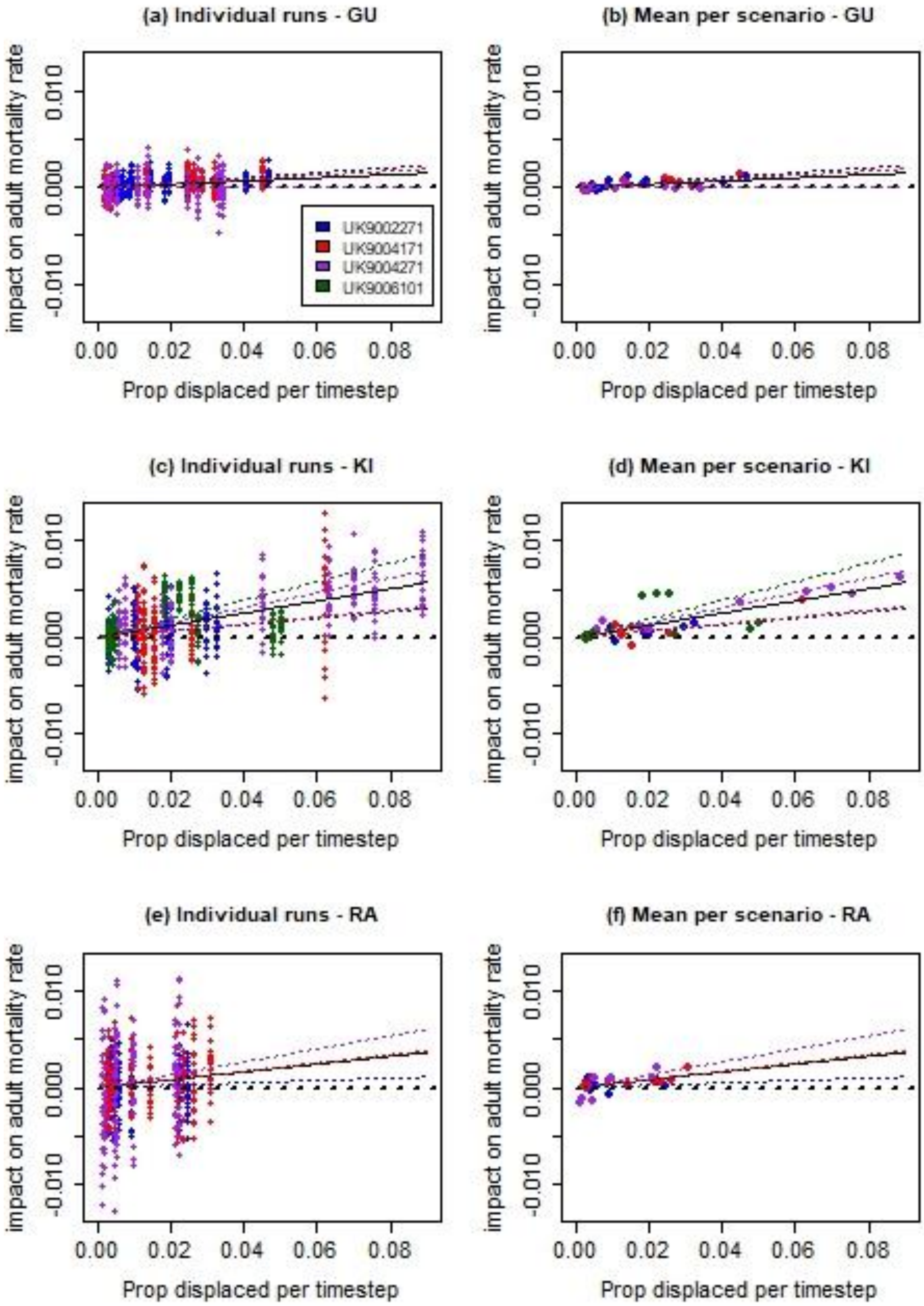
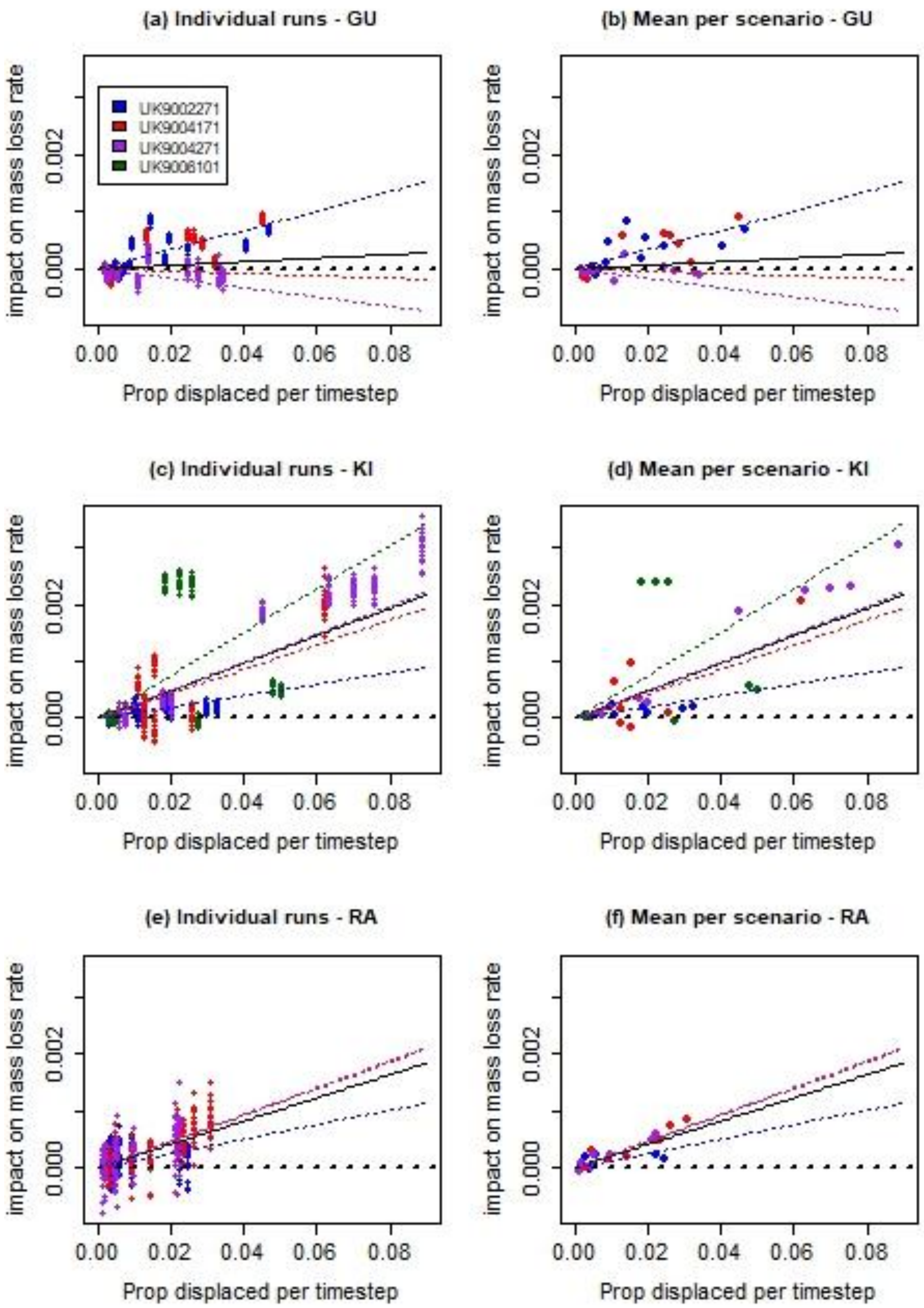


Figure 6. Scatterplots of proportion of birds displaced per timestep (ptdisp) against simulated SeabORD impacts on adult mass loss for each species, with different colours representing different SPAs, with fitted emulation models shown. Left hand plots show raw SeabORD impacts (points) together with predicted values from the pooled mixed model M1 (thick lines) and SPA-specific mixed model M11 (dotted thick lines). Right hand plots show mean values per scenario with predicted values from the pooled regression model R1 (thick lines) and SPA-specific regression model R11 (dotted thick lines). The thick dashed black line represents impact of zero.



3.4. Emulator goodness of fit and cross validation

We examined goodness of fit of the emulator, and explore the potential consequences of lack of fit, using a range of different approaches.

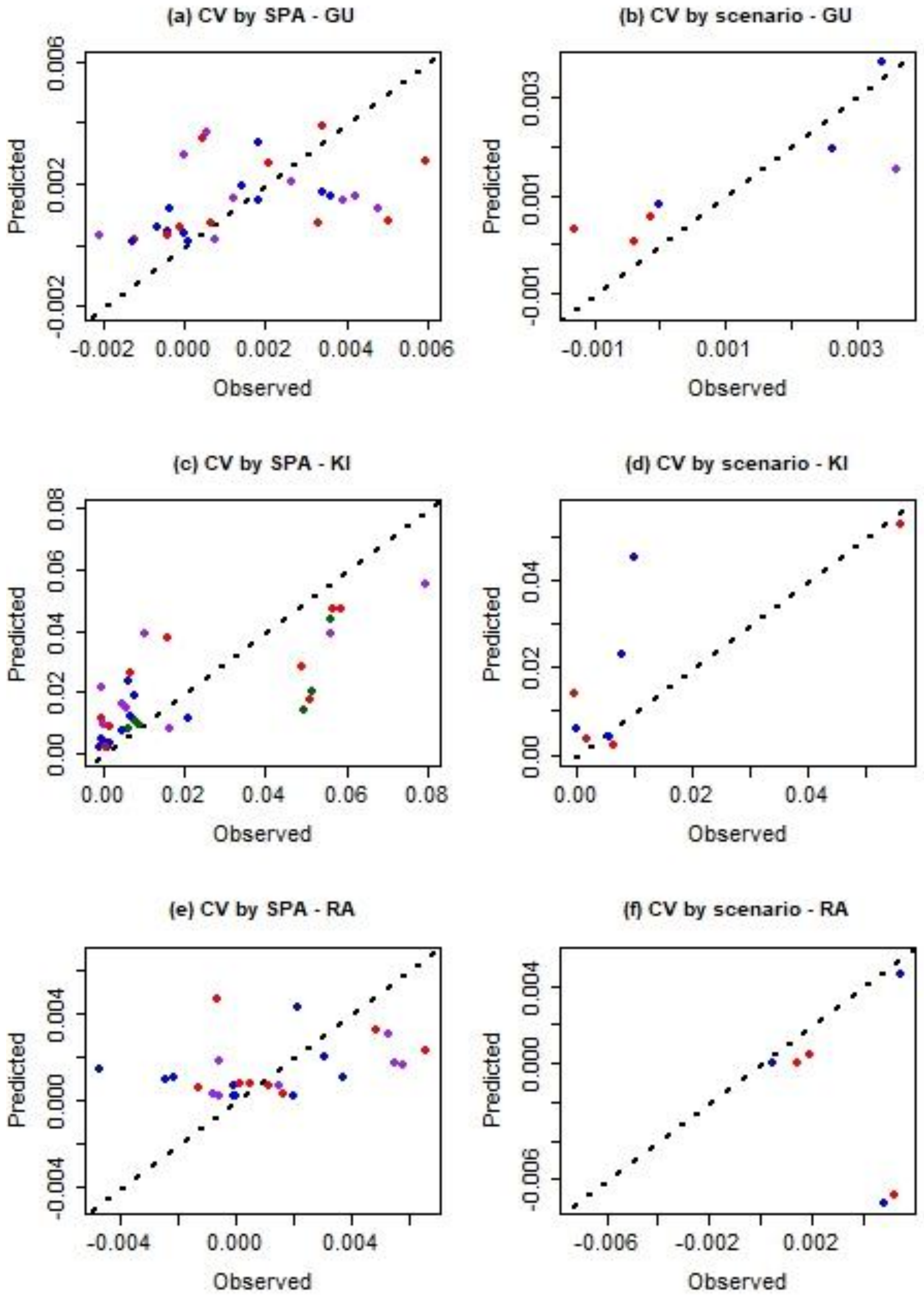
Firstly, we use standard visual diagnostic tools for statistical models to try to detect empirical properties of the data that may be inconsistent with model assumptions, focusing on the models containing all explanatory variables. A common problem with all emulation models was that the QQ-plots (a standard diagnostic tool for examining the goodness of fit relative to the assumed distribution for the response variable, not shown) were quite right-skewed, although this was less pronounced for mass change than for adult and chick survival. A slight right skew was also seen in histograms of the response variables (not shown). Various approaches were investigated in order to try to address this issue. Since values of the response variable can take both negative and positive values (in all cases) many standard approaches to transformation (log, square root, Box-Cox) cannot be applied directly to the "raw data" (i.e., in this context, the SeabORD model outputs). Response variable values were shifted by max(response) to be strictly positive, and log, square root and Box-Cox transformations were then applied using Gaussian GLMs with appropriate link functions. A Gamma GLM was also investigated as a potential alternative. None of these potential alternatives led to any substantive improvements to the Q-Q plots however (which remained right-skewed, or only showed a very slight improvement), and the impact on the percentage of variance explained was negligible. A cube root transformation of the response variables leads to rather greater improvements to the QQ plots, but since the improvements to goodness-of-fit were not substantial, we present other results in relation to the models of untransformed response variables, given that these are more readily interpretable.

Our earlier summaries of model performance of model performance were based on R-squared values, but we noted that these were *within-sample* (i.e., not predictive) assessments of performance. We therefore also use an alternative, cross-validation, approach that involves fitting models to a subset of the data, and then assessing performance against the data that was not included in the subset used for fitting. Because the "test data" are not used to fit the model, this avoids the problem of over-fitting (which will frequently be an issue when, for example, only R-squared is considered).

Within this context, we focus on two possible ways of performing the cross validation: a) we fit the model to 2 of the 3 SPAs, and then test using the remaining SPA, and b) fitting the model to75% of the data (randomly selected) and then testing on the remaining 25% (Figure 4). We focus on the null model, because this model is of importance in relation to displacement mortality rates (WP4) and because the best model according to AICc cannot meaningfully be examined by cross-validation, given the low numbers of observations per SPA.

The results (Figures 7-9) show some positive relationship between observed and predicted values in all cases, but with very high levels of noise around this relationship, indicating relatively poor performance of the models when used to predict data points that have been omitted from model fitting. There are no clear differences between species or impact metrics. Performance is generally worse when cross-validating by SPA than by scenario, reflecting the apparent unexplained differences in impacts between SPAs. Overall, the cross-validation results suggest that the R-squared values obtained above may give an over-optimistic sense of model performance, with the ability of the models to predict for new windfarm scenarios or SPAs being relatively poor.

Figure 7. Results of cross validating the pooled regression model R1 of mean SeabORD impacts of chick mortality per scenario, for each species. Left hand plots involve removing each SPA and then fitting the model using the other two SPAs and using this to predict the response for this SPA, for each of the SPAs. Right hand plots involve randomly removing 25% of the scenarios and then fitting the model using the remaining 75% and using this to predict the response for the omitted 25%, for each of the three SPAs. Colours are used to distinguish between SPAs: colour scheme is as in Figure 6.



to

Figure 8 Results of cross validating the pooled regression model R1 of mean SeabORD impacts of adult mortality per scenario, for each species. Left hand plots involve removing each SPA and then fitting the model using the other two SPAs and using this to predict the response for this SPA, for each of the SPAs. Right hand plots involve randomly removing 25% of the scenarios and then fitting the model using the remaining 75% and using this to predict the response for the omitted 25%, for each of the three SPAs. Colours are used to distinguish between SPAs: colour scheme is as in Figure 6.

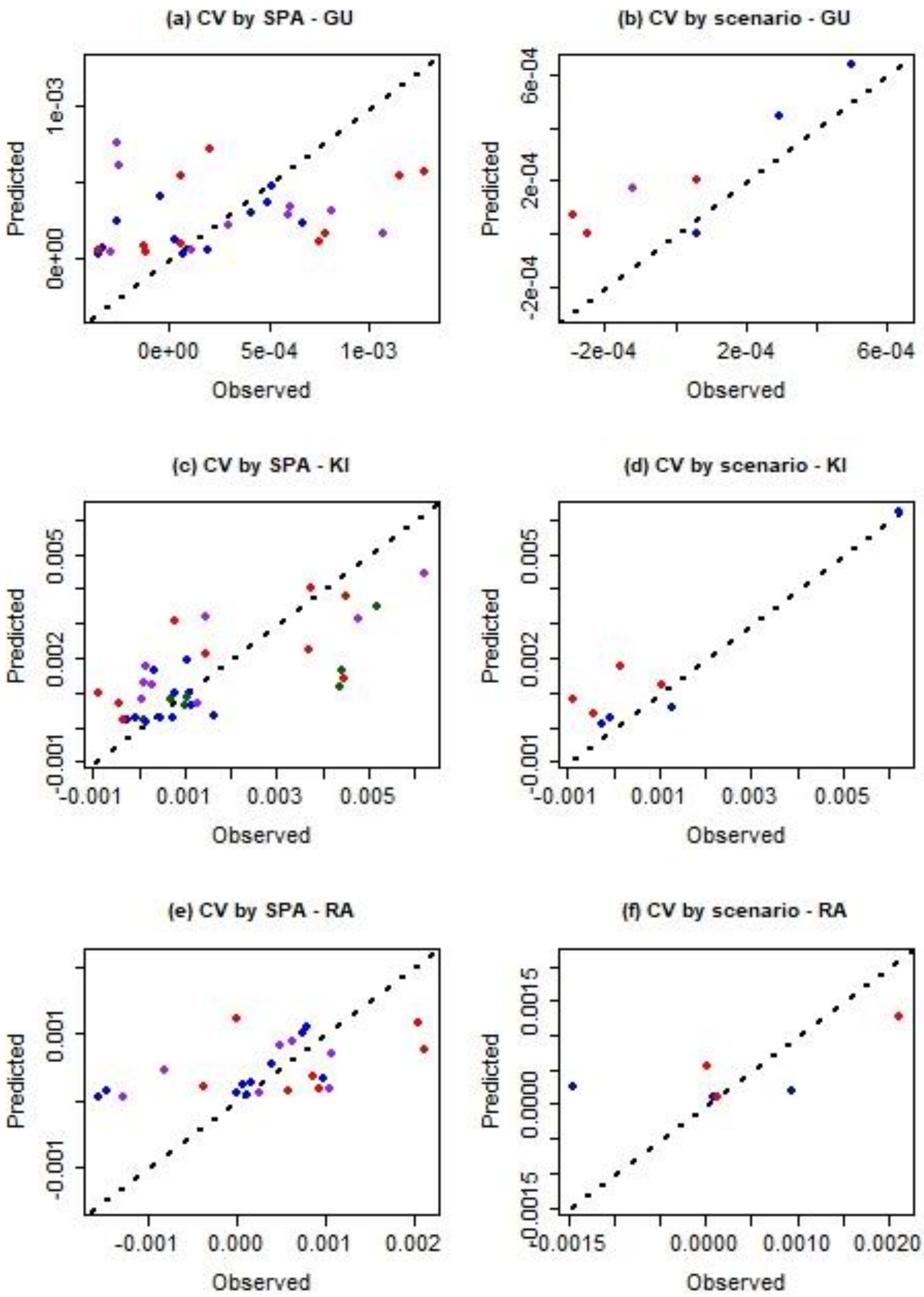
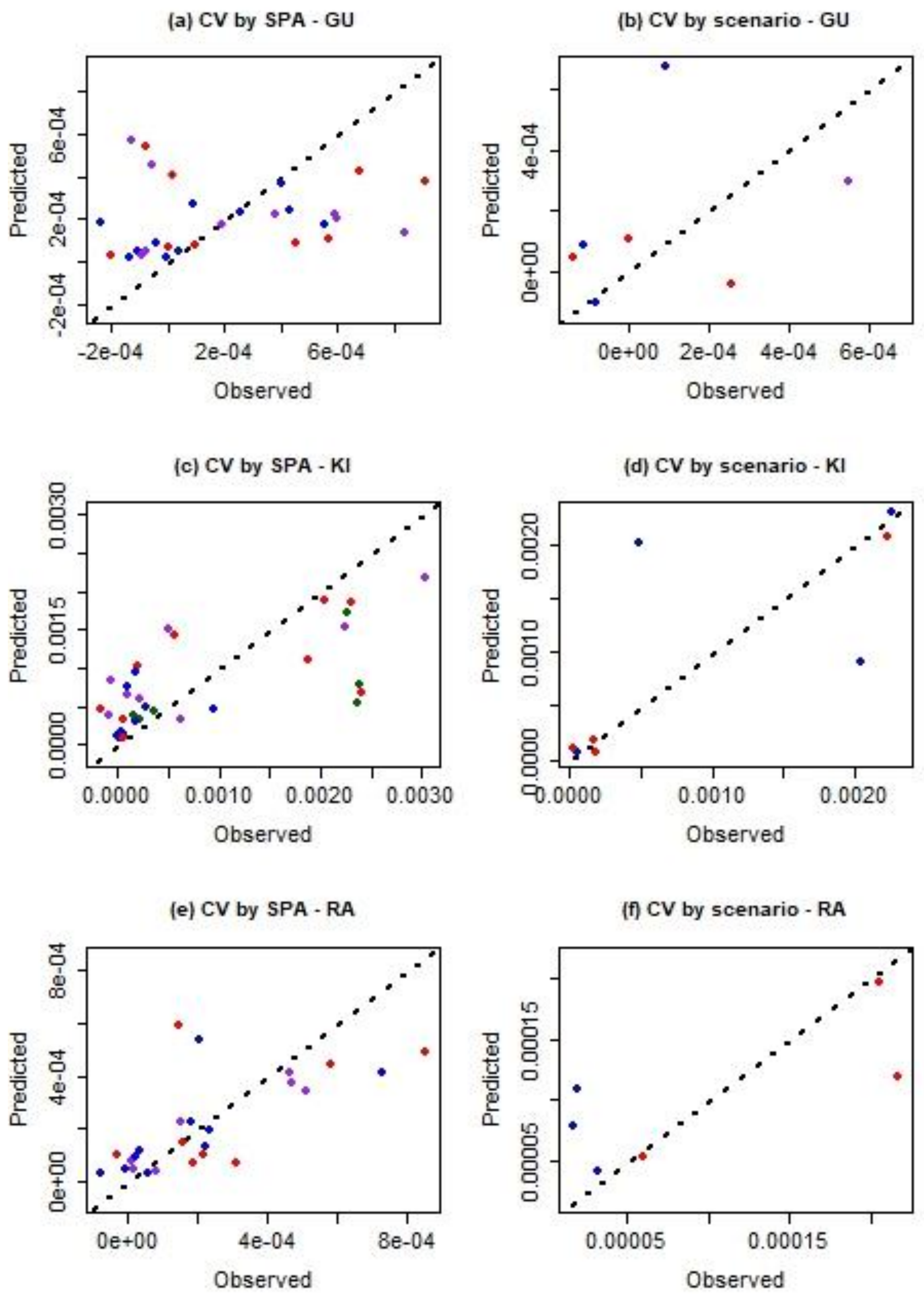


Figure 9. Results of cross validating the pooled regression model R1 of mean SeabORD impacts of adult mass loss per scenario, for each species. Left hand plots involve removing each SPA and then fitting the model using the other two SPAs and using this to predict the response for this SPA, for each of the SPAs. Right hand plots involve randomly removing 25% of the scenarios and then fitting the model using the remaining 75% and using this to predict the response for the omitted 25%, for each of the three SPAs. Colours are used to distinguish between SPAs: colour scheme is as in Figure 6.



4. Discussion

4.1. Key findings

The results need to be interpreted very cautiously, because they are based on a relatively small training set of SeabORD runs (and, in particular, a relatively small number of SPAs), and because of the high levels of noise (inter-run and inter-scenario variation) relative to the strength of the estimated relationships, but the key findings of the modelling were that:

a. The mean simulated impacts SeabORD of windfarm scenarios on adult mass loss, adult mortality and chick mortality show evidence of a positive relationship with the expected proportion of bird displaced per timepoint ("ptdisp"), which can be derived by multiplying the displacement susceptibility rate (often called "displacement rate") by the proportion of the bird distribution that lies within a footprint or within 2km of a footprint
b. Emulators have been constructed for each of the three key SeabORD outputs (impact on adult mass loss, impact on chick mortality, impact on adult mortality) for each of the three species (guillemot, kittiwake and razorbill)
c. A number of potential different emulation models have been considered, including a model that allows a common linear relationship with "ptdisp" in all circumstances, models that description variations in the linear relationship in relation to explanatory variables, models that assume a non-linear rather linear relationship, and models that assume separate relationships for each SPA. The results of model comparison show a lot of variation between both species and metrics, but a lack of consistency in the evidence as to which factors explain this variation.

We investigated a range of different emulation models, but these overall qualitative findings remained consistent across all of the models considered. The high levels of uncertainty regarding the models, and the high levels of unexplained variation, suggest that the results of the emulation cannot currently be used to predict displacement mortality for new SPA or windfarm scenarios.

4.2. Caveats and limitations

There are a number of key caveats and limitations to be aware of when interpreting the results.

Many of the key caveats and limitations around the current work arise from the fact that SeabORD remains a highly computationally intensive model to use, and continues to require manual calibration of prey levels for each population (colony and species) for which it will be run – exploratory work within the Scottish Government Marine Directorate CEF project to try to automate the calibration process have suggested that this is challenging, and not easily achieved, and although this is being explored more fully within the OWEC PrePARED project, such work in ongoing so the need for manual calibration currently remains. The work within this project has also occurred in parallel with the development of SeabORD-R within the Marine Scotland CEF project, which has presented logistical challenges. These constraints have meant that within this project it has only been feasible to run SeabORD for a relatively small number of colonies per species (3-4), scenarios per colony (between 6 and 11, with a mean of 9) and prey levels per scenario when considering wind farm impacts (20). These runs still represent a substantial amount of overall computational effort, but the computational constraints mean that the results need to be treated with caution and interpreted carefully. In particular, the small number of colonies per species means that it has not been possible to thoroughly evaluate the extent to which model-based estimates of displacement mortality rates can generalise between colonies, in terms of the properties of SeabORD itself. The computational constraints have also made it difficult to evaluate the assumptions that underpin the emulator, and thereby the extent to which the emulation results can generalise to colonies and scenarios other than those for which SeabORD has been run.

The results are dependent upon the assumptions that underpin SeabORD, and upon the values of the species-level biological parameters used within the model (which are taken from Searle et al., 2018, Appendix B). Note that the emulation is designed to be an approximation to SeabORD, so all of the assumptions within SeabORD effectively also become assumptions within the emulator.

We have assumed that bird distributions are derived from the maps of Wakefield et al. (2017), scaled up to SPA level. Fairly strong assumptions are required in order to analyse GPS tracking data in such a way as to produce a model that can predict spatial distributions for all colonies (Wakefield et al., 2017), and, for colonies with extensive GPS tracking data, it is likely to be possible to produce more defensible and realistic maps using a colony-specific analysis. A key specific caveat around the use of the Wakefield et al. (2017) maps within this context is the fact that the maps incorporate all behaviours, but are assumed when used here as inputs to SeabORD to relate to foraging locations, which may lead, through the inclusion of locations relating to transit, to a bias towards foraging being simulated to occur closer to the colony than would be the case if foraging-only maps were used. However, the maps of Wakefield et al. (2017) currently provide a feasible approach for colonies that lack GPS tracking data, and avoid the need for colony-specific analyses, so are suited to broad-scale multi-colony analyses such as being conducted here. The approach used to relate the Seabird 2000 subsites used in Wakefield et al. (2017) to the SPAs for which we run SeabORD (i.e., to produce dataset #609 in the CEF Data Store) is provisional, in the absence of an alternative generic way of aligning Seabird 2000 colony definitions with SPA boundaries but has been flagged as a dataset that requires feedback from stakeholders and may be revised within subsequent iterations of the CEF.

The results assume that the SeabORD runs within the scenarios considered for each species span a range of conditions that can be used to estimate general properties of the SeabORD model. For guillemot and razorbill, in particular, this assumption may be problematic, given that the even the highest levels of displacement captured by the scenarios are relatively modest. This occurs even though the process of selecting scenarios was deliberately designed to capture those colony-windfarm combinations with the highest levels of overlap. Possible explanations are likely to relate to a combination of (a) lower spatial overlap with windfarms for these species, (b) characteristics of the spatial distribution maps in Wakefield et al. (2017) and (c) the scenarios only considering those windfarms currently included in the CEF Data Store.

We have assumed here that prey is uniformly distributed across space within the foraging range of each colony due to a lack of availability for prey maps from relevant geographical regions. Work within the OWEC PrePARED project will improve the representation of prey within SeabORD.

The performance of the emulator that we have constructed for SeabORD depends heavily upon the choice of potential explanatory variables to try to capture the characteristics of the model inputs. The key inputs to SeabORD (aside from the biological parameters, which have been held fixed across all runs here, and the prey map, which has been assumed here to be uniform) are the prey level parameter, the colony size, the bird distribution maps, and the footprints. We include the first two of these as potential explanatory variables, and construct seven additional variables that relate to the characteristics of the footprints in relation to the colony and bird distribution map. Some of these variables are defined in relatively *ad hoc* ways, however, and there are clearly many other possible additional metrics that could be constructed.

One potentially very useful metric would be the proportion of individuals per timestep that experience barrier effects, with this is a key mechanism of impact within SeabORD – this was not considered here because we focused on metrics that could readily be calculated for any scenario, and this metric is moderately computationally intensive to calculate, but it would be useful to investigate this further.

We have focused here upon impacts of OWF and SPA characteristics in relation to the level of displacement ("ptdisp"), but it is possible that the effects of these characteristics are also non-additive in other ways, and that interactions between these characteristics may therefore exist. It is also possible that interactions between characteristics and "SPA" may exist, if the impacts of characteristics vary between SPAs in ways that cannot be captured by simple additive and linear relationships.

We have focused here on relatively simple statistical methods for emulation, multiple regression and mixed modelling. This has some important advantages, in relation to ease of implementation and interpretation, but many alternative methods for emulation are available, and there could well be methods that have better performance than multiple regression, particularly in situations where the training set is small. The assumption that the response variables are normally distributed is not particularly realistic, and could potentially be overcome through alternative models – it is difficult to construct a distribution on the wind farm impacts, due to the relatively complicated constraints on the values of these impact variables, so an alternative approach would be to model the baseline and windfarm SeabORD runs separately, and to capture the difference between these within the model.

4.3. Future work

We outline the potential for future work in this area, and, in particular, the potential for future work to overcome some of the caveats and limitations that we have highlighted. We distinguish in this section between (a) future work that is already funded and scheduled to take place within the NERC ECOWINGS project, (b) other future work that is already funded and scheduled to take place within the OWEC PrePARED project, and (c) broader ideas for future work that are not currently funded.

4.3.1. Scheduled and funded future work in ECOWINGS WP4

The ECOWINGS project will build heavily upon the work undertaken within this project, here we identify the elements of future work that are already scheduled to take place within that project. ECOWINGS also aims to use emulation to provide a computationally fast statistical approximation to SeabORD, but it builds on, rather than duplicating the work undertaken here, and aims to extend the work undertaken within this project in four key ways:

1. It is designed to upscale estimates of displacement impacts to a regional scale, so will consider how emulation can be used to scale estimates of displacement mortality up to larger numbers of windfarms than those currently considered in SeabORD scenarios.
2. By conducting runs that cover a more comprehensive set of SPA and windfarm – e.g., by using hypothetical windfarm footprints to allow windfarm-SPA characteristics that are not reflected in actual windfarms-SPA configurations to be considered, allowing the results of the emulation to be generalized more readily.
3. It will involve a more comprehensive comparison and evaluation of different emulation methods.

4. It will focus on an iterative process of model approximation and model refinement and/or simplification, in which the emulator is used to identify ways in which SeabORD can be simplified without substantial loss of accuracy, as well providing an approximation to SeabORD.

Initial work within ECOWINGS is heavily focused on producing outputs on the potential for strategic net gain, so will focus primarily on (a). This will involve re-fitting the current emulation models to a larger training set of SeabORD model runs, which includes a larger set of colonies (thereby removing some of the key caveats around generalisability of the current emulation results), and using the results to develop an emulator that can rapidly predict impacts across a region (the North Sea). As part of the re-fitting there will also be more exploration of emulation model selection and predictive performance, including a more comprehensive assessment of out-of-sample performance (something that was heavily restricted here, due to the small number of colonies used for the SeabORD runs).

The emulation work within ECOWINGS will exploit experience gained through this project by focusing on emulating impacts on adult mass and chick survival and not attempting to directly emulate impacts on adult survival. The mass-survival relationship within SeabORD is a simple statistical relationship, based on published literature, so we will link an emulator of the remaining parts of SeabORD to the mass-survival model, rather than attempting to capture the mass-survival relationship within the model.

Later statistical work within ECOWINGS will investigate the potential to use more sophisticated emulation methods. Work will begin by reviewing the available methods, before investigating the performance of the most promising in providing accurate approximations to SeabORD outputs. Key potential alternative approaches include (a) mixed models that account for the structure of the model runs (and, in particular, the structure of the model in relation to colonies through the inclusion of a “colony” random effect, (b) Gaussian processes, which unlike multiple regression can account for autocorrelation in the unexplained variation within the emulator and (c) machine learning approaches, such as random forests or neural networks, which can have more flexibility in capturing non-linearities.

The emulation research within ECOWINGS will be undertaken through work to identify model simplifications and refinements within SeabORD to improve computational efficiency without substantively compromising accuracy. This strand of work has already begun, through the use of a simplified energetics-based model to estimate displacement risk at a North Sea scale, but this work will become more closely linked to the emulation component of ECOWINGS as the project progresses, so that the two strands of work will inform each other. Within the context of the emulator, there will be further exploration regarding the choice of explanatory variables (e.g., windfarm and colony characteristics), and exploration as to how these relate to quantities used in the simplified mechanistic models of energetics. In particular, metrics relating to the proportion of birds that experience barrier effects will be considered, to see if these metrics are useful in explaining the remaining variability in impacts.

During later stages of ECOWINGS empirical learning from the novel predator-prey data collection and data analysis taking place within the project will also be used to improve the structure of the model, to better reflect the fundamental mechanisms that underpin predator-prey-windfarm interactions.

4.3.2. Scheduled and funded future work in OWEC PrePARED

Work within the OWEC PrePARED project will develop SeabORD to include a new mechanism for the redistribution of prey during construction and operation of offshore windfarms based on new empirical data and modelling in the Forth-Tay and Moray Firth, adaptation of SeabORD to work with the MS sandeel habitat suitability model predictions across the North Sea, development of SeabORD to work with new prey availability maps within the Forth-Tay, and development of SeabORD to work with new joint predator-prey distribution maps within the Forth-Tay. The project will also consider the development of new methods for simulating more realistic foraging tracks within SeabORD based on recent GPS tracking data.

4.3.3. Other potential future work

This work is dependent upon using GPS-based maps, which have been derived by upscaling the maps of Wakefield et al. (2017) to SPA level, to capture the spatial distributions of birds from each colony. Colony sizes are determined via Seabird 2000, as in Wakefield et al. (2017). Seabird 2000 currently provides the most recent census of colony-level abundance but is now more than 20 years old. The new seabird census will shortly supersede this, and there would be advantage in updating the analyses outlined here to reflect these new estimates of colony size. As well as amending the colony sizes that are used directly as inputs to SeabORD, this would also require the analyses of Wakefield et al. (2017) to be updated to incorporate these data. There is also a need to update the analyses of Wakefield et al. (2017) to incorporate more recent GPS tracking data, and to exploit recent developments in spatial statistics. For colonies with extensive tracking data there may be advantages in producing local models of spatial distribution that rely on less strong assumptions than those required for multi-colony analyses that involve a generalisation to untracked colonies.

A more immediate issue lies in the upscaling of the Wakefield et al. (2017) maps to SPA level: the CEF data that provide the spatial linkage between Seabird 2000 subsites and SPAs are provisional, and need to be reviewed by stakeholders, and may potentially need to be updated as a result of this review.

We have focused here on the three SeabORD species for which GPS-based maps are available. The fourth species included in SeabORD, puffin, lacks such maps, due to the relatively scarcity of GPS tracking data. Distance-decay maps can still be produced for puffin, but initial work within this project suggested that the resulting maps were difficult to develop rapidly to produce biologically plausible estimates and such work was beyond the scope of this project. Further work is therefore needed to quantify the spatial distributions for puffin.

The impacts on adult mortality are heavily dependent upon the mass-survival relationship within SeabORD. This relationship represents a key step is translating estimates of displacement effects on behaviour and energetics into the consequences for demography, but it is a non-mechanistic component of SeabORD, and dependent upon published relationships from Oro & Furness (2002) and Erikstad et al. (2009) whose transferability to UK SPAs for the species considered here is not necessarily clear – particularly in the case of guillemot and razorbill, as the mass-survival relationship used for these species within SeabORD is based on a published relationship derived for another species (puffin). More recent analyses (Daunt et al., 2020) could allow these relationships to be updated, in a way that utilises more recent and directly relevant data and more comprehensively captures uncertainty, and there is potential for SeabORD to be updated to use the outputs of these analyses (Searle et al., 2022). In the interim, we recommend that studies using SeabORD report changes in adult mass, as well as changes in adult survival/mortality, to futureproof against improvements to the representation of this relationship within the model.

SeabORD currently relates only to interactions between seabirds and windfarms during the chick rearing period, but it would be possible to extend this to cover the entirety of the breeding season (Searle et al., 2022). The ORJIP DISNBS project, which has just begun, will construct comparable models for the non-breeding season. This is substantially more challenging, because of (a) the relative lack of empirical data on spatial distribution and movement in the non-breeding season and (b) the differing biological constraints during the non-breeding season (the mechanisms within SeabORD are heavily focused around the central place foraging constraint, which is typically much stronger in the breeding season than the non-breeding season).

Calibration is a key step within SeabORD in ensuring that the values of “prey level”, which is a key parameter that lacks biological information and which effectively captures a range of processes not currently represented within the model, are specified so as to produce baseline results for productivity, adult mass loss and other metrics during the chick rearing period that are consistent with empirical data. The approach taken to calibration is currently manual and uses a visual assessment to select the plausible range of prey values, so there would be benefits in moving to a more automated and systematic approach, such as Approximate Bayesian Computation (ABC) or History Matching.

SeabORD also contains a range of other biological parameters, whose values are derived from the published literature or from expert judgement. Uncertainty in these parameter values is not currently accounted for, but in order for SeabORD outputs to provide a more complete quantification of uncertainty it would be valuable to also consider uncertainty in these parameters. Expert elicitation may provide a mechanism to capture the level and form of uncertainty in each of these parameters.

Finally, there is potential to extend SeabORD to include other species (Searle et al., 2022).

4.4. Conclusions

This work has illustrated the potential to use SeabORD as a way of estimating wind farm impacts on adult mass loss, and adult and chick mortality, and of estimating variability in these rates between colonies and wind farm scenarios. It has also demonstrated the potential to use statistical methods to provide an approximation to SeabORD, thereby enabling displacement mortality to be estimated for a much larger number of scenarios than would be computationally feasible within SeabORD. Emulation methods also provide a useful tool for studying the properties of mechanistic models like SeabORD, and, in this case, have revealed useful insights regarding the model.

The results of the emulation work- should be treated with substantial caution, given the high levels of noise relative to the strength of the fitted relationship, and the current results do not provide a sufficient basis for predicting displacement mortality for new SPAs or windfarms.

The work has also highlighted that, although emulation can overcome some of the computational challenges imposed by the use of a mechanistic individual-based model such as SeabORD, it is still fundamentally constrained by the computational time required to run the mechanistic model. This is because the emulator needs to be trained using a reasonably large number runs of the mechanistic model. A key constraint lies in the ability to run SeabORD for large numbers of colonies, due to the continuing need to calibrate SeabORD manually for each colony. The results presented here should therefore be interpreted cautiously because the limited number of scenarios considered may mean that the results do not generalise to SPAs or windfarms for which SeabORD has not yet been run – this is particularly true because the model selection and cross validation indicate evidence for variations in the behaviour of SeabORD between SPAs (presumably due to differing bird distribution maps and differing forms of interaction between these maps and the windfarm footprints).

SeabORD, and emulators of SeabORD, provide model-based estimates of displacement mortality rates that can be compared against those obtained via expert elicitation (WP2), and are helpful in highlighting explainable sources of variation in these rates – this is explored further in WP4. The results obtained here can, in principle, provide a basis for quantifying displacement mortality rates, and variations in these rates in relation to colony and windfarm characteristics. The results obtained here tentatively suggest that SeabORD-based variations in displacement mortality within and, especially, between, colonies cannot be explained solely using the nine colony and windfarm characteristics considered within this emulator, suggesting that the emerging properties of SeabORD are complex and subtle, which is to be expected from a large process-based mechanistic model. As such, the results obtained here are useful in sense-checking, and providing context for, the results of the expert elicitation (EE), and we explore this further in WP4 by considering how the SeabORD and emulation results may be consistent with the definition of displacement mortality rate used in the EE. Neither the EE nor model-based results are derived from empirical estimates of displacement mortality, however, so both sets of results need to be interpreted cautiously.

Work within the ECOWINGS project will further extend the work undertaken here and overcome some of the limitations of the current work by using a larger training set of SeabORD runs and considering more sophisticated emulation methods.

References

Erikstad, K.E. Sandvik, H., Fauchald, P. & Tveraa, T. (2009) Short- and long-term consequences of reproductive decisions: an experimental study in the puffin. Ecology, 90: 3197-3208

Daunt, F, Z Fang, R Howells, M Harris, S Wanless, KR Searle and D Elston (2020) Improving estimates of seabird body-mass survival rates. *Scottish Marine and Freshwater Science Vol 11 No 13. DOI: 10.7489/12329-1*

Oro, D. & Furness, R.W. (2002) Influences of food availability and predation on survival of kittiwakes. Ecology 83: 2516-2528

Searle, K. R., E. L. Jones, M.I. Bogdanova, D. Elston, Z. Fang, Daunt, F. & A. Butler (2022) Study to examine the feasibility of extending SeabORD to the entire breeding season. *Marine Scotland Science* https://www.gov.scot/publications/study-examine-feasibility-extending-seabord-entire-breeding-season/

Searle, K.R.; Butler, A.; Mobbs, D.; Trinder, M.; Waggitt, J.; Evans, P.; Daunt, F. (2020) Scottish Waters East Region Regional Sectoral Marine Plan Strategic Ornithology Study: final report (Report No. NEC07184). *Report by Natural Environment Research Council (NERC). Report for Marine Scotland Science.* https://www.gov.scot/publications/scottish-waters-east-region-regional-sectoral-marine-plan-strategic-ornithology-study-final-report/

Searle KR, D C Mobbs, A Butler, R W Furness, M N Trinder and F Daunt. (2018) Finding out the Fate of Displaced Birds. *Scottish Marine and Freshwater Science Vol 9 No 8. DOI: 10.7489/12118-1*

Searle KR, A Butler, D Mobbs, M Bogdanova, S Wanless, M Bolton and F Daunt. (2015) At-Sea Turnover of Breeding Seabirds. *Scottish Marine and Freshwater Science Volume 6 Number 10. DOI: 10.7489/1622-1*

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.

SNCBs (2022) Joint SNCB Interim Displacement Advice Note. https://data.jncc.gov.uk/data/9aecb87c-80c5-4cfb-9102-39f0228dcc9a/joint-sncb-interim-displacement-advice-note-2022.pdf

Wakefield, E.D., Owen, E., Baer, J., Carroll, M.J., Daunt, F., Dodd, S.G., Green, J.A., Guilford, T., Mavor, R.A., Miller, P.I., Newell, M.A., Newton, S.F., Robertson, G.S., Shoji, A., Soanes, L.M., Votier, S.C., Wanless, S. and Bolton, M. (2017), Breeding density, fine-scale tracking, and large-scale modeling reveal the regional distribution of four seabird species. Ecol Appl, 27: 2074-2091. https://doi.org/10.1002/eap.1591

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: 4th Floor, Dorset House, 27-45 Stamford Street, London SE1 9NT.

Published in the UK: 2025