CrossMark

OFFSHORE WIND FARM IMPACTS

# Assessing seabird displacement at offshore wind farms: power ranges of a monitoring and data handling protocol

Nicolas Vanermen · Thierry Onkelinx · Pieter Verschelde · Wouter Courtens · Marc Van de walle · Hilbran Verstraete · Eric W. M. Stienen

**Abstract** Prior to the construction of an offshore wind farm at the Belgian Thorntonbank, local seabird abundance was studied by means of ship-based surveys. 'Seabirds at sea' count data, however, exhibit extreme spatial and temporal variation, impeding the detection of human impacts on seabird abundance and distribution. This paper proposes a transparent impact assessment method, following a before–after control–impact design and accounting for the statistical challenges inherent to 'seabirds at sea' data. By simulating a broad range of targeted scenarios based on empirical model coefficients, we tested its efficacy in terms of power and investigated how the chance of statistically detecting a change in numbers is affected by data characteristics, monitoring period and survey intensity. Because of high over-dispersion and/or zero inflation, the power to detect a 50% decrease in numbers was generally low, but did reach 90% within less than 10 years of post-impact monitoring for northern gannet (*Morus bassanus*) and common guillemot (*Uria aalge*).

N. Vanermen (✉) · T. Onkelinx · P. Verschelde · W. Courtens · M. Van de walle · H. Verstraete · E. W. M. Stienen
Research Institute for Nature and Forest, Kliniekstraat 25, 1070 Brussels, Belgium
e-mail: nicolas.vanermen@inbo.be

## Introduction

In order to meet the targets set by the European Directive 2009/28/EC on renewable energy, the Belgian government reserved an offshore area of 238 km$^2$ for the production of electricity. Current plans are to construct seven wind farms, comprising an estimated number of 530 turbines. In 2008, C-Power installed the first six wind turbines (30 MW) at the Thorntonbank, located 27 km offshore, followed by the construction of 48 more turbines throughout 2012 and 2013 (295 MW) at the same location. A before–after control–impact (BACI) monitoring programme was implemented to assess the changes in seabird abundance following the wind farm construction at the Thorntonbank. This area hosts a bird community dominated by offshore species such as common guillemot (*Uria aalge*), razorbill (*Alca torda*), northern gannet (*Morus bassanus*) and black-legged kittiwake (*Rissa tridactyla*). Also relatively high densities of little gull (*Hydrocoloeus minutus*) and Sandwich tern (*Sterna sandvicensis*), both listed in the Annex I of the Bird Directive (2009/147/EC), are present during the migration seasons (Vanermen et al., 2013).

🖉 Springer

Because of high mobility and clustering behaviour, 'seabirds at sea' (SAS) data are characterised by a high proportion of zeros and a minority of strongly varying positive numbers (Maclean et al., 2013; Leopold et al., 2013). Further to this, SAS data tend to be subject to spatial dependence and/or temporal auto-correlation (Schneider & Duffy, 1985; Schneider, 1990; Pebesma et al., 2000; Huettmann & Diamond, 2006; Karnovsky et al., 2006; Pérez-Lapeña et al., 2010). This paper proposes a method accounting for these statistical challenges. We tested its efficacy by performing an extensive power analysis on SAS count data collected prior to wind farm construction in the impact area as well as in a carefully delineated control area, the power being the chance that a certain change in seabird numbers is identified as statistically 'significant'. By means of a broad range of targeted data simulations, we examined how this chance is affected by species-dependent data characteristics, monitoring period and survey intensity. While impact level thresholds are (in)directly defined by the legislator, a power analysis may inform whether these impact levels can in fact be statistically detected after a pre-set period of time. It further allows the environmental researcher to maximise power when designing a monitoring programme and may help to learn if and how an on-going programme can be improved.

## Methods

### BACI monitoring set-up

In this study, the impact area is considered to be the zone where effects of turbine presence can be expected, and was delimited by surrounding the Thorntonbank wind farm area by a buffer zone of 3 km (Fig. 1). This buffer distance is based on the avoidance distances as found for scoters (*Melanitta* sp.) and long-tailed ducks (*Clangula hyemalis*) during the Danish research project at the Nysted offshore wind farm (OWF) (Petersen et al., 2006). Next, a control area was delineated, harbouring comparable numbers of seabirds and showing a similar range in water depth and distance to the coast (Vanermen et al., 2006, 2010). The distance between control and impact area was chosen to be small enough to be able to count both areas on the same day by means of a research vessel, thus minimising variation resulting from short-

term temporal changes in seabird abundance and observation conditions.

The first turbines were erected in the course of 2008, and the reference period includes all available data prior to April 2008. Starting in 2005, 3 years before the installation of the first turbines, the study area was surveyed almost monthly by counting seabirds along a fixed monitoring route, which could be completed in 1 day and covered the control as well as the impact area. We also included historical survey results collected before 2005, provided that both areas were visited on the same day. As such, 66 surveys were included in the reference dataset—with two counts per survey (one for each area) this results in a sample size ($N$) of 132 (Tables 1, 2). The mean effort per area per survey was 5.5 km$^2$ (SD 3.6 km$^2$).

For the analyses, we used data on 12 seabird species occurring regularly in the Thorntonbank wind farm area, these being northern fulmar (*Fulmarus glacialis*), northern gannet, little gull, common gull (*Larus canus*), lesser black-backed gull (*Larus fuscus*), herring gull (*Larus argentatus*), great black-backed gull (*Larus marinus*), black-legged kittiwake, Sandwich tern, common tern (*Sterna hirundo*), common guillemot and razorbill.

### Ship-based seabird counts

Ship-based seabird counts were conducted according to a standardised and internationally applied method, combining 'transect counts' for birds on the water and 'snapshot counts' for flying birds, applying a transect width of 300 m (Tasker et al., 1984). Seabird surveys were only executed during favourable conditions, defined by good visibility, a calm to moderate wind force (<6 Bft) and a significant wave height of less than 2 m.

Seabird counts are usually aggregated per 2–10 min tracks. However, to avoid the potential auto-correlation between subsequent counts and to minimise overall variance, we aggregated the count data per area (control–impact) and per monitoring day, resulting in day totals for both zones. Taking into account the distance travelled, these day totals can be transformed to seabird densities.

### Data modelling

When a counted subject is randomly dispersed, count results tend to be Poisson distributed, in which the
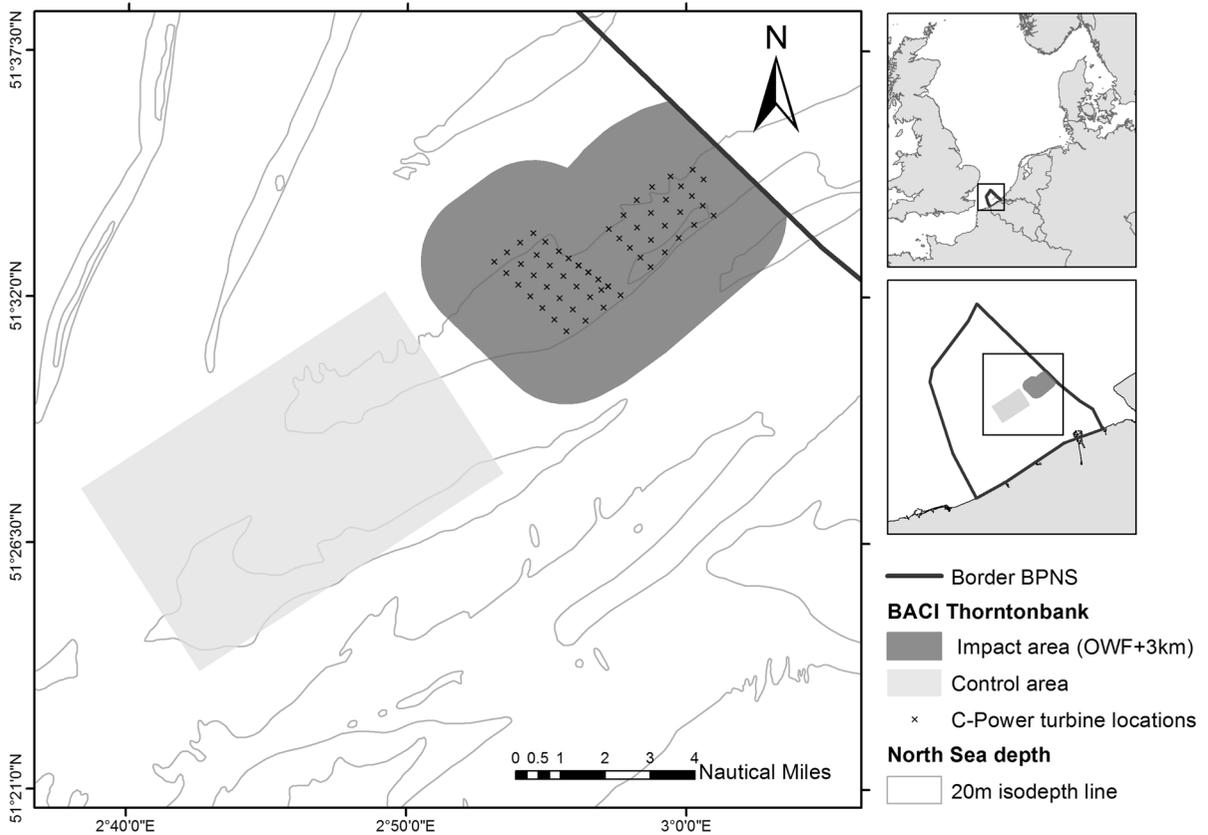
**Fig. 1** Control and impact area at the Thorntonbank study area

mean equals the variance (McCullagh & Nelder, 1989). Seabirds, however, often occur strongly aggregated in (multi-species) flocks, typically resulting in count data with a high proportion of zeros, relatively few but sometimes very large positive numbers and a high variance exceeding the mean. Such over-dispersed data can be analysed by a generalised linear model with a quasi-poisson (QP) or a negative binomial (NB) distribution (Ver Hoef & Boveng, 2007). Here we applied a NB distribution as this distribution is to be preferred over a QP in case of high over-dispersion (Zuur et al., 2009). The variance in a negative binomial distribution is a function of the mean ($\mu$) and the dispersion parameter theta ($\theta$):

$$V(\mu) = \mu + \frac{\mu^2}{\theta}. \tag{1}$$

In case data appeared to exhibit more zeros than can be predicted by a NB distribution, a zero inflated NB (ZINB) model was used (Potts & Elith, 2006; Zeileis et al., 2008). This type of model consists of two parts:

(1) a 'count component' modelling the data according to a NB distribution and (2) a 'zero component' modelling the excess in zero counts.

The response variable (Y) equals the total number of birds observed (inside the transect and during snapshot counts) per survey in either the control or the impact area. To account for varying monitoring effort, the number of $km^2$ counted was included in the model as an offset-variable. Whether data were collected in the control or impact area was included in the model by the factor variable CI. Lastly, the continuous variable month (m) was used to model seasonal fluctuations by fitting a cyclic sine curve, described by a linear sum of sine and cosine terms (Stewart-Oaten & Bence, 2001; Onkelinx et al., 2008), as for example in

$$\log(Y) = \text{offset}\left(\log\left(km^2\right)\right) + a_1 + a_2 \cdot \sin\left(2\Pi \frac{m}{12}\right)$$
$$+ a_3 \cdot \cos\left(2\Pi \frac{m}{12}\right) + a_6 \cdot \text{CI}. \tag{2}$$

In Eq. 2, seasonality is modelled applying a single sine curve with a period of 12 months. In this case, the

**Table 1** Number of surveys and number of km² monitored per month in the control and impact area, before wind farm construction in April 2008

| Month | Control | | Impact | |
|-------|---------|---------|--------|---------|
| | $n$ | $\Sigma$ of km² | $n$ | $\Sigma$ of km² |
| Jan | 4 | 13.6 | 4 | 25.3 |
| Feb | 11 | 80.3 | 11 | 55.1 |
| Mar | 8 | 24.9 | 8 | 38.8 |
| Apr | 3 | 9.3 | 3 | 32.4 |
| May | 2 | 7.7 | 2 | 16.6 |
| Jun | 3 | 8.9 | 3 | 27.0 |
| Jul | 6 | 30.5 | 6 | 34.5 |
| Aug | 6 | 23.2 | 6 | 35.4 |
| Sep | 5 | 35.0 | 5 | 34.4 |
| Oct | 7 | 28.9 | 7 | 30.8 |
| Nov | 6 | 25.3 | 6 | 41.0 |
| Dec | 5 | 29.0 | 5 | 34.1 |
| Total | 66.0 | 316.3 | 66.0 | 405.6 |

**Table 2** Number of surveys per year in the control area and impact area, before wind farm construction in April 2008

| Year | Control | Impact | $\Sigma$ of km² |
|------|---------|--------|-----------------|
| 1993–2004 | 38 | 38 | 355.9 |
| 2005 | 14 | 14 | 210.5 |
| 2006 | 7 | 7 | 86.9 |
| 2007 | 6 | 6 | 60.9 |
| 2008 | 1 | 1 | 7.6 |

amplitude of the seasonal variation equals the square root of the sum of squares of coefficients $a_2$ and $a_3$. For certain species, however, seasonal variation might be captured better by adding another sine curve with a period of 6 or 4 months, on top of the one with a period of 12 months, thus allowing to model more than one peak in density per year, as for example in

$$\log(Y) = \text{offset}\left(\log\left(\text{km}^2\right)\right) + a_1 + a_2 \cdot \sin\left(2\Pi\frac{\text{m}}{12}\right)$$
$$+ a_3 \cdot \cos\left(2\Pi\frac{\text{m}}{12}\right) + a_4 \cdot \sin\left(2\Pi\frac{\text{m}}{6}\right)$$
$$+ a_5 \cdot \cos\left(2\Pi\frac{\text{m}}{6}\right) + a_6 \cdot \text{CI}. \qquad (3)$$

In general, this method performed better compared to the inclusion of 'month' as a factor variable, as this

splits the data in 12 subsets and resulted in highly unreliable coefficient estimates. We did not account for interaction between CI and seasonality since differences in seasonal patterns are not expected to occur at such a small spatial scale.

When applying a ZI model, the zero-component was limited to an intercept ($b_1$), which is linked to the response $p_{ZI}$ (the additional chance of encountering no birds) by a logit function.

Data simulations

In order to be able to perform power analyses, we simulated BACI datasets with predefined characteristics (i.e. the model parameters as found by modelling the reference dataset), but imposing a hypothetical change on the post-construction numbers of the impact area.

First, we modelled the reference data applying one and the same *base model* for all species, being a ZINB model including a 12-month period sine curve and the factor variable CI (count component equalling Eq. 2). This exercise revealed empirical ranges of the maximum density, the amplitude of the seasonal variation, the area effect, the NB dispersion parameter theta ($\theta$) and the amount of zero inflation. We then defined 137 scenarios varying all of these parameters within their empirical range. At this stage, the monitoring set-up characteristics were held constant, with a reference and impact period of both 5 years, one survey per month with an effort of 10 km² per area, a post-construction decrease in numbers of 50% and a significance level of 10%. We simulated 1,000 datasets for each scenario.

In a second step, we performed data simulations based on the coefficient estimates of 12 *species-specific models*, obtained by testing for zero inflation, selecting the best model fit for seasonality and testing whether or not to include the area factor. Models were selected through backward model selection, using a Wald (in case of ZI models) or a likelihood ratio test and considering the respective AIC values. This time, we varied monitoring set-up characteristics, namely the decrease in numbers in the impact area to be detected (25, 50 and 75%), the monitoring period (5 years before versus 1, 3,…, 15 years after impact) and the level of effort per area per survey, i.e. 5, 10 and 20 km². Again, each scenario was simulated 1,000 times.

## Power analysis

For each scenario, we ran a specific impact model on every one of the 1,000 simulated datasets. The impact model is a simple extension of the model used for data simulation by the factor variables BA (before–after turbine construction) and/or T (turbine absence–presence), depending on whether the factor variable CI was retained in the data simulation model. Hence, two general types of impact model were used:

$$\log(Y) \sim \text{seasonality} + \text{BA} + \text{CI} + \text{BA:CI}, \quad (4)$$

$$\log(Y) \sim \text{seasonality} + \text{BA} + \text{T}. \quad (5)$$

In Eqs. 4 and 5, the change in seabird numbers expected to be due to wind farm presence is estimated by the coefficient of the interaction term BA:CI and the factor variable T, respectively. As the power reflects the chance of statistically detecting a certain change in seabird numbers, power was estimated by the percentage of model outputs with a significant $P$ value for the BA:CI or T term. In this study, we applied a significance threshold $\alpha$ of 0.10, being twice the conventional 0.05, which will be reasoned in the discussion.

## Statistics

All data handling and modelling were performed in R.3.0.1 (R Core Team, 2013a), making use of the following packages: lmtest (Zeileis & Hothorn, 2002), MASS (Venables & Ripley, 2002), reshape (Wickham, 2007), pscl (Zeileis et al., 2008; Jackman, 2011), foreign (R Core Team, 2013b) and RODBC (Ripley & Lapsey, 2013).

## Results

### Base models: empirical coefficient estimates

Applying the same *base model* (see Eq. 2) to all species provided us with empirical coefficient ranges. Based upon these, the amplitude was varied by setting $a_3$ to zero and varying $a_2$ by 0, 1, 2, 3 and 4. The maximum natural log abundance (intercept $a_1$ + amplitude $a_2$) of the count component was varied from $-1$ to 3, again in discrete steps of 1 (see Fig. 2, left panel). These are all untransformed coefficient values, linked to the response Y by the natural log, and natural log

abundances of $-1$ and 3, for example, correspond to densities of 0.4 and 20.1 birds/km$^2$, respectively.

The base models further revealed a correlation between theta and the amount of zero inflation. For data showing no zero inflation ($b_1 \leq -8$), theta was rather low, varying between 0.18 and 0.68. Otherwise, for data subject to zero inflation ($b_1 > 0$), theta values were clearly higher, ranging from 0.41 to 1.33, suggesting that the over-dispersion was partly captured by the zero-component. With regard to data simulations, we therefore combined a $b_1$ of $-10$ ($\sim p_{ZI} = 0\%$) with a theta of 0.2, 0.4 and 0.6, and a $b_1$ of 1 ($\sim p_{ZI} = 73\%$) with a theta of 0.6 and 1.2 (see Fig. 2, right panel).

To investigate the role of zero inflation and the possible benefit of using a ZINB model versus its non-ZI analogue, we varied $b_1$ by $-10$, 0 and 1, corresponding to $p_{ZI}$ values of 0, 50 and 73%, for a limited selection of four scenarios ($a_1 = \{-1, -2\}$, $a_2 = \{1, 2\}$, $a_3 = 0$, $\theta = 0.6$), and compared power values obtained by either a NB or a ZINB model.

Until now, the area factor CI was not included in the data simulation, and the impact model applied for power analysis thus followed Eq. 5. Base models, however, revealed the area coefficient to vary between $-0.98$ and 1.27, and we calculated the effect of varying $a_6$ with $-1$, 0 and 1, again for a limited selection of four scenarios ($a_1 = \{-1, -2\}$, $a_2 = \{1, 2\}$, $a_3 = 0$, $b_1 = -10$, $\theta = 0.6$).

Combining all of the above parameters, we ended up with 137 theoretical scenarios, enabling us to isolate and explore the effect of these model parameters on the power of the impact analysis.

### Power analysis: effect of model parameters

Zero inflation had a strong negative effect on the power of the impact study. Figure 3 shows the power results obtained for a scenario with a theta of 0.6 ($a_1 = -1$, $a_2 = 2$, $a_3 = 0$, $\theta = 0.6$, $\alpha = 0.10$). When inducing zero inflation by setting $b_1 = 0$ and $b_1 = 1$, power dropped with 33 and 45% compared to the value of 82% in the non-ZI scenario ($b_1 = -10$). Despite overall power being low, the ZINB model performed slightly better than the NB model when considering ZI data (Fig. 3).

The dispersion parameter theta is inversely related to the data variance (Eq. 1) and strongly influenced the power of the impact analysis (Fig. 4, scenario $a_1 = -1$, $a_2 = 2$, $a_3 = 0$, $\alpha = 0.10$). For the non-ZI

**Fig. 2** Untransformed parameter estimates as found by modelling the reference data of 12 seabird species collected at the Thorntonbank study area (*cross symbol*) and the model coefficients used for data simulations (*filled circle*)
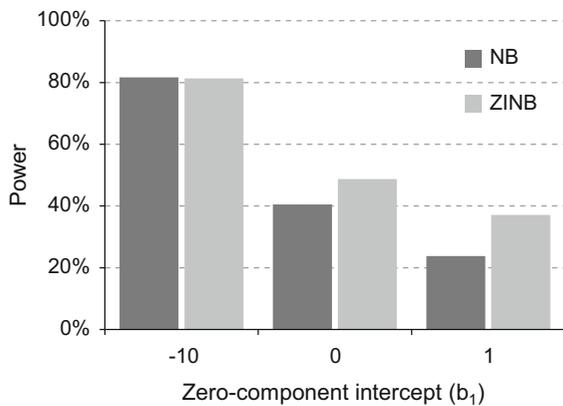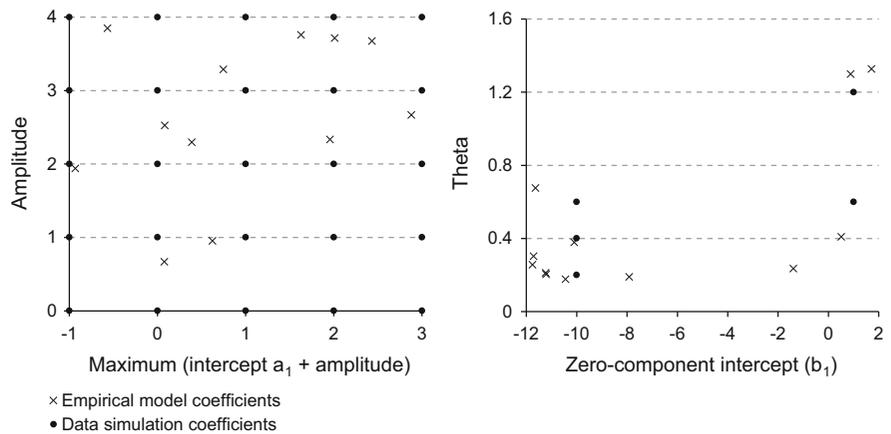


× Empirical model coefficients
● Data simulation coefficients



**Fig. 3** Effect of zero inflation on the power to detect a 50% decrease in numbers, either using a NB or a ZINB model ($a_1 = -1$, $a_2 = 2$, $a_3 = 0$, $\theta = 0.6$, $\alpha = 0.10$)
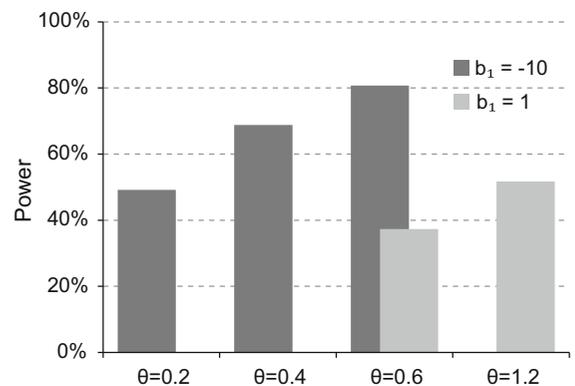


**Fig. 4** Effect of the dispersion parameter theta on the power to detect a 50% decrease in numbers, for scenarios with zero inflation ($b_1 = 1$, ZINB impact model) and scenarios without zero inflation ($b_1 = -10$, NB impact model) ($a_1 = -1$, $a_2 = 2$, $a_3 = 0$, $\alpha = 0.10$)

scenarios, a theta of 0.2 or less inevitably resulted in low power (30–54%), while applying a theta of 0.6 resulted in power values of 47–91% after 5 years of post-impact monitoring. As mentioned, base modelling showed that the reference data of several species combined zero inflation with a higher theta. The positive effect of a higher theta as demonstrated above, however, could not compensate for the loss of power due to zero inflation, and power values for all ZI scenarios remained well below 70%.

When simulating constant numbers throughout the year (amplitude = 0), the effect of abundance on the power appeared to be limited (see Fig. 5, $\theta = 0.6$, $b_1 = -10$). More precisely, power ranged between 82 and 91% when assuming a theta of 0.6 and a constant natural log abundance between $-1$ and 3 (corresponding to densities of 0.4 and 20.1 birds/km², 

respectively). However, when increasing the amplitude from 0 to 4, power values decreased by 15–35%, the effect of amplitude being strongest for the scenario with lowest abundance. Hence, despite accounting for seasonality in the models, seasonal variation negatively affected power, most probably due to prolonged periods of absence or low abundance.

Finally, we investigated the effect of the area factor CI and simulated datasets with varying CI coefficients ($a_6$). When there is no statistical difference in the numbers of seabirds occurring in the control and impact area during reference years, the area factor CI can be excluded from the models, thus gaining a degree of freedom (compare for example Eqs. 4 and 5). Results informed that scenarios without an area effect showed better power, e.g. a value of 82%
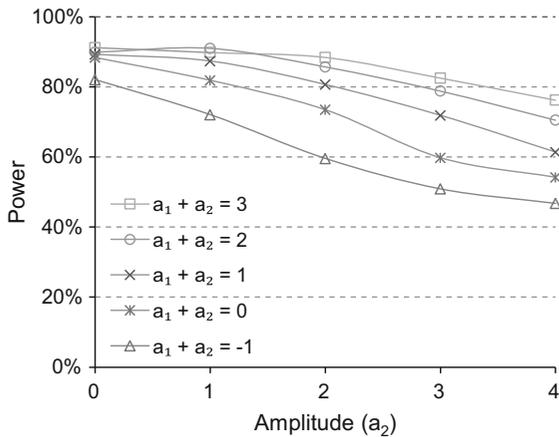
Fig. 5 Effect of the seasonal amplitude ($a_2$) on the power to detect a 50% decrease in numbers, for various levels of maximum natural log abundance ($a_1 + a_2$) (NB impact model, $a_3 = 0$, $\theta = 0.6$, $b_1 = -10$, $\alpha = 0.10$)



Fig. 6 Effect of the area factor CI ($a_6$) on the power to detect a 50% decrease in numbers (NB impact model, $a_1 = -1$, $a_2 = 2$, $a_3 = 0$, $\theta = 0.6$, $b_1 = -10$, $\alpha = 0.10$)

compared to power values of 55 and 61% for the same scenario but with an area effect of $-1$ and $+1$, respectively (see Fig. 6, scenario $a_1 = -1$, $a_2 = 2$, $a_3 = 0$, $\theta = 0.6$, $b_1 = -10$, $\alpha = 0.10$).

Species-specific models: empirical coefficient estimates

Having isolated and analysed the effects of the data characteristics, we built species-specific models, taking into account species-dependent seasonality patterns and area effects. Table 3 shows all coefficients obtained through backward model selection and Fig. 7 illustrates the seasonally varying model predictions for four seabird species.

For eight out of twelve species, model selection resulted in a double sine curve to fit seasonal variation, while the occurrence of four species (all gulls) was best described applying a single sine curve with a period of 12 months. We retained a significant area effect for only one species, i.e. common gull. Backward model selection showed that significant zero inflation prevailed in the data of Sandwich tern ($p_{ZI} = 53.0\%$) and common tern ($p_{ZI} = 77.0\%$). For the latter, the ZINB model fit showed a very high theta of 24.6 suggesting that most of the over-dispersion was accounted for by the covariates and the zero-component intercept. As a result, model selection eventually led to a ZI Poisson (ZIP) model. In all other species, zero inflation was very close to 0%, and the
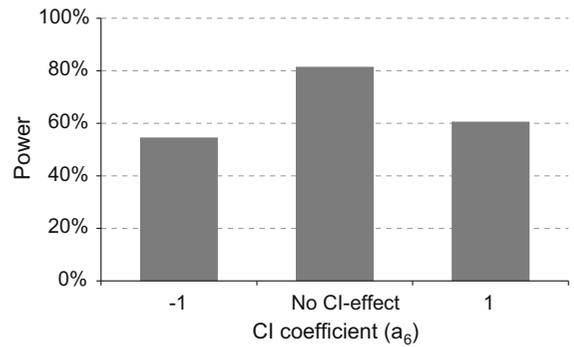
zero-component could be excluded from the model, thereby continuing with a NB model.

Interestingly, all gull species were characterised by low theta values ranging between 0.17 and 0.27, reflecting their often highly aggregated numbers. On the other hand, the more homogenously distributed common guillemot and razorbill showed clearly higher theta values of 0.83 and 0.55, respectively.

Power analysis: species-dependent effects of monitoring period and survey intensity

By applying species-specific models, we found that for none of the 12 seabird species under study, we will be able to detect a decline of 25% with a power of more than 60%, not even after 15 years of post-impact monitoring (Fig. 8). However, a 50% change in numbers should be detectable within less than 10 years with a chance of >90% for northern gannet and common guillemot. The models of these two species showed a favourable combination of moderate over-dispersion ($\theta > 0.3$) and no zero inflation, together with a high peak abundance (6.5 birds/km$^2$) in case of common guillemot, and lower peak abundance (1.9 birds/km$^2$) but year-round presence in case of northern gannet (see Fig. 7). Within the same timeframe of 10 years of post-impact monitoring, we should be able to detect a decrease in numbers of 75% with a power of >90% in all species, except for common gull. The reference data collected for the latter species were indeed far from ideal, showing a significant area effect and a low theta of 0.2 (Table 3).

**Table 3** Estimated (untransformed) model coefficients and theta values of the species-specific models

| Species | Model type | Count component | | | | | | | | | Zero-component |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Intercept | Sin (12m) | Cos (12m) | Sin (6m) | Cos (6m) | Sin (4m) | Cos (4m) | CI | Theta ($\theta$) | Intercept |
| Northern fulmar | NB | −1.75 | −0.93 | 0.14 | −1.18 | 0.07 | | | | 0.11 | |
| Northern gannet | NB | −0.85 | −0.61 | 0.26 | −0.60 | −0.53 | | | | 0.34 | |
| Little gull | NB | −2.42 | 0.99 | 2.32 | | | | | | 0.17 | |
| Common gull | NB | −4.41 | 1.98 | 3.30 | | | | | 1.27 | 0.20 | |
| LBB gull | NB | −0.39 | 0.67 | −2.17 | | | −1.19 | 0.90 | | 0.27 | |
| Herring gull | NB | −2.75 | 1.74 | 0.80 | | | | | | 0.19 | |
| GBB gull | NB | −1.80 | −0.41 | 2.65 | | | −0.68 | 0.70 | | 0.19 | |
| BL kittiwake | NB | −0.62 | −1.04 | 1.78 | | | | | | 0.25 | |
| Sandwich tern | ZINB | −9.22 | 0.45 | −11.39 | 1.17 | −6.72 | | | | 4.57 | 0.12 |
| Common tern | ZIP | −10.04 | −1.37 | −13.03 | −1.07 | −6.92 | | | | | 1.21 |
| Common guillemot | NB | −3.99 | 1.77 | 7.51 | −1.42 | −2.00 | | | | 0.83 | |
| Razorbill | NB | −6.71 | 0.43 | 9.36 | −0.48 | −3.48 | | | | 0.55 | |

**Fig. 7** The modelled densities of four seabird species, based on SAS count data collected in the Thorntonbank study area, prior to the construction of the wind farm (<04/2008)
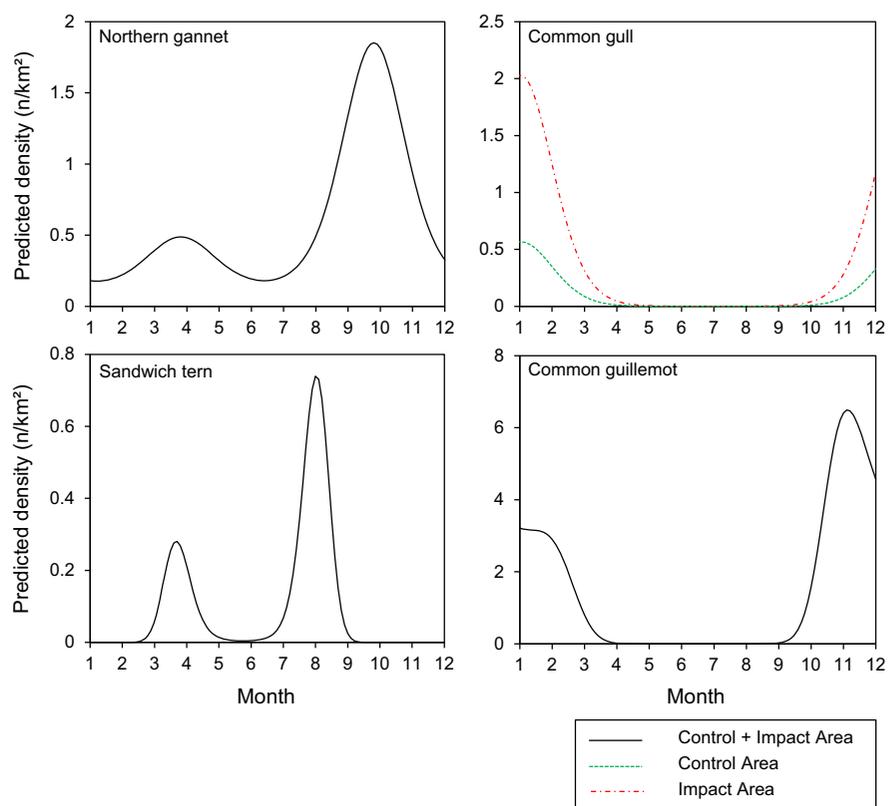


In order to investigate the effect of survey intensity on the statistical power, we ran simulations and subsequent analyses imposing efforts of 5, 10 and 20 km$^2$ per area per survey. Power was positively correlated with effort, but the effect was highly species-specific. Variation in power to detect a 50%

change in numbers after 10 years of impact monitoring resulting from varying effort (5–20 km$^2$) was very small for the common and moderately over-dispersed common guillemot (increase by only 2%), while for most other species, power improvement was between 4 and 10%. The two species subject to zero inflated data, however, showed a strong improvement in power after 10 years of monitoring resulting from the increased effort, i.e. 69 versus 93% for Sandwich tern and 56 versus 97% for common tern.

## Discussion

Since data on seabirds at sea are typically characterised by high over-dispersion and strong spatio-temporal variation, impact studies may be confronted with low power for detecting changes in abundance (Maclean et al., 2013). Meanwhile a lot of effort, time and money is spent on monitoring programmes, aiming to assess the effects of anthropogenic activities on the numbers and presence of seabirds. In order to test the
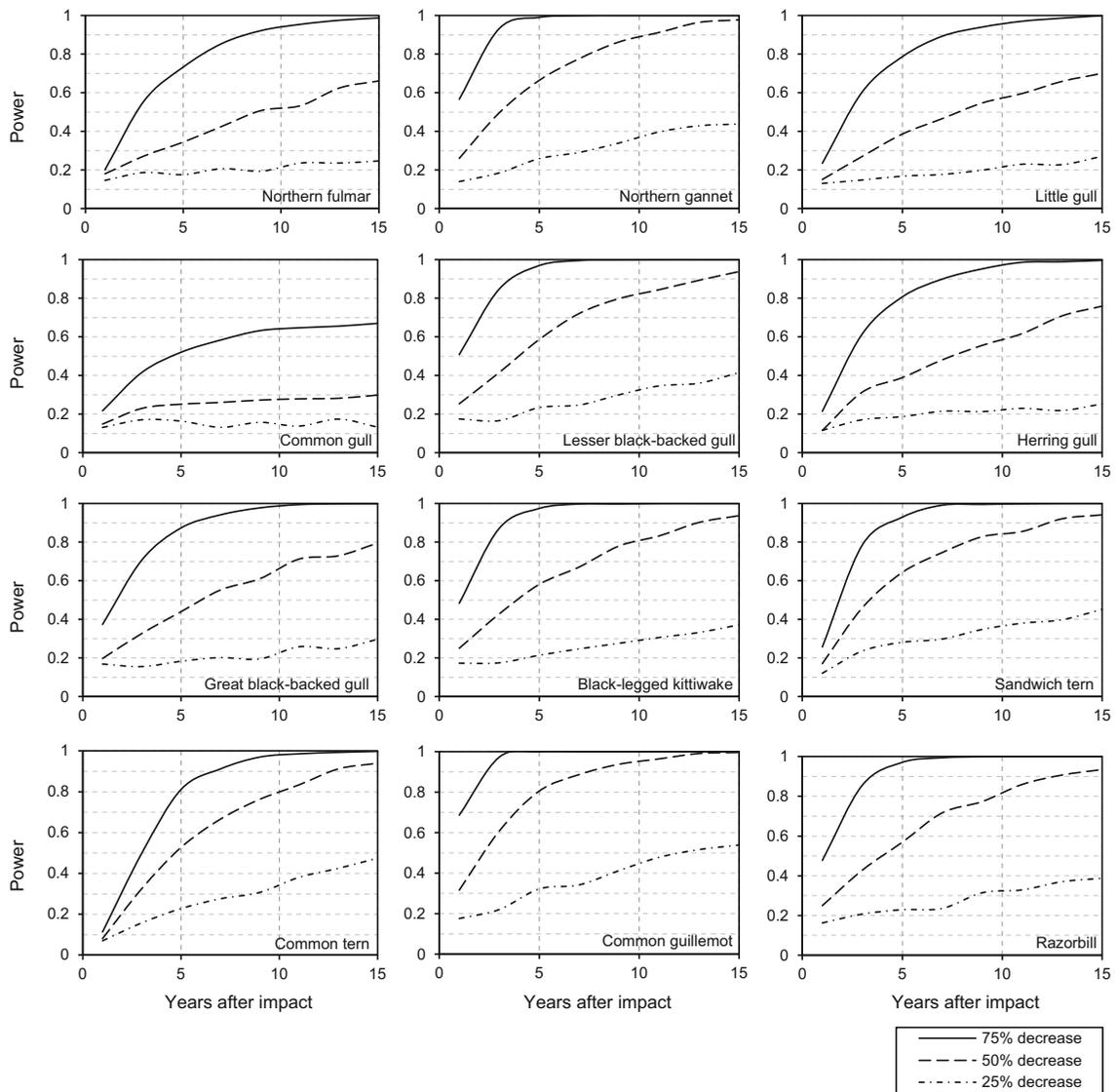


**Fig. 8** Power results for 12 seabird species for an impact study with one survey of 10 km$^2$ per month per area and 5 years of reference monitoring ($\alpha = 0.10$)

efficacy of a BACI monitoring programme designed to detect seabird displacement following OWF constructions at the Belgian part of the North Sea, we analysed the SAS data collected at the Thorntonbank during reference years, and performed a twofold power analysis. First, we ran the same ZINB model on all species to determine empirical coefficient ranges. These ranges were used to investigate how and to which extent data characteristics may influence power. Next, we determined species-specific models and varied monitoring set-up parameters, i.e. the decrease in numbers to be detected, the monitoring period and the amount of effort per area per survey.

Our study showed that in general, the SAS data collected at the Thorntonbank are of poor statistical value because of zero inflation and/or high over-dispersion. Power was further reduced when reference data appeared to be subject to a significant area effect, implying that bird densities were significantly different between impact and control area already before impact. On the other hand, for a simulated decrease in numbers of 50%, power did exceed the 90% threshold level for two species, being northern gannet and common guillemot, after 7 and 10 years of post-impact monitoring, respectively. Maclean et al. (2013) conducted a comparable study on long-time series of aerial survey count data of four seabird taxa (scoters, northern gannet, divers and Sandwich tern) collected in UK waters, which also resulted in low power results. For a monitoring set-up with 12 surveys per year, statistical power to detect a 50% change in bird numbers after 4 years of monitoring stayed well below 80% for all species under study, even for a significance level of 0.20. The authors state that the primary reason for low power in analysing SAS data is the fact that seabird numbers fluctuate greatly at any given location, and that the only way likely to improve the power would be for some of this variance to be explained, by for example hydrodynamic data.

The power of any statistical analysis is known to be affected by (1) the variance in measurements, (2) the sample size and (3) the probability of a type I error ($\alpha$) (Underwood & Chapman, 2003). All three aspects will be discussed in the next paragraphs.

Being inversely related to the data variance (Eq. 1), low theta values were demonstrated to have a highly negative effect on the power of the impact analysis. While theta is actually supposed to reflect species-specific aggregating behaviour, with low values

indicating strong spatial aggregation in numbers, it is important to note that any unexplained temporal and spatial variability will result in a decreasing theta, demonstrating the importance of modelling as much biologically relevant variation as possible. To reduce the variation in seabird count results in this study, the impact and reference areas were always counted on the same day, thus avoiding day-to-day variation in seabird abundance and observation conditions. Seasonal variation was modelled through a cyclic sine curve, and differences between both areas were accounted for by a factor variable. Despite seabird surveys being performed in favourable circumstances only, weather and observation conditions are still likely to affect the detection probability of inconspicuous seabird species, and accounting for this through multivariate distance analysis (Marques & Buckland, 2003) might reduce yet unexplained temporal variance. In this study, we carefully modelled seasonality, but variation between years was not accounted for. Possibly, the inclusion of the year as a random variable might further reduce temporal variance. Also seabird numbers have been demonstrated to relate to tidal state (Schwemmer et al., 2009; Embling et al., 2012; Cox et al., 2013; Scott et al., 2013), and the inclusion of this variable is another unexplored but possibly rewarding route towards a better power.

When count data collected along a fixed monitoring route are aggregated to day totals per area, one no longer needs to account for within-area spatial variability, while between-area variability should be fully captured by the area factor CI. Further to this, lumping the data to day totals avoids the potential auto-correlation between subsequent transect counts. While the applied count unit in SAS research is usually the result of a 2–10 min track (Tasker et al., 1984), these rather short and consecutive transect counts are likely to be pseudo-replicates (Schneider, 1990; Huettmann & Diamond, 2006; Karnovsky et al., 2006). Ignoring 'auto-correlation' between count data would result in an underestimation of the standard errors, in turn increasing the risk of wrongly identifying an effect of the wind farm on bird numbers ('type I error'—Pérez-Lapeña et al., 2010).

Apart from its simple design and transparent modelling approach, the monitoring design presented here is distinguished by the small scale of the study area, as opposed to research programmes carried out in the Danish, Dutch and UK waters (Petersen et al.,

2006; Leopold et al., 2013; Walls et al., 2013; Petersen et al., 2014). In a BACI set-up, changes observed in the impact area are weighed against the trends observed in the control area in order to come to a decision on displacement, assuming seabird densities in the control area are representative of those in the wider area and are not affected by any other (uncontrolled) source of variation. A small-scale study area offers the advantage of monitoring to be more cost- and time-effective. Moreover, it offers the possibility to delineate a control area comparable to the impact zone in terms of environmental characteristics and expected seabird densities, thus reducing between-area variability in the collected data. On the other hand, the smaller the control area, the more likely the observed trends will not adequately reflect the true and wider scale trends in seabird numbers. While we can predefine the tolerance for type I and type II errors, it is important to bear in mind that within the proposed monitoring set-up, this source of uncertainty stays beyond the control of the researcher. Furthermore, in cases where habitats are much more complex, or where other anthropogenic impacts impede a correct assessment of the 'background' densities of seabirds, it may prove impossible to delineate a suitable control area. In this case, a before–after monitoring of a wide area including the OWF and spatially explicit modelling of the resulting data is preferred over a classic BACI design (Mackenzie et al., 2013).

As mentioned, power is enhanced by a larger sample size, thus by counting longer or more frequently (Underwood & Chapman, 2003; Pérez-Lapeña et al., 2011; Maclean et al., 2013). This also means that the time needed to reach a certain power can be halved by performing two monitoring surveys each month instead of just one, provided that surveys are sufficiently spread over time to avoid temporal autocorrelation. In contrast, for most species, doubling the effort by counting 20 km$^2$ per survey per area (instead of 10 km$^2$) gives only moderate power improvement. This is due to the fact that model coefficients are assumed to remain constant, and to be independent of the time spent in the study area. This is probably a fair reflection of reality provided that effort is high enough to assure sufficient encounters with the species of interest, which might not be the case for scarce species or seabirds characterised by a strongly clustered distribution. For such species, spending more time in the study area increases the chance of an actual

observation, opposed to yet another zero count. Indeed, the only species for which a strong effect of effort on power was found were both subject to zero inflation and low abundance. On the other hand, increased effort might, for example, temper the data variance, in turn positively affecting theta. By assuming constant model coefficients, this is not accounted for in this study, and it would therefore be very interesting to investigate how the count effort per survey relates to the variation and robustness of the parameter estimates.

By far the easiest way to enhance the power of any impact analysis is to apply a higher significance threshold ($\alpha$). In this context, $\alpha$ equals the chance of wrongly concluding that the turbines are causing an impact, while in fact they are not ('type I error'). However, a stringent significance level goes at the expense of power, resulting that certain impact effects may go unnoticed ('type II error') (Underwood & Chapman, 2003). For decision-making, ecological studies commonly set the threshold value for a type I error ($\alpha$) to 5%, and for a type II error ($\beta$) to 20%. This choice, however, tends to be arbitrary and implies that the acceptable risk of committing a type II error is four times higher than the risk of a type I error (Pérez-Lapeña et al., 2011). This is contradictory to the fact that most impact studies are meant to function as an early warning system and to detect potential negative effects as soon as possible. In this paper, the risk of making a type I error $\alpha$ was therefore relaxed to 10% (instead of the conventional 5%), while we use 90% as a boundary for 'sufficient' power ($1 - \beta$), thus equalling the tolerance for both risks ($\alpha = \beta$). Clearly, it would even be better for these values to be determined by predefined management objectives (Pérez-Lapeña et al., 2011) and to set the values for $\alpha$ and $\beta$ based on economic, political, environmental and social costs (Mapstone, 2005).

Crucially, this study shows how the power to detect changes in seabird numbers is largely determined by species-dependent data characteristics. On top of this, any monitoring programme is designed within certain logistical and budgetary constraints, again reflected in an upper power limit. Both researcher and policy-maker should be aware of the fact that not being able to statistically detect a change in seabird numbers should not be mistaken for 'no change'. On the contrary, for some seabird species, this study showed that whatever the change in density, statistical evidence is likely to

run short, illustrating that wind farms should not be regarded innocent until proven guilty.

# References

Cox, S. L., B. E. Scott & C. J. Camphuysen, 2013. Combined spatial and tidal processes identify links between pelagic prey species and seabirds. Marine Ecology Progress Series 479: 203–221.

Embling, C. B., J. Illian, E. Armstrong, J. van der Kooij, J. Sharples, C. J. Camphuysen & B. E. Scott, 2012. Investigating fine-scale spatio-temporal predator–prey patterns in dynamic marine ecosystems: a functional data analysis approach. Journal of Applied Ecology 49: 481–492.

Huettmann, F. & A. W. Diamond, 2006. Large-scale effects on the spatial distribution of seabirds in the Northwest Atlantic. Landscape Ecology 21: 1089–1108.

Jackman, S., 2011. pscl: A Package of Classes and Methods for R Developed in the Political Science Computational Laboratory. Stanford University, Stanford.

Karnovsky, N. J., L. B. Spear, H. R. Carter, D. G. Ainley, K. D. Amey, L. T. Ballance, K. T. Briggs, R. G. Ford, G. L. Hunt Jr, C. Keiper, J. W. Mason, K. H. Morgan, R. L. Pitman & C. T. Tynan, 2006. At-sea distribution, abundance and habitat affinities of Xantus's Murrelets. Marine Ornithology 38: 89–104.

Leopold, M. F., R. S. A. van Bemmelen & A. Zuur, 2013. Responses of local birds to the offshore wind farms PAWP and OWEZ off the Dutch mainland coast. Report C151/12, Imares, Texel.

McCullagh, P. & J. A. Nelder, 1989. Generalized Linear Models, 2nd ed. Chapman and Hall, London.

Mackenzie, M. L., L. A. Scott-Hayward, C. S. Oedekoven, H. Skov, E. Humphreys, & E. Rexstad, 2013. Statistical Modelling of Seabird and Cetacean data: Guidance Document. Report SB9 (CR/2012/05), Centre for Research into Ecological and Environmental Modelling, University of St Andrews, St Andrews.

Maclean, I. M. D., M. M. Rehfisch, H. Skov & C. B. Thaxter, 2013. Evaluating the statistical power of detecting changes in the abundance of seabirds at sea. Ibis 155: 113–126.

Mapstone, B. D., 2005. Scalable decision rules for environmental impact studies: effect size, type I, and type II errors. Ecological Applications 5: 401–410.

Marques, F. F. C. & S. T. Buckland, 2003. Incorporating covariates into standard line transect analyses. Biometrics 53: 924–935.

Onkelinx, T., G. Van Ryckegem, D. Bauwens, P. Quataert & E. Van den Bergh, 2008. Potentie van ruimtelijke modellen als beleidsondersteunend instrument met betrekking tot het voorkomen van watervogels in de Zeeschelde. Report INBO.R.2008.34, Research Institute for Nature and Forest, Brussels.

Pebesma, E. J., A. F. Bio & R. N. M. Duin, 2000. Mapping seabird densities on the North Sea: combining geostatistics and generalized linear models. In Kleingeld, W. J. & D. G. Krige (eds), Proceedings of the Sixth International Geostatistics Congress, Cape Town.

Pérez-Lapeña, B., K. M. Wijnberg, S. J. M. H. Hulscher & A. Stein, 2010. Environmental impact assessment of offshore wind farms: a simulation-based approach. Journal of Applied Ecology 47: 1110–1118.

Pérez-Lapeña, B., K. M. Wijnberg, A. Stein & S. J. M. H. Hulscher, 2011. Spatial factors affecting statistical power in testing marine fauna displacement. Ecological Applications 21: 2756–2769.

Petersen, I. K., T. K. Christensen, J. Kahlert, M. Desholm & A. D. Fox, 2006. Final results of bird studies at the offshore wind farms at Nysted and Horns Rev, Denmark. NERI report, National Environmental Research Institute.

Petersen, I. K., R. D. Nielsen & M. L. MacKenzie, 2014. Post-construction evaluation of birds abundances and distributions in the Horns Rev 2 offshore wind farm area, 2011 & 2012. Technical Report, Aarhus University, Danish Centre for Environment and Energy, Aarhus.

Potts, J. M. & J. Elith, 2006. Comparing species abundance models. Ecological Modelling 199: 153–163.

Ripley, B. & M. Lapsley, 2013. RODBC: ODBC Database Access. R package version 1.3-7. Available on internet at http://CRAN.R-project.org/package=RODBC/.

R Core Team, 2013a. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna. Available on internet at http://www.R-project.org/.

R Core Team, 2013b. Foreign: Read Data Stored by Minitab, S, SAS, SPSS, Stata, Systat, dBase, etc. R package version 0.8-55. Available on internet at http://CRAN.R-project.org/package=foreign/.

Schneider, D. C., 1990. Spatial autocorrelation in marine birds. Polar Research 8: 89–97.

Schneider, D. C. & D. C. Duffy, 1985. Scale-dependent variability in seabird abundance. Marine Ecology Progress Series 25: 211–218.

Scott, B. E., A. Webb, M. R. Palmer, C. B. Embling & J. Sharples, 2013. Fine scale bio-physical oceanographic characteristics predict the foraging occurrence of contrasting seabird species; Gannet (*Morus bassanus*) and

storm petrel (*Hydrobates pelagicus*). Progress in Oceanography 117: 118–129.

Schwemmer, P., S. Adler, N. Guse, N. Markones & S. Garthe, 2009. Influence of water flow velocity, water depth and colony distance on distribution and foraging patterns of terns in the Wadden Sea. Fisheries Oceanography 18: 161–172.

Stewart-Oaten, A. & J. R. Bence, 2001. Temporal and spatial variation in environmental impact assessment. Ecological Monographs 71: 305–339.

Tasker, M. L., P. H. Jones, T. J. Dixon & B. F. Blake, 1984. Counting seabirds at sea from ships: a review of methods employed and a suggestion for a standardised approach. Auk 101: 567–577.

Underwood, A. J. & M. G. Chapman, 2003. Power, precaution, type II error and sampling design in assessment of environmental impacts. Journal of Experimental Biology and Ecology 296: 49–70.

Vanermen, N., E. W. M. Stienen, W. Courtens & M. Van de walle, 2006. Referentiestudie van de avifauna van de Thorntonbank. Report INBO.A.2006.22, Research Institute for Nature and Forest, Brussels.

Vanermen, N., E. W. M. Stienen, W. Courtens, T. Onkelinx, M. Van de walle & H. Verstraete, 2010. Monitoring seabird displacement: a modelling approach. Report INBO.R.2010.12, Research Institute for Nature and Forest, Brussels.

Vanermen, N., E. W. M. Stienen, W. Courtens, T. Onkelinx, M. Van de walle & H. Verstraete, 2013. Bird monitoring at offshore wind farms in the Belgian Part of the North Sea—assessing seabird displacement effects. Report INBO.R.2013.755887, Research Institute for Nature and Forest, Brussels.

Venables, W. N. & B. D. Ripley, 2002. Modern Applied Statistics with S, 4th ed. Springer, New York.

Ver Hoef, J. M. & P. L. Boveng, 2007. Quasi-poisson vs. negative binomial regression: how should we model overdispersed count data? Ecology 88: 2766–2772.

Walls, R., S. Canning, G. Lye, L. Givens, C. Garrett & J. Lancaster, 2013. Analysis of Marine Environmental Monitoring Plan Data from the Robin Rigg Offshore Wind Farm, Scotland (Operational Year 1). Technical Report 1022038, Natural Power Consultants.

Wickham, H., 2007. Reshaping data with the reshape package. Journal of Statistical Software 21. Available on internet at http://www.jstatsoft.org/.

Zeileis, A. & T. Hothorn, 2002. Diagnostic checking in regression relationships. R News 2: 7–10. Available on internet at http://cran.r-project.org/doc/Rnews/.

Zeileis, A., C. Keibler & S. Jackman, 2008. Regression models for count data in R. Journal of Statistical Software 27. Available on internet at http://www.jstatsoft.org/.

Zuur, A. F., E. N. Ieno, N. J. Walker, A. A. Saveliev & G. M. Smith, 2009. Mixed Effects Models and Extensions in Ecology with R. Springer, New York.