Natural England Commissioned Report NECR237

Offshore wind farms and birds: incorporating uncertainty in collision risk models: a test of Masden (2015)

First published 28 September 2017

www.gov.uk/natural-england

NATURAL ENGLAND

Foreword

Natural England commission a range of reports from external contractors to provide evidence and advice to assist us in delivering our duties. The views in this report are those of the authors and do not necessarily represent those of Natural England.

Background

Operational offshore wind farms are known to have a number of potential impacts on birds and these include mortality from collision with turbine blades and ancillary structures (moving and stationary). Offshore windfarm developers routinely use collision risk models (CRMs) to assess this potential impact on birds when undertaking environmental impact assessments. In the UK, for offshore windfarms, the most frequently used avian collision risk model is the Band model (Band 2012).

The Band (2012) model requires a number of input parameters, including information on the density of birds in the windfarm area, bird avoidance rates, flight speed, flight height and size information for the bird species involved and various turbine parameters like rotor diameter, pitch and operational time. All of these input parameters have variability and uncertainty associated with them and since the predicted collision risk from the Band model is sensitive to the input parameters, variability in the input parameters can have a significant effect on predicted collision risk.

However, consideration of this variability in the key input parameters is not routinely included when collision risk modelling is undertaken as part of the Environmental Impact Assessment (EIA) process, and uncertainty/variability around the collision predictions is rarely presented in environmental statements from offshore windfarm (OWF) developers.

For these reasons a project was undertaken to develop the Band (2012) model using a simulation approach to incorporate variability and uncertainty in the collision risk modelling process. The output of this project was the development of a stochastic version of the Band (2012) collision risk model (Masden 2015) which allows variability around input parameters to be entered in the model and used to calculate a distribution of collision risk estimates which reflects the variability in the input parameters.

Natural England, as part of its statutory advice responsibilities in relation to Nationally Significant Infrastructure Projects (NSIPs) in the offshore environment, would like developers to take account of variability and uncertainty in their assessment of potential collision impacts, and the stochastic version of the Band model developed by Masden (2015) offers a means of doing that. However, there has been limited testing of the application of this stochastic version of the Band model to datasets typically used by developers for collision risk modelling. Therefore Natural England commissioned this project to review and test the stochastic version of the model to determine the best way to parameterise the model using data available from EIAs, and to compare outputs derived from the stochastic version of the model against those generated by the Band (2012) model.

Natural England will use the results of this project to inform our advice to offshore windfarm developers and the Planning Inspectorate regarding the assessment and significance of potential collision impacts to birds as part of the Environmental Impact Assessment (EIA) and Habitats Regulations Assessment (HRA) processes.

The results of this Natural England project will also be used in a project commissioned by Marine Scotland that is developing an updated version of the stochastic Band model that builds on the work undertaken to date and will address the gaps and issues identified in the current version by industry and statutory agencies..

This report should be cited as: TRINDER, M., (2017*) Offshore wind farms and birds: incorporating uncertainty in collision risk models: a test of Masden (2015)* Natural England Commissioned Reports, Number 237. York.

Natural England Project Officer – Dr Melanie Kershaw, **Melanie.Kershaw@naturalengland.org.uk**

Contractor – Dr Mark Trinder, MacArthur Green, **mark.trinder@macarthurgreen.com**

Key words – Nationally Significant Infrastructure Project (NSIP), Habitats Regulations, HRA, Environmental Impact Assessment, EIA, marine, seabird, collision risk modelling, CRM, Band model, uncertainty, stochastic modelling, offshore renewables, development planning, sustainable development, marine casework

Further information

This report can be downloaded from the Natural England Access to Evidence Catalogue: **http://publications.naturalengland.org.uk/** . For information on Natural England publications contact the Natural England Enquiry Service on 0300 060 3900 or e-mail **enquiries@naturalengland.org.uk** or **MENE@naturalengland.org.uk**

ISBN 978-1-78354-447-9

Incorporating Uncertainty in Collision Risk Models: a test of Masden (2015)

Prepared by:	Dr Mark Trinder
Authorised by:	Prof Bob Furness
Date:	28/04/2017
Tel:	0141 342 5404
Email:	mark.trinder@macarthurgreen.com
Web:	www.macarthurgreen.com
Address:	95 South Woodside Road | Glasgow | G20 6NT

CONTENTS

1 INTRODUCTION

Natural England would like offshore wind farm developers to be able to present robust collision mortality estimates for birds which reflect parameter uncertainty.

Offshore windfarm developers routinely use collision risk models (CRMs) to assess the potential impacts of wind turbines on birds when undertaking environmental impact assessments. In the UK, the most frequently used avian collision risk model is the Band model (Band 2000, Band et al. 2007), which was subsequently updated to be applicable to the offshore environment for a Strategic Ornithological Support Services (SOSS) project (Band 2012).

The Band (2012) model requires a number of input parameters, including information on the density of birds in the windfarm area, bird avoidance rates, flight speed, flight height and size information for the bird species involved and various turbine parameters like rotor diameter, pitch and operational time. All of these input parameters have variability and uncertainty associated with them and since the predicted collision risk from the Band model is sensitive to the input parameters, variability in the input parameter can have a significant effect on predicted collision risk.

To address this issue, the Band (2012) update of the model includes guidance about how to express uncertainty around the model input parameters when reporting a predicted collision risk. However, this approach is relatively simplistic and is only statistically valid when the sources of variability are independent of one another (Masden 2015). Furthermore, as the approach to considering uncertainty is not an intrinsic part of the modelling process, it is not routinely followed when collision risk modelling is undertaken as part of the EIA process, and uncertainty/variability around the collision predictions is rarely presented in environmental statements from offshore windfarm (OWF) developers.

For these reasons, a stochastic version of the (deterministic) Band (2012) Collision Risk Model (CRM) for birds was developed by Masden (2015). This simulation based model (hereafter referred to as 'the Masden model') was implemented in the R programming environment and used by Masden (2015) to investigate the magnitude of variation in mortality estimates obtained using realistic levels of parameter variance and to perform a sensitivity analysis.

Natural England is interested to understand how the Masden model operates and if it can be parameterised and run using the format of data typically available in reporting for offshore wind farm assessments. The aim of the current project was to review, test and set out options for incorporating variability and uncertainty in CRM input parameters into the Masden (2015) collision risk model update in a statistically and ecologically appropriate way and to compare outputs from the Masden (2015) model with those derived using the Band (2012) model.

This has included consideration of the way in which parameters are inputted to the Masden model and an investigation of methods for quantifying the variability and/or uncertainty around the input parameters. For the purposes of this review only Band model Option 1 results have been compared.

2 ESTIMATING COLLISION MODEL INPUT PARAMETERS FROM SURVEY DATA

The following section provides an overview of data analysis methods which are appropriate for generating robust input parameters for stochastic collision modelling. The methods proposed are based on an understanding of the type of data most likely to be collected (e.g. repeat samples providing a sequence of counts). Alternative methods may also be suitable, however a key factor of relevance to the current project is that under most circumstances the survey data are very unlikely to be well suited to statistical methods based on the normal distribution.

Density of flying birds

Observations of seabirds in flight at a wind farm site are collected using a form of snapshot sampling (the data are conceptually very similar for either boat or digital aerial survey methods). Count data should be analysed using an appropriate method, such as a Generalized Linear Model (GLM) or a General Additive Model (GAM; if spatial covariates are to be included) with a Poisson error structure (ideally, the method should also allow for over-dispersion, with options to use quasi-Poisson errors). Categorical explanatory variables can be used (e.g. month, year, season or survey ID) to obtain density estimates with an appropriate temporal scale (omitting an intercept term makes the outputs simpler to interpret as a single coefficient is produced for each period specified in the model). If spatial covariates such as distance to coast and sea depth are available a GAM type of model can be used. These may also be structured to account for auto-correlation (using modelling approaches such as MRSea developed by the Centre for Research in Ecological and Environmental Modelling[1]).

Most recent wind farm assessments (e.g. Forewind 2013, SmartWind 2015) have undertaken modelling using methods similar to those described above, although in the past a more basic method of density estimation was often applied, with the total number of individuals observed during a survey divided by the total area of snapshot samples. The advantage of the modelling approach is that the results include measures of parameter uncertainty (e.g. SE and confidence intervals), which are lacking from the simple approach. These are informative in their own right, but also enable subsequent assessment to explicitly consider uncertainty.

Offshore wind farm baseline surveys to inform environmental impact assessments for birds are typically conducted each month for a period of two years. Thus, there will be two density estimates available for calculating collision risk in each month. The common currencies when discussing collisions are the estimate for each month and the annual total predicted collisions. To obtain these, the number of collisions can either be calculated for each monthly survey separately and then averaged by month across the two years, or the average monthly density of birds across the two years can be estimated as a first step from which a single monthly collision is estimated. While averaging means is straightforward, it is less simple to combine estimates which include uncertainty. The simplest solution is to avoid the need to do this by fitting a GLM (or similar) to the counts with month as an explanatory variable, but not year (see Annex 1 for an example GLM summary from analysis of snapshot count data). The resulting monthly estimates will accommodate inter-annual variation (albeit derived across only two years), and measures of variance around the estimates can be calculated. The alternative is to use a method for averaging variables which have been estimated with

[1] https://creem2.st-andrews.ac.uk/download/mrsea-guidance/

uncertainty in order to obtain a joint mean and joint uncertainty (e.g. the delta method; see Annex 2 for an example for how to calculate the overall variance for two sample variances). Use of the first approach removes the need to consider such options. For completeness in an assessment, stochastic collision estimates could also be presented using the individual monthly density estimates (e.g. 24 values), but with the former monthly averaged values used for the actual impact assessment.

If collision modelling is being conducted deterministically (e.g. using the Band model) then an indication of the range of collision estimates can be obtained by using the upper and lower confidence interval density estimates as well as the mean density. However, this provides no indication of the probability distribution of collisions which can be derived from a stochastic collision model using randomised parameter values.

If deterministic collisions are being calculated then the method used to estimate the mean and SD density has no impact on how collisions are estimated. However, if a stochastic collision model is being used (such as Masden) it is critically important that the method used in the CRM to generate the simulated (random) density estimates shares the same statistical properties as that used to estimate the densities from the survey data. For example, if an over-dispersed Poisson model has been used for data analysis, random number generation should also use this distribution. Although it is possible to back-calculate a standard deviation (SD) from model coefficients derived with a Poisson error structure (e.g. SD = square-root(n) x (upper C.I. – lower C.I.)/3.92) this makes the assumption that the confidence intervals are symmetrical around the mean. This assumption of symmetrical confidence intervals is often violated for Poisson data, particularly at lower values.

Normal or truncated normal random numbers generated from a back-calculated SD of an asymmetrical confidence interval will therefore be biased to the right (i.e. over-estimated). To illustrate this effect, consider an example with 1,000 random numbers generated using a Poisson random number function with lambda=1 (i.e. mean=SD=1), modelled as an intercept only model (i.e. to obtain the mean) using a GLM with Poisson errors, from which a symmetrical SD is back-calculated and used to generate 1,000 truncated normal random numbers:

```
rnd.pois = rpois(n=1000, lambda = 1)
model1 = glm(rnd.pois~1, family="poisson")
ci = exp(confint(model1))
sd = sqrt(1000) * (((ci)[2]-(ci)[1])/3.92)
rnd.rt.norm = rtnorm(1000, mean=1, sd=sd, lower=0, upper=Inf)
mean(rnd.rt.norm) = 1.309
mean(rnd.pois) = 0.98
```

The mean of the 1,000 truncated normal random numbers was 1.309, 30% higher than the mean of the original data (1) for the underlying process. This effect will be more pronounced if the underlying distribution is over-dispersed. In this case, using this approach (truncated normal numbers estimated for a Poisson variable) to generate densities for CRM would produce a mean estimate 30% higher than it should be.

This is relevant for the current project because the Masden model generates random density values using a truncated normal distribution which uses a mean and SD (like a normal distribution) but also upper and lower limits (the lower limit in this case set to zero). Although the lower limit prevents 'impossible' (i.e. negative) values, there is still an underlying assumption of symmetry. The consequence is that the 'centre' of the distribution is shifted away from the limits (in this case zero).

Using a different probability distribution for random number generation than that which best fits and is used for data analysis is likely to result in a poor match between the resulting random draws and the original data. Further discussion on this is provided in a later section with respect to parameterising the Masden model.

Proportion at collision height

A similar data analysis approach can be used for calculating the proportion of individuals at collision height (PCH), using a GLM with binomial errors (e.g. a binary response of 'at PCH / not at PCH'). Explanatory variables can include month and year, although the temporal resolution that can be used will depend on the sample sizes available. Thus, if sufficient data on flight heights are available in all months (or surveys) then monthly (or survey) PCH can be estimated, but if sample sizes are small, seasonal or annual estimates may be more appropriate.

As for density estimates, randomised values are most appropriately estimated using the same probability distribution (e.g. binomial) to ensure reasonable correspondence between data and simulations. While the Masden model uses a normal distribution to simulate PCH, the risk of generating skewed values is lower because the mean is typically farther away from the constraints of 0 and 1 which apply to proportional data.

If site specific data are not considered suitable for estimating PCH (e.g. insufficient observations) then an alternative is to use the modelled estimates presented in Johnston et al. (2014). This is incorporated in the Masden model (Option 2) and discussed below.

3 MASDEN MODEL REVIEW

Model structure

To convert the deterministic Band model to a stochastic one it is necessary to run the model multiple times with the input parameters for each run drawn at random from appropriate probability distributions. Each iteration of the model generates a different result and summary outputs can be obtained from the multiple iterations that are run (e.g. the mean and confidence intervals). The Masden model generates stochastic mortality estimates by nesting the calculations within a loop. New random numbers are drawn at the beginning of each run through the loop and the outputs of the model are stored at the end of each iteration. The number of simulations (i.e. runs through the loop) is user defined.

Input parameters

Input parameters (e.g. mean and SD) for the Masden model are entered in pro-forma text files ('.csv'). Table 1 lists the input parameters and the file name where they are entered. The three input files listed in Table 1 (CountData.csv, BirdData.csv, TurbineData.csv) can have multiple rows; CountData and BirdData have a row for each species and TurbineData has a row for each specification of turbine.

As can be seen in Table 1, the proportion at collision height (PCH) is modelled as a single value and multiple values (e.g. for different months) cannot be entered (without modifying the script) into the Masden model.

Table 1. Input parameters required for the Masden model. For most parameters the mean is entered in the cell with parameter name and the SD is identified with a suffix ('SD'). Further details on parameter inputs are provided in Masden (2015). Note some parameters are also entered in the model code (e.g. wind speed).

Filename	Parameter	Value	Note
CountData.csv	Monthly density (labelled as Jan-Dec)	Mean & SD	Density of birds in flight in each month
BirdData.csv	AvoidanceBasic	Mean & SD	Option 1 & 2 avoidance rate
	AvoidanceExtended	Mean & SD	Option 3 & 4 avoidance rate
	Body_Length	Mean & SD	From literature
	Wingspan	Mean & SD	From literature
	Flight_Speed	Mean & SD	From literature
	Nocturnal_Activity	Mean & SD	Value in range 0-1
	Flight	Flapping / Gliding	
	Prop_CRH_Obs	Mean & SD	Single value (i.e. not monthly, etc.)
TurbineData.csv	TurbineModel	Name (e.g. output in MW)	
	Blades	Integer	No. of blades
	RotationSpeed	Mean & SD	RPM
	RotorRadius	Mean & SD	
	HubHeightAdd	Mean & SD	Distance between lower rotor tip and highest astronomical tide (HAT). (NB: added to rotor radius this equals hub height).
	BladeWidth	Mean & SD	Max. width (at c. 25% along length from hub)
	Pitch	Mean & SD	Angle of the blade from plane of rotation, degrees
	(Jan-Dec)Op	Mean	% wind availability in each month
	(Jan-Dec)OpMean	Mean & SD	% maintenance downtime in each month

Probability distributions

The Masden model makes use of two probability distributions to generate the random parameter values for each simulation: the normal distribution and the truncated normal distribution. The truncated normal distribution is used when it is necessary to generate random numbers which are constrained by lower and/or upper limits (e.g. a lower limit of 0 prevents negative values being generated). However, the truncated normal distribution is based on the standard normal distribution and therefore it is not appropriate for parameters in the CRM which are poorly represented by the normal distribution (see previous section on density estimation).

The key aspect is that there is no straightforward method for converting a Poisson distribution to the truncated normal (as required for input to the Masden model). This limits the reliability of the outputs obtained from the Masden model, since biased density estimates will result in biased collision estimates. Further consideration of this aspect is provided in a later section.

In addition to these statistical considerations, there are two instances where the Masden model in its original state (i.e. as downloaded from the Marine Scotland website) has errors in how the random number functions are used. The truncated normal distribution function used to generate seabird densities has an upper limit set at 2 (i.e. seabird densities cannot exceed 2 birds per km^2). While this may not be of concern at some sites, there may be instances when this would cause densities to be under-estimated. The second error is the use of the normal distribution for generating random proportions of birds at collision height, rather than the truncated normal with a lower limit of zero. This error means it is possible to obtain negative values, which will in turn result in negative collision estimates (since collisions are calculated as the product of this and other variables). Guidance on how to correct these errors is provided in a later section.

Turbine parameters

Turbine hub height is modelled as a random addition to the rotor radius, measured from Highest Astronomical Tide (HAT). This is simulated as a normal random number. Surveys are likely to have been conducted over a range of tidal states, so the proportion of birds at collision height would be expected to approximate to Mean Sea Level (MSL; this will depend on the extent to which height observations are pooled, although even across a single survey the span of heights may cover several hours). Thus, to accommodate the difference between HAT and MSL the Masden model includes an offset term in the script (i.e. this is not specified in the input tables but is embedded in the model code) which has a pre-set value of 2.5m. The end-user needs to modify this for their wind farm location.

The rationale for modelling hub height and the other turbine dimensions as random variables is that this captures the uncertainty about turbine model selection which may be present at the assessment stage of wind farm development (note this does not simulate tidal variation as this follows a ‘u’ shaped distribution, not a normal distribution). However, while the final turbine design may not be determined when the collision analysis is undertaken, there will be one or more candidate models. Collision modelling, as with all other aspects of the assessment, proceeds on the basis of the ‘worst case scenario’ for any given feature, following the Rochdale Envelope approach. In the case of collision modelling this requires that each candidate turbine is used in the model in order to establish which produces the highest (and hence most precautionary) collision estimates.

It is therefore unnecessary to model these fixed turbine parameters as random variables since for any given turbine they will be known with certainty (or at least have a fixed range of alternative values). Making these random is also inconsistent with the Rochdale Envelope assessment approach. Adapting the Masden model to 'fix' these parameters to be constant is straightforward, by setting the SDs for rotor radius, hub height and blade width to be zero in the TurbineData.csv file.

However, other turbine parameters in the model (RPM and blade pitch) vary in relation to wind speed and it is therefore appropriate to model these as random variables. In its unmodified form the Masden script derives values for RPM and blade pitch from a table which relates these to wind speed (e.g. '*windpower_6.csv*' and '*windpower_8.csv*' are included with the model code for 6MW and 8MW turbines respectively). This table is automatically read into the R workspace during model execution. Values for wind speed (mean and SD) are entered directly into the model script (i.e. these are not included in the tables of input data), from which normal random variables are generated. During each simulation the value for random wind speed is used to obtain the corresponding RPM and blade pitch for use in that simulation. Note that the wind speed is specified as an annual value, not monthly.

Modelling RPM and blade pitch as related functions of wind speed is a sensible approach. However, the values for this relationship have not been derived from any specific turbine model but are instead generic estimates based on expert opinion (during the current project an approach was made to turbine manufacturers to ask if this relationship could be supplied, but these requests were declined on commercial grounds). Thus, it is impossible to be certain if the tables in Masden are suitable for CRM.

In acknowledgement of this, it is stated in Masden (2015) that if mean and standard deviations for RPM and blade pitch are entered in '*TurbineData.csv*' these will be used instead of the *windpower* relationships. However, review of the model code and testing this aspect found that there is no mechanism to enable this switch, and in fact the model always defaults to use the tabulated relationship in the *windpower_6.csv* and *windpower_8.csv* files??, irrespective of RPM and blade pitch values being entered *in TurbineData.csv*.

Flight height distributions

The Masden model generates outputs using Options 1, 2 and 3 of the Band model. For the current comparisons the focus was on Option 1 (site specific flight heights). For Option 1 the Masden model uses the mean and standard deviation of the proportion of birds at collision height (*Prop_CRH_Obs*) in the *BirdData.csv* file to simulate from a normal distribution, which in most cases will provide a reasonable approximation to the underlying proportion data (although see note above about the potential for negative values). For option 2 the overlap between rotor height and bird height (i.e. PCH) is calculated from a pre-defined sample of bird flight heights using data stored in species-specific files (e.g. *Black_legged_kittiwake_ht.csv*). In Masden (2015) it is stated that these were generated by the BTO from the modelling in Johnston et al. (2014). Each species file contains 200 bootstrap samples (200 columns) of the proportion of birds in 1m height intervals between 0 and 300m (300 rows). During each simulation one column is selected at random from the table and the proportion at collision height calculated as the overlap with rotor heights. This approach is considered robust and appropriate and will not result in the generation of negative PCH values.

4 MODEL COMPARISONS

As noted above, the unedited Masden model always uses the *windpower.csv* relationships (wind speed : RPM & blade pitch) even when these parameters are entered in the *TurbineData.csv* file. For the purposes of comparing the Masden model outputs against the Band model (i.e. to run the Masden model as a deterministic model) it was therefore necessary to provide an alternative *windpower.csv* file. This contained constant RPM and blade pitch values (i.e. these had the same value at all wind speeds) to ensure these parameters could not vary.

A second related modification was required to permit comparison of stochastic outputs from the Masden model with Band model outputs derived from upper and lower parameter values (e.g. as presented in SmartWind 2015). This required editing of one of the model scripts ('*sampleturbineparams.txt*'), to allow the alternative sampling method to be used (i.e. use of the mean and SD for rotor speed and blade pitch values in the turbine data sheet to generate normal random variables, rather than the relationship in *windpower.csv*). This was necessary to ensure that RPM and blade pitch varied in a predictable manner around their means, rather than the non-linear relationships specified in *windpower.csv*.

It is worth noting that modelling RPM and blade pitch as independent variables in this manner is expected to inflate the variance of collision model outputs because these variables are actually related to one another (as noted by Masden, and hence the tabulated approach). However, in the absence of manufacturer data this covariance cannot be estimated and it is therefore necessary to model these as independent variables. For interest, outputs using the wind speed version are also presented for comparison, using the *windpower_6.csv* provided with the Masden script.

Deterministic comparison - Masden Model outputs compared to Band Model

The generic bird parameters and turbine parameters in Tables 2 and 5 were made up for the purposes of this comparison. The bird densities (Tables 3 and 6) were estimated from a snapshot boat survey dataset, modelled using a GLM with quasi-Poisson errors (see Appendix 1 for model details). The mean densities for use with the Masden model were the monthly coefficients from the model, while the SDs were calculated from the model confidence intervals (using sqrt(n) x (upper c.i. – lower c.i.)/3.92; where n was the number of snapshots). As discussed above, this makes the assumption that the confidence intervals were symmetrical around the mean, which is unlikely to be the case. However, this method was used here to illustrate the potential influence of this assumption on the outputs obtained.

The input parameter values used are provided in tables 2 and 3. The results obtained from each models are provided in Table 4.

Table 2. Generic bird parameters and wind farm parameters used in the Masden and Band models for deterministic comparison.

Category	Parameter	Masden		Band
		Mean	SD	
Bird (generic)	Body length	0.39	0	0.39
	Wing span	1.08	0	1.08

Category	Parameter	Masden		Band
		Mean	SD	
	Flight speed	13.1	0	13.1
	Nocturnal activity	50	0	3
	Flight type	Flapping	NA	Flapping
	Avoidance rate	98.9	0	98.9
	PCH	0.20	0	0.20
Wind farm	Latitude	55.80	NA	55.80
	Wind farm capacity	600	NA	NA
	Turbine capacity	6	NA	NA
	No. of turbines	Calculated from previous 2 values	NA	100
	Rotor radius	80.00	0	80.00
	No. of blades	3.00	NA	3.00
	RPM	11.00	0	11.00
	Blade pitch	15.00	0	15.00
	Max. blade width	5.50	0	5.50
	Hub height	NA	NA	106.5
	Hub height addition	26.50	0	NA

Table 3. Monthly bird density and wind farm operational parameters for deterministic comparison. Note that the Operation values for the Band model are Operation minus OperationMean for the Masden model (e.g. for January 96.28 - 6.3 = 89.98)

Model	Parameter	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Masden	Density	0.13	0.31	1.03	0.86	0.77	1.274	0.57	0.11	0.18	0.87	0.48	0.09
	Operation	96.28	96.53	95.83	92.78	90.86	92.22	89.11	89.92	93.71	96.14	97.14	96.41
	OperationMean	6.3	6.3	6.3	6.3	6.3	6.3	6.3	6.3	6.3	6.3	6.3	6.3
	operationSD	0	0	0	0	0	0	0	0	0	0	0	0
Band	Density	0.13	0.31	1.03	0.86	0.77	1.274	0.57	0.11	0.18	0.87	0.48	0.09
	Operation	89.98	90.23	89.53	86.48	84.56	85.92	82.81	83.62	87.41	89.84	90.84	90.11
	OperationMean	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
	operationSD	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

Table 4. Deterministic collision modelling results obtained from the Masden model (with all variance =0) and Band model.

Model	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Total
Masden	4.9	10.9	42.9	35.5	33.9	56.6	25.1	4.6	7.2	34.9	17.7	3.3	277.5
Band	4.9	10.9	42.9	35.5	33.9	56.6	25.1	4.6	7.2	34.9	17.7	3.3	277.5

With all parameter variances set to zero and RPM and blade pitch fixed (i.e. not taken from the *windpower.csv* table) the Masden model produces identical results to the Band Model. This is the expected result, since the Masden model was derived from the Band model (however as noted above this could only be confirmed following code modifications to allow all parameters to be fixed).

Stochastic comparison - Masden Model outputs compared to Band Model

Masden Model in original format

The following simulations were conducted without making any adjustments to the Masden script. The input parameters used are provided in Tables 5 and 6. Note that rotor RPM and blade pitch used in the Band model were derived from the calculations using wind speed in the Masden model. In order to obtain the same mean values for use in the Band model it was necessary to run the Masden model first and then extract the mean RPM and blade pitch from the outputs.

A mean wind speed of $16 ms^{-1}$ (and SD of 3.2) was entered in the Masden code as this corresponded to a blade angle (in the original *windpower.csv* table) of 15 degrees and an RPM of 10.2, which were considered to be similar to typical values used in collision modelling. Following completion of the Masden simulations the actual mean RPM and mean blade pitch generated during simulations were 9.87 and 13.3 respectively, and these were used in the Band model.

Table 5. Generic bird parameters and wind farm parameters used in the Masden and Band models for stochastic comparison.

Category	Parameter	Masden		Band
		Mean	SD	
Bird (generic)	Body length	0.39	0.005	0.39
	Wing span	1.08	0.04	1.08
	Flight speed	13.1	1.5	13.1
	Nocturnal activity	50	0.0045	3
	Flight type	Flapping	NA	Flapping
	Avoidance rate	98.9	0.001	98.9
	PCH	0.20	0.033	0.20
Wind farm	Wind speed	16	3.2	NA
	Latitude	55.80	NA	55.80
	Wind farm capacity	600	NA	NA
	Turbine capacity	6	NA	NA
	No. of turbines	Calculated from previous 2 values	NA	100
	Rotor radius	80.00	0	80.00
	No. of blades	3.00	NA	3.00
	RPM	NA	NA	9.87
	Blade pitch	NA	NA	13.3
	Max. blade width	5.50	0	5.50
	Hub height	NA	NA	106.5
	Hub height addition	26.50	2	NA

Table 6. Monthly bird density and wind farm operational parameters for stochastic comparison. Note that the Operation values for the Band model are Operation minus OperationMean for the Masden model (e.g. for January 96.28 - 6.3 = 89.98)

Model	Parameter	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Masden	Density (mean)	0.13	0.31	1.03	0.86	0.77	1.274	0.57	0.11	0.18	0.87	0.48	0.09
	Density (SD)	0.10	0.15	0.28	0.25	0.23	0.31	0.21	0.09	0.12	0.24	0.16	0.08
	Operation	96.28	96.53	95.83	92.78	90.86	92.22	89.11	89.92	93.71	96.14	97.14	96.41
	OperationMean	6.3	6.3	6.3	6.3	6.3	6.3	6.3	6.3	6.3	6.3	6.3	6.3
	operationSD	2	2	2	2	2	2	2	2	2	2	2	2
Band	Density	0.13	0.31	1.03	0.86	0.77	1.274	0.57	0.11	0.18	0.87	0.48	0.09
	Operation	89.98	90.23	89.53	86.48	84.56	85.92	82.81	83.62	87.41	89.84	90.84	90.11
	OperationMean	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
	operationSD	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

The results obtained from the original Masden model and the Band model are provided in Table 7.

Table 7. Stochastic collision modelling results obtained from the unmodified Masden model (with input variances as defined in tables 4 and 5) and Band model.

Model		Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Total
Masden	Mean	5.6	10.6	40.9	34.0	32.9	54.2	24.5	5.2	7.5	34.0	16.9	3.9	270.2
	SD	3.4	5.2	14.6	13.0	11.9	17.0	10.6	3.2	4.5	11.9	6.8	2.5	
	CV	61.9	48.9	35.6	38.1	36.2	31.3	43.0	62.6	60.5	34.9	40.2	64.5	
	Median	5.0	10.2	39.3	32.7	32.0	53.0	23.6	4.9	7.0	32.9	16.4	3.6	260.6
	IQR	4.8	7.0	18.9	16.6	15.9	23.3	14.6	4.4	6.0	15.8	8.9	3.4	
Band		4.6	10.2	40.1	33.2	31.7	52.9	23.5	4.3	6.8	32.7	16.5	3.1	259.6
Band as percentage of Masden	Mean	82.7	96.0	98.1	97.7	96.3	97.7	95.8	82.1	90.6	96.2	98.0	77.6	96.1
	Median	91.9	100.0	102.1	101.4	99.0	99.8	99.8	87.7	96.6	99.4	100.8	85.7	99.6

Using the parameters detailed in Tables 5 and 6 the unmodified Masden model produced slightly higher mean collision estimates (c. 4% higher), although the median outputs were very similar (<0.5% higher).

Masden Model modified to correct misspecifications

For the following comparison the Masden code was edited to remove the upper limit on bird density and to allow rotor RPM and blade pitch to be entered as independent variables. The input parameters were the same as those used for the unmodified Masden model (Tables 5 and 6).

Table 8. Stochastic collision modelling results obtained from the modified Masden model (with variances as defined in tables 4 and 5) and Band model.

Model		Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Total
Masden	Mean	5.4	10.9	40.2	34.4	32.3	54.8	24.5	5.1	7.6	33.7	17.2	3.9	270.0
	SD	3.3	5.3	14.3	12.9	11.9	18.7	10.6	3.2	4.5	12.1	6.9	2.6	
	CV	61.1	48.5	35.5	37.5	36.8	34.1	43.2	62.5	59.9	35.9	40.1	66.4	
	Median	5.0	10.5	38.2	33.2	31.3	52.4	23.2	4.7	7.0	32.9	16.7	3.5	258.6
	IQR	4.4	7.0	18.5	17.3	16.0	24.6	14.1	4.3	6.2	15.3	9.0	3.6	
Band		4.6	10.2	40.1	33.2	31.7	52.9	23.5	4.3	6.8	32.7	16.5	3.1	259.6
Band as percentage of Masden	Mean	85.2	94.0	99.9	96.6	98.1	96.5	96.0	83.7	89.3	96.9	96.2	77.5	96.1
	Median	93.0	97.4	105.0	99.9	101.2	100.9	101.4	91.1	96.6	99.4	99.0	87.2	100.4

A visual comparison of the results in Table 8 is provided in Figure 1. The Masden model produced mean collision estimates that were consistently higher than the Band model, by up to 23%, although the absolute differences were comparatively small with the annual total only 4% higher. The median estimates were closer to the Band outputs. In both cases the magnitude of difference in each month between Band and Masden is negatively related to the CV of seabird density. Thus, the greater the relative uncertainty on density (i.e. larger CV), the greater the difference between the Masden mean (or median) estimate and the Band output. While greater uncertainty should be reflected in less precise estimates, in this case the difference is one of reduced accuracy (not precision), due to the introduction of positive bias in the resampled densities resulting from use of the truncated normal distribution: the mean of the 1,000 resampled densities for each month were larger than the input means in 10 of the 12 months, by up to 2.3%.

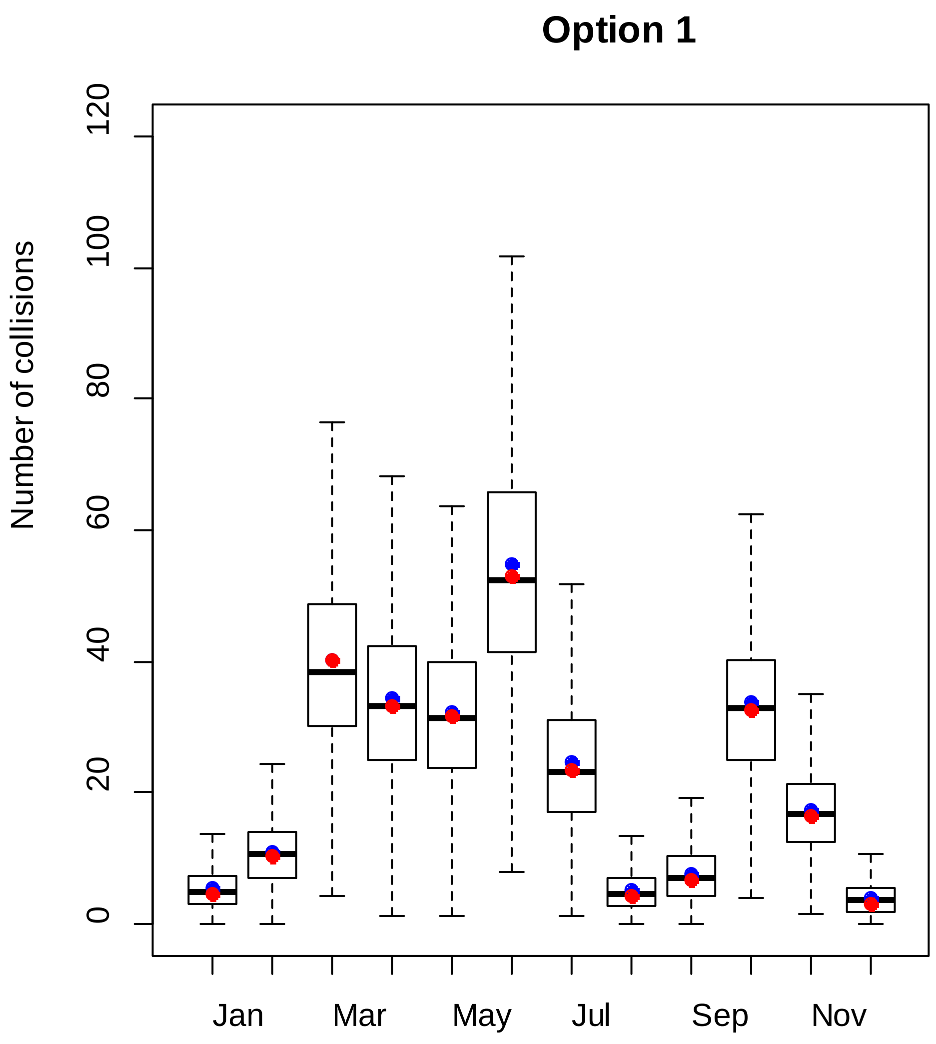

Figure 1. Box and whisker plot of Masden model outputs (in Table 8) using the parameters listed in Tables 5 and 6. The heavy horizontal lines are the medians, the boxes the 25th and 75th percentiles and the whiskers represent the range (the default setting for boxplot as used in the Masden model). The mean Masden values (blue dots) and Band model outputs (red dots) have been overlaid (note the March blue dot is hidden under the red dot).

For the dataset used in this analysis the modified Masden model produced the same results as the unmodified version. However, this would not have been the case if the data contained higher density estimates (i.e. >2/km^2) which would be truncated by the unedited Masden model by the upper limit of 2 defined for that parameter. In addition, the wind speed, RPM and blade pitch values were all standardised across the model runs (to ensure comparisons were based on the same data). However, ensuring the unedited Masden model and the Band model had the same values for RPM and blade pitch can only be achieved through a process of trial and error or by modifying the wind speed table (e.g. setting all RPM and blade pitches to the same value, although this removes the stochastic aspect for these parameters).

An alternative option to present uncertainty in collision predictions without using a stochastic model such as Masden is to calculate Band outputs using the upper and lower values for selected input parameters (e.g. SmartWind 2015). This can't provide a probability distribution of outputs, but does indicate the range over which estimates could lie. The Band model results obtained using upper and lower confidence values for seabird density (i.e. 95% confidence interval values obtained from the GLM of survey data derived using the 'confint' function) on their own and also with the avoidance rate set to upper and lower levels (i.e. +/- 0.002) are provided in Table 9 and Figure 2.

Table 9. Collision modelling results obtained from the modified Masden model (with variances as defined in Tables 5 and 6) and Band model using upper and lower 95% confidence seabird density estimates obtained from a GLM and also with recommended upper and lower avoidance rates (98.7 - 99.1%).

Model		**Jan**	**Feb**	**Mar**	**Apr**	**May**	**Jun**	**Jul**	**Aug**	**Sep**	**Oct**	**Nov**	**Dec**	**Total**
Masden	Mean	5.4	10.9	40.2	34.4	32.3	54.8	24.5	5.1	7.6	33.7	17.2	3.9	270.0
	SD	3.3	5.3	14.3	12.9	11.9	18.7	10.6	3.2	4.5	12.1	6.9	2.6	
	CV (%)	61.1	48.5	35.5	37.5	36.8	34.1	43.2	62.5	59.9	35.9	40.1	66.4	
	Median	5.0	10.5	38.2	33.2	31.3	52.4	23.2	4.7	7.0	32.9	16.7	3.5	258.6
	IQR	4.4	7.0	18.5	17.3	16.0	24.6	14.1	4.3	6.2	15.3	9.0	3.6	
Band	Mean	4.6	10.2	40.1	33.2	31.7	52.9	23.5	4.3	6.8	32.7	16.5	3.1	259.6
Density range	Lwr 95%	0.7	3.6	17.8	20.5	16.8	32.9	11.0	0.8	0.9	18.1	8.4	0.3	131.71
	Uppr 95%	15.3	22.4	75.2	50.7	52.9	79.8	43.3	12.3	23.0	54.0	30.7	11.7	471.37
Density range & Avoidance rate range	Lwr 95% & 99.1% AR	0.6	2.9	14.5	16.8	13.7	26.9	9.0	0.7	0.7	14.8	6.9	0.2	107.76
	Uppr 95% & 98.7% AR	18.0	26.5	88.9	60.0	62.5	94.4	51.2	14.6	27.2	63.8	36.3	13.8	557.08

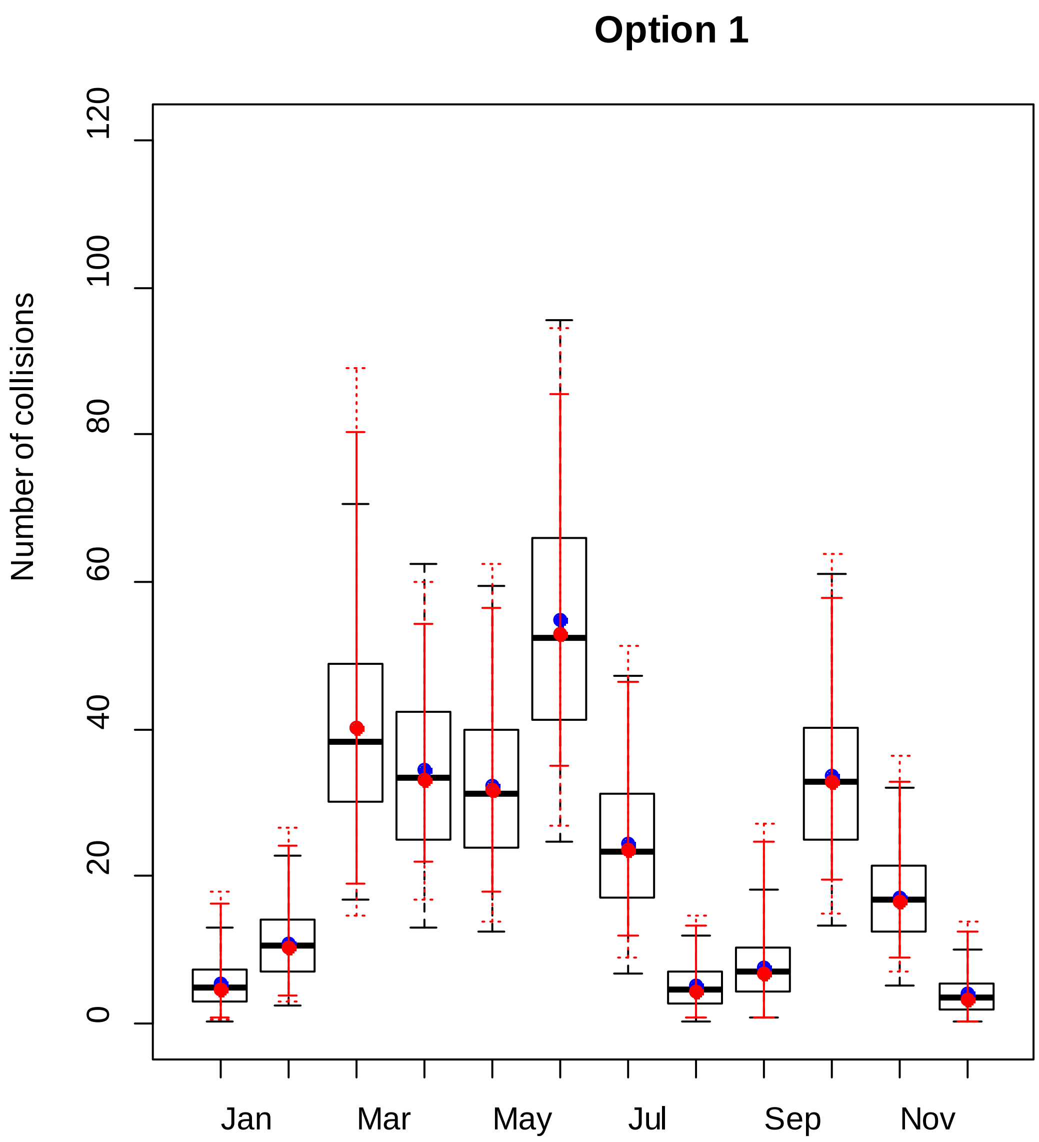

Figure 2. Box and whisker plot of Masden model outputs (in Table 9) using the parameters listed in Tables 5 and 6. The black horizontal lines are the median, the boxes the 25th and 75th percentiles and the black whiskers the 95% confidence interval (i.e. 2.5% - 97.5%). The mean Masden values (blue dots) and Band model outputs (red dots) have been overlaid. The solid red lines indicate the upper and lower Band outputs using 95% confidence intervals (i.e. 2.5% - 97.5%) from the seabird density GLM, the dotted red lines also include +/- 0.002 applied to the avoidance rate (i.e. 98.7 - 99.1%). It should be noted that for this figure the boxplot function has been modified from that defined in the Masden model to generate whiskers (black dotted lines) at the 95% confidence range for comparison with the intervals presented around the Band model outputs (red whiskers).

Comparing the Band model upper and lower estimates with those from the Masden model, it can be seen that the 95% confidence ranges generated by the Masden are generally fairly similar, although there is no consistent pattern (i.e. in some months the Band model outputs are wider and in others the Masden model outputs are wider). It should be noted that for comparison the boxplot function used in the Masden model has been modified slightly for Figure 2 to obtain the equivalent 95% confidence range as that presented for the Band model outputs.

It is interesting to note that the extent of the Band model ranges was more influenced by the uncertainty in the density estimates than the avoidance rate, the latter contributing a maximum of 30% to the range of collision estimates (peaking for the higher absolute collision estimates).

5 RECOMMENDATIONS FOR EDITING THE MASDEN MODEL

The Masden model script without modification (i.e. as downloaded) produces mean collision estimates which may be different (depending on input parameter values) from those obtained by the Band model for the following reasons:

- The simulated proportion at collision height can generate negative values (depending on the mean and SD entered),
- The simulated seabird densities are capped at 2/km^2, and
- Rotor RPM and blade pitch are simulated as a function of a randomly generated wind speed variable, using a tabulated relationship which is not based on actual turbine parameters (it should be noted that the reports which accompany the Masden model state that this relationship is over-ridden if a mean and SD for rotor speed and blade pitch are entered, however this is incorrect as the model code does not include a mechanism to perform this switch).
- The mean of the density values generated from the normal (or truncated normal) distribution may differ from the input mean values, due to inherent differences between the underlying distribution and the normal or truncated normal distributions.

As a consequence, the model should not be used for wind farm assessment without modification. The following steps can be taken to correct the above aspects. These modifications were applied to obtain the outputs in Table 9.

- The R script '*sampleCRH.R*' should be changed from:

```
sampleCRH <- function(meanCRH, sdCRH) {
  rnorm(1, meanCRH, sdCRH)
 }
```

To :

```
sampleCRH <- function(meanCRH, sdCRH) {
  rtnorm(1, meanCRH, sdCRH,lower=0,upper=1)
 }
```

This constrains the resampled collision height estimates to lie between 0 and 1. Note this is only necessary when using site specific flight height data (e.g. Option 1).

- The R script '*samplecount.R*' should be changed from:

```
sampleCount <- function(meancount, sdcount){
  rtnorm(1, meancount, sdcount,0,2)
 }
```

To:

```
sampleCount <- function(meancount, sdcount){
  rtnorm(1, meancount, sdcount,lower=0,upper=Inf)
 }
```

This removes the upper seabird density cap of 2.

- The text file '*sampleturbineparams.txt*' should be modified as follows.

Lines 3 to 10 (inclusive) shown below, should be commented out – add '#' at the beginning of each line. This prevents these lines from being used by R.

```
 ####ROTOR SPEED (related to wind speed)####
source("scripts\\get_rotor_plus_pitch_auto.txt")
randomSample<-sample(length(rotorSpeed),1)
sampledRotorSpeed[i]<-rotorSpeed[randomSample]

 ###PITCH (related to wind speed and linked to above)####
sampledRotorPitch[i]<-rotorPitch[randomSample]
Pitch = sampledRotorPitch[i]*pi / 180 #### Transform Pitch, needed for Collision Risk Sheet
```

Becomes:

```
####ROTOR SPEED (related to wind speed)####
#source("scripts\\get_rotor_plus_pitch_auto.txt")
#randomSample<-sample(length(rotorSpeed),1)
#sampledRotorSpeed[i]<-rotorSpeed[randomSample]

###PITCH (related to wind speed and linked to above)####
#sampledRotorPitch[i]<-rotorPitch[randomSample]
#Pitch = sampledRotorPitch[i]*pi / 180 #### Transform Pitch, needed for Collision Risk Sheet
```

- The following lines should then be pasted in below the commented lines:

```
## Modified script to generate resampled rotor speed and blade pitch from input data in TurbineData.csv
ifelse(!is.na(TurbineData$RotationSpeedSD[t]), rotorSpeed<-sampleRotorRadius(TurbineData$RotationSpeed[t], TurbineData$RotationSpeedSD[t]), rotorSpeed<-TurbineData$RotationSpeed[t])
  sampledRotorSpeed[i]<-rotorSpeed

ifelse(!is.na(TurbineData$PitchSD[t]), rotorPitch<-sampleRotorRadius(TurbineData$Pitch[t], TurbineData$PitchSD[t]), rotorPitch<-TurbineData$Pitch[t])
  sampledRotorPitch[i]<-rotorPitch
  Pitch=sampledRotorPitch[i]*pi / 180 #### Transform Pitch, needed for Collision Risk Sheet
```

This ensures that the Masden model will sample the RPM and blade pitch from the mean and SD values entered in the *TurbineData.csv* file rather than the windpower.csv file.

There is an option in the Masden script which allows the initial state for the random number generator to be set to a fixed value (this is set to 100 in the code: *'set.seed(100')*). The advantage of this is that results are repeatable (i.e. the same sequence of 'random' numbers is generated on each run of the model). However, failing to switch this off (or alternatively, setting the seed to a new value each time (e.g. using the CPU clock: '*set.seed(as.numeric(Sys.time()))*') will lead to unexpected outputs (e.g. identical results on every simulation).

The above aspects of the code are relatively straightforward to correct through editing of the Masden model code, however this requires an understanding of the R programming language.

More fundamentally, in its current state without modification (i.e. as available on the Marine Scotland Datasets webpage[2]) the Masden model uses inappropriate probability distributions for some parameters. As a consequence, there is a high likelihood that use of the Masden model will result in erroneous collision estimates (i.e. estimates which do not accurately reflect input parameters due to errors in the model code and the way data are simulated).

6 OTHER CONSIDERATIONS FOR STOCHASTIC COLLISION MODELLING

The Masden model in its unedited state samples rotor RPM and blade pitch jointly using wind speed. This approach correctly identifies that these turbine parameters are not independent of one another, but are closely related and jointly dependent on wind speed. However, while this is an appropriate method to model these variables, the relationship between wind speed and the turbine rotor operation has not been made available by the turbine manufacturers, therefore the accuracy of the relationship is unknown. Thus, to permit comparison of outputs with the Band model it was necessary to derive the mean values for RPM and blade pitch from the ones generated by running the Masden model (using the RPM, blade pitch, and windspeed relationships table provided with the Masden model). The alternative is to set the mean and SD using turbine data and modify the code (as described

[2] http://marinedata.scotland.gov.uk/dataset/developing-avian-collision-risk-model-incorporate-variability-and-uncertainty-r-code

above) to make these variables independent of one another. This allows closer comparability with the Band model, but will inflate the overall variance of the outputs. Furthermore, this highlights the fact that there are several other components of the collision model which are related and which should therefore covary in a stochastic model.

A key example of this is the avoidance rate. Seabird avoidance rates have been estimated from long term datasets (Cook et al. 2014). The estimates are therefore mean values for the study periods used, and equivalent mean parameter estimates should be used for the other model input parameters (e.g. flight speed, proportion at collision height, etc. should be derived over similar time frames). It therefore follows that simulating each parameter around its mean value should ensure that the mean collision estimate obtained will correspond to the individual input parameter means. However, unless the parameters have been combined within each model iteration in such a way as to avoid inappropriate combinations the variance around the mean collision estimate will be inflated. Incorporating covariance in the model is an important consideration for development of a reliable stochastic model.

This is important, since the main objective of a stochastic collision model is to improve understanding of the variance around the mean estimates. As demonstrated above, the Masden model produces mean and median values which are very similar to those from the Band model. But because the parameters are simulated independently the overall 'parameter space' generated will be inflated to an unknown extent with a result that the collision estimates will also have wider confidence intervals than if the input parameters were simulated with realistic levels of covariance.

The proportion of birds at collision height can only be entered in the Masden model as a single value (mean and SD) which is applied as an annual average (although the model could be run for a single month or months to apply seasonal variation in this and other parameters). It would be appropriate to model collisions using a monthly value for this parameter if it can be estimated for a given location. This would require considerably more editing of the current scripts and is beyond the scope of the current project.

The simplest robust option for producing randomised density estimates for input to a stochastic collision model is to bootstrap the snapshot counts for a given month. The drawback of this approach is that for low density species there may be a limited number of non-zero counts from which to draw (i.e. there may be a very small range of possible outputs). A more flexible approach is to use a function such as *generateNoise* (MRSea Power) which uses the outputs from a model of the snapshot counts (e.g. GLM or GAM), including any over-dispersion parameter. Unlike bootstrapping, this method is not constrained by the original observations. For example, if the original sample only included snapshot counts of 1, 2 and 5 individuals, the bootstrap resamples will have the same three count sizes. In contrast, resamples obtained using *generateNoise* can take any integer value within the range defined by the model. In both cases the output is a vector of counts the same length as the original number of surveyed samples. The column sum divided by the total area of snapshots is a random density estimate for input to the CRM. Repeating this process generates bird density estimates that can be used to produce collision estimates incorporating uncertainty in species density at the project site in a statistically robust way.

As discussed above, seabird counts used to derive densities are poorly represented by the normal or truncated normal distribution. Thus, a stochastic CRM either needs to permit random number generation using different distributions (e.g. Poisson) or alternative parameter inputs (e.g. external generation of multiple densities using bootstrap methods which can be used in simulations as outlined in the paragraph above).

As described above, one option is to use the results of a Poisson GLM to generate random resamples which correspond to the observed distribution. However, there is no means in the unedited Masden script to specify alternative random number generation or alternative density inputs (the user must supply mean and SD values for use with a truncated normal random number generator).

The best option currently available is to calculate the mean and SD from the resampled GLM data (as above) and use these as Masden model inputs. The drawback to this is that a Poisson (or over-dispersed Poisson) process is likely to be poorly represented by the truncated normal distribution that the Masden model uses to sample densities from. The magnitude of difference between the underlying (over-dispersed Poisson) process and that obtained using the truncated normal as described above, depends on how close the mean density is to zero. At low mean densities (e.g. <0.5 birds / km^2) the truncated normal estimates are biased high (Figure 3), although this bias decreases as the mean increases and is effectively undetectable at higher (e.g. >1 bird / km^2 (Figure 4).

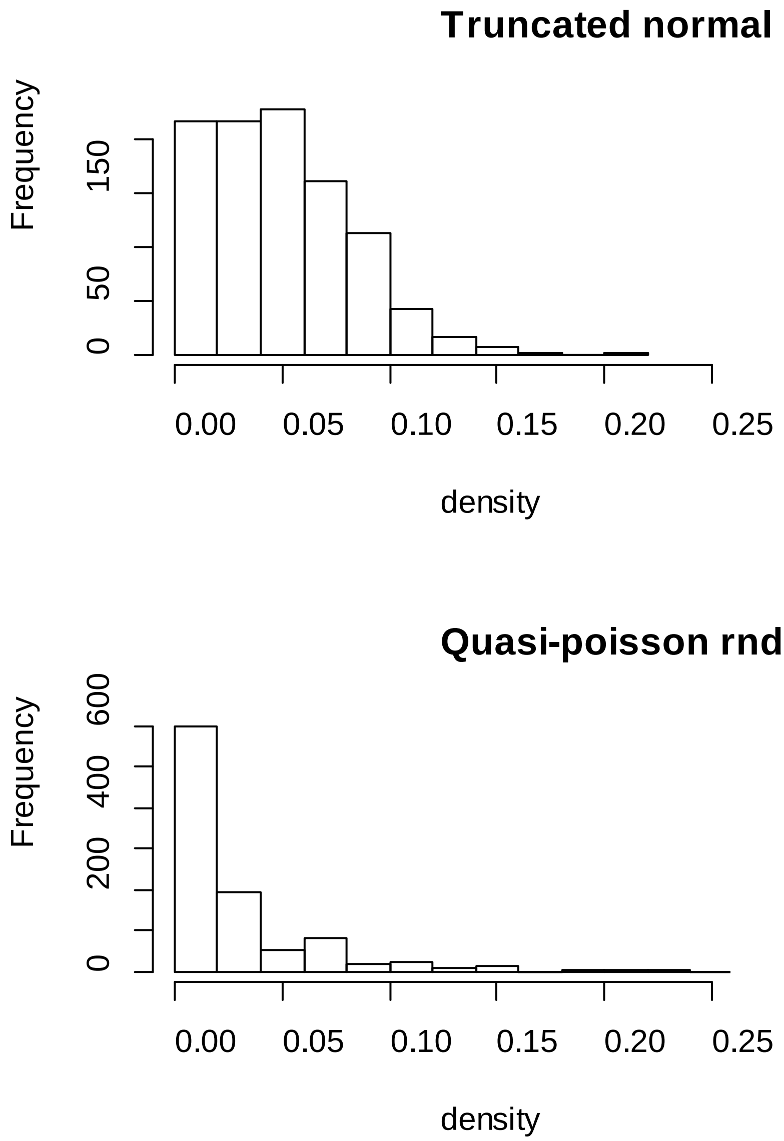

Figure 3. Low density resampled seabird densities, the values in the lower plot have been generated directly from an over-dispersed Poisson GLM using the generateNoise function. The values in the upper plot have been obtained using the mean and standard deviation (of the samples in the lower plot) as inputs to the rtnorm function (truncated normal random numbers). The truncated normal random deviates are shifted to the left compared with the underlying distribution. The original distribution (lower plot) has a mean (sd) of 0.031 (0.044) while the mean (sd) of the truncated normal distribution in the upper plot is 0.049 (0.032).

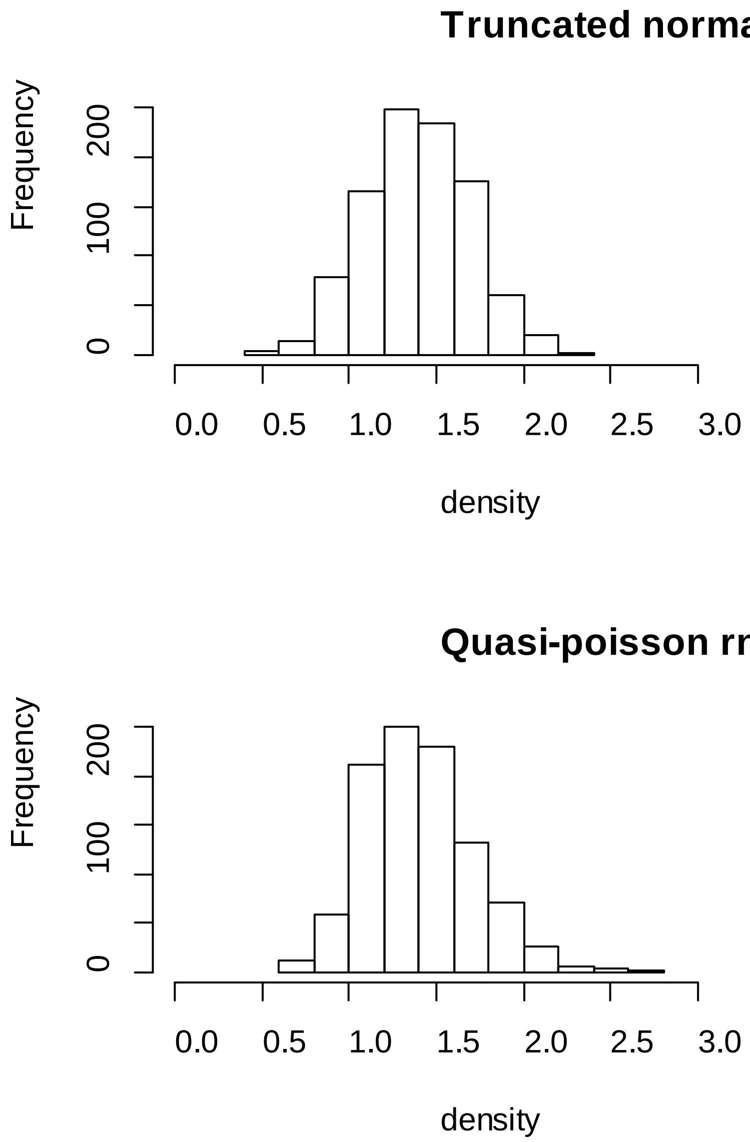

Figure 4. Medium density resampled seabird densities, the values in the lower plot have been generated directly from an over-dispersed Poisson GLM using the generateNoise function. The values in the upper plot have been obtained using the mean and standard deviation (of the samples in the lower plot) as inputs to the tnorm function (truncated normal random numbers). These illustrate that as density increases the bias declines to undetectable levels. The original distribution (lower plot) has a mean (sd) of 1.39 (0.304) while the mean (sd) of the truncated normal distribution in the upper plot is 1.39 (0.0297).

As noted above, the truncated normal distribution is used to obtain normal random numbers which are constrained by lower and/or upper limits. While this can prevent inappropriate values, the results are not necessarily a good match for the underlying process. In addition, the bird densities generated in the unedited Masden model using the truncated normal distribution have an upper limit which has been fixed at 2 birds / km^2. It is assumed that this is an error in the relevant script which if uncorrected risks generating incorrect densities for abundant species.

A GLM approach is also a robust method for estimating the proportion of birds at collision height with variability. The first step is to convert observed flight heights to a binary state (0 = not at PCH, 1 = at PCH). These data can then be modelled using a GLM with binomial errors. As for density estimates, these can be resampled directly to be used as CRM inputs. In order to use PCH data modelled in this manner with the unedited Masden model the mean and SD can be calculated across the resampled values. However, there is a potentially important error in the Masden script when using option 1 and site specific flight height data: the proportion of birds at collision height is simulated using a normal distribution (i.e. these are not truncated at zero) and it is therefore possible to obtain negative values for this parameter if the mean PCH is low, or the SD is large (or both). Using a negative value for PCH will result in a negative collision estimate, and reducing the summary values obtained. Unless there are a lot of negative values (i.e. resulting in a negative lower confidence interval) this is unlikely to be obvious in the summary outputs. This should be corrected (see section 5) prior to use of the Masden model.

On a practical level, the Masden model generates stochastic mortality estimates by nesting the calculations within a loop. New random numbers are drawn at the beginning of each run through the loop and the outputs of the model are stored at the end of each iteration. While this approach is conceptually straightforward, it is inefficient (i.e. the model runs slowly). Simulations can be undertaken much more efficiently through the use of vectorisation. This minimises the use of loops by generating multiple random values for each parameter in a single step and then multiplying these together to obtain tables of outputs which are the same as those obtained at the end of a looped process.

It is important to state that regardless of the method used (looped or vectorised), the results obtained are the same. Therefore, although the Masden code is slow compared with vectorised script, this does not preclude its use (although the time saving may be significant: as an example, to complete 1,000 simulations for a single species the run time for the unedited Masden code was 45 minutes, while a vectorised version achieved the same outputs in less than 4 seconds).

REFERENCES

Band, W. (2000). Guidance WINDFARMS AND BIRDS: calculating a theoretical collision risk assuming no avoidance action. Guidance notes series 2000. Scottish Natural Heritage, Battleby.

Band, W., Madders, M., & Whitfield, D.P. (2007). Developing field and analytical methods to assess avian collision risk at wind farms. In: de Lucas, M., Janss, G.F.E. & Ferrer, M. (eds.) *Birds and Wind farms: Risk Assessment and Mitigation*, pp. 259-275. Quercus, Madrid

Band, W. (2012). Using a collision risk model to assess bird collision risks for offshore wind farms. Report to The Crown Estate Strategic Ornithological Support Services (SOSS), SOSS-02. http://www.bto.org/science/wetland-and-marine/soss/projects

Cook, A.S.C.P., Humphries, E.M., Masden, E.A., and Burton, N.H.K. (2014). The avoidance rates of collision between birds and offshore turbines. BTO research Report No 656 to Marine Scotland Science

Forewind (2013) Dogger Bank Creyke Beck Environmental Statement Chapter 11 Appendix A - Ornithology Technical Report

Masden, E. (2015). Scottish Marine and Freshwater Science Vol 6 No 14: Developing an avian collision risk model to incorporate variability and uncertainty. Published by Marine Scotland Science. DOI: 10.7489/1659-1. http://www.scotland.gov.uk/Resource/0048/00486433.pdf

Johnston, A., Cook, A.S.C.P., Wright, L.J, Humphreys, E.M. & Burton, N.H.K. (2014) Modelling flight heights of marine birds to more accurately assess collision risk with offshore wind turbines – Corrigendum. Journal of Applied Ecology 51: 1126-1130.

SmartWind (2015) Collision Risk Modelling; Addressing uncertainty clarification note. Appendix J to the Response submitted for Deadline 1. Application reference: EN010053.

ANNEX 1.

Bird density modelling. Note that no surveys were conducted in November in the example dataset. For the CRM tests density parameters for November (mean, c.i.) were averaged across October and December. The original data comprises 22 surveys across a two year period, with regular snapshot counts (range: 362 – 461) collected by boat survey.

> summary(mod1)
Call:
glm(formula = Numbers ~ as.factor(Month) - 1 + offset(log(area)), family = quasipoisson)

Deviance Residuals:

Min	1Q	Median	3Q	Max
-0.4781	-0.3944	-0.3220	-0.1577	21.1670

Coefficient	Estimate*	Std. Error	t value	Pr(>|t|)	Lower c.i. #	Upper c.i. #
Month1	0.1372	0.7360	-2.690	0.00716	0.02088	0.4424
Month2	0.3129	0.4604	-2.538	0.01118	0.10758	0.6814
Month3	1.0340	0.3630	0.075	0.93983	0.45868	1.9414
Month4	0.8727	0.2302	-0.635	0.52571	0.53050	1.3145
Month5	0.7795	0.2899	-0.926	0.35425	0.40813	1.2846
Month6	1.2580	0.2253	1.060	0.28911	0.78830	1.9160
Month7	0.5705	0.3448	-1.600	0.10959	0.26898	1.0572
Month8	0.1071	0.6657	-3.361	0.00078	0.02026	0.3119
Month9	0.1847	0.7806	-2.195	0.02822	0.02342	0.6111
Month10	0.8798	0.2760	-0.476	0.63391	0.4838	1.4401
Month12	0.0851	0.9014	-2.700	0.00695	0.0074	0.3446

* Note these estimates have been converted using ‘exp()’ to obtain values on the response scale.
# The confidence intervals were obtained using function ‘confint()’

(Dispersion parameter for quasipoisson family taken to be 4.874795)
Null deviance: 3789.4 on 8812 degrees of freedom
Residual deviance: 3360.7 on 8801 degrees of freedom
AIC: NA
Number of Fisher Scoring iterations: 7

ANNEX 2.

The following sets out a method for calculating an overall (or average) variance for two variables which have their own mean and variances (i.e. the average variance for two monthly densities which each have their own average and variance).

For a two-sample calculation, the input parameters are:

- n1 and n2 (sample sizes, e.g. n1= 300, n2 = 400)
- x1 and x2 (sample means, e.g. x1= 25, c2 = 15)
- x.bar (mean of x1 and x2, e.g. x.bar = 20)
- v1 and v2 (variance estimates, e.g. v1 = 25, v2 = 9)

Calculate the overall error sum of squares:

ESS.total = (v1 * n1-1) + (v2 * n2-1)

Calculate the overall group sum of squares:

GSS.total = ((x1 – x.bar)^2 *n1-1) + ((x2 – x.bar)^2 *n2-1)

Calculate the overall variance:

Overall variance = (ESS.total + GSS.total) / ((n1+n2)-1)

Using the example values the following distributions are obtained: