User documentation for

SCRAM

Developed by:

Biodiversity Research Institute

The University of Rhode Island

U. S. Fish and Wildlife Service

With funding by:

The Bureau of Ocean Energy Management

17 March 2022

What is this?
We are providing online access to a model that simulates collision risk to birds from existing or planned wind development. We adapted the underlying model from the widely used framework developed by Band (2012), which is often referred to as a collision risk model (CRM). This CRM is run in the open source computing software R (R Core Team 2021) using code that we adapted from Masden (2015) and Trinder (2017). We also developed a user interface for running the CRM in a web browser, similar to the application developed by Marine Scotland (McGregor et al. 2018), but specific to the western Atlantic. Collectively, we are calling this adaptation of the CRM and user interface, SCRAM.

What is the goal?
Our goal is to provide user-friendly access to a web application that generates evidence-based projections of collision risk using a transparent and defensible model. While SCRAM can be run for any species for which a user has data, it also includes default options that contain all of the required species data for three birds that are often the focus of species impact assessments in the western Atlantic: Roseate Tern (*Sterna dougallii*), Red Knot (*Calidris canatus*), and Piping Plover (*Charadrius melodus*). SCRAM can provide decision support for both environmental assessments and research related to collision risk from offshore wind. For more specialized applications, the underlying code for our adaptation of the Band, Masden, and Trinder CRM is available on GitHub.

What is a Collision Risk Model (CRM)?
At its core, a CRM is a simulation of a) how many individuals of a given species will pass through the airspace of an existing or planned turbine array and b) how many of these individuals will collide with the turbines' blades. The rules of the simulation are determined by first principles – e.g. envision throwing a bird-sized beanbag through a spinning fan – as well as basic ecological models that estimate the likelihood of birds being in the vicinity of turbines (see Appendix I). The simulation relies on array-specific data for physical turbine characteristics (e.g. number of blades, rotor speed and height, and the number of turbines) and the characteristics of target species (e.g. typical flight speed, passage rates through the area of interest, bird size, and avoidance behavior). This type of collision risk model includes all of the major components that are likely to influence risk for a proposed or existing array, but it does not integrate information from other arrays in the region. It is, therefore, best suited for detailed, array-specific assessments of risk, such as endangered species assessments and environmental impact statements. One major difference between our adaptation of this CRM framework and previous efforts is that SCRAM parameterizes bird passage rates using publicly accessible data from the *Motus* wildlife tracking system, a network of automated telemetry locations. See Appendix I for more information on the differences between SCRAM and previous implementations of Band's (2012) collision risk framework. For additional technical details of the modeling framework used in SCRAM see Band (2012), Masden (2015), and Trinder (2017), as well as Appendices I, III, IV, and V.

How do I use the web application?

SCRAM requires two types of data to run: 1) "Turbine data", which is provided via a single spreadsheet of turbine and array characteristics, and 2) "Species data", which is baked into the model for the three target species. Users may, however, supersede these defaults for target species or specify data for additional species by providing species data, via three simple spreadsheets. Examples of each data input can be downloaded from the application interface (and are shown in Appendix II). Once the data for turbine and array characteristics are compiled and formatted appropriately (see examples on interface and Appendix II), we anticipate that it should take 3-5 minutes to finish setting up SCRAM to run for a target species. The run time once the data are uploaded will depend on which version of the model specification is selected. The general steps for running SCRAM are discussed below. The interface does not require that these steps are done sequentially, but we recommend following this order to ensure that steps are not missed:

1) Select the species of interest or select the option for providing your own species data.
2) Upload species data (optional if selecting from among the three target species).
3) Check the species data for correct values.
4) Upload turbine and array data.
5) Check the wind farm data for correct values.
6) Choose which version of the CRM to run.
7) Select the number of iterations (100-10,0000).
8) Set a threshold for the maximum acceptable number of collisions.
9) Run CRM.
10) Run sensitivity analyses (optional).
11) Generate summary report and/or download results for each iteration.
12) Check the CRM results.

Each step is discussed in more detail below.

1) *Select the species of interest or select the option for providing your own species data.* Select at least one target species or "Use your own data." If the latter option is selected, you must provide all of the required species data using the "Upload species data" box (see step 5).
2) *Upload species data (optional if selecting from among the four target species).* If "Use your own data" is selected, the application requires three types of species data as .csv files: species characteristics (see Table 2 in Appendix II for an example with appropriate units), flight height distributions (Table 3 of Appendix II), and estimates of passage rates for the array, which can be estimated using a variety of methods including boat surveys or tracking studies (Table 4 of Appendix II). The files for species characteristics and passage rates contain a row for each species of interest, while separate files for flight height distributions are required for each species. If one or more of these species data files are provided for target species, the user-provided files will supersede the application's default data, making it possible to use a combination of user-specified and default data. See Appendix II for more information on the required format for species data. Users can download an example of each file by clicking "Download example input files" on the interface. The naming

conventions of the files do not matter, as long as the appropriate fields are included and correctly named.

3) *Check the species data tables for correct values.* Make sure the values you uploaded to SCRAM are appropriate for these models. Examine the flight height graph to see if the flight height data make sense and are appropriate for the model.
4) *Upload turbine and array data.* Data for turbine and array characteristics are required to run any version of SCRAM. Turbine and array characteristics include the physical and geographic characteristics of the wind farm, including the dimensions of the turbine model, power targets, and the width and geographic coordinates of the turbine array. The application accepts these data as a single .csv file that has alternate options for arrays specified as rows. SCRAMR runs for a single geographic location at a time, but the specified array options can vary by other parameters including power, size, and turbine model specifications. If the rows for the turbine and array data include more than one set of geographic coordinates, only the coordinates from the first row will be used. See Table 1 in Appendix II for an example with appropriate units or click "Download example input file" on the application interface. The naming convention of the file itself does not matter, as long as the appropriate fields are included and correctly named. Turbine specifications, such as blade pitch and width, can be found on the manufactures web pages (e.g. models from General Electric) or the U.S. Geological Survey's Wind Turbine Database.
5) *Check the wind farm data tables for correct values*. Make sure the values you uploaded to SCRAM are appropriate for these models.
6) *Choose which version of the CRM to run*. The principles that the CRM uses to simulate collision risk are simple, but there are two flavors for how these principles are executed that differ by how the underlying calculations use the input data. Band, Masden, and Trinder provided several options, which we synthesized to provide two options – one that we have shown performs best and one that gives approximate estimates in much less time. In SCRAM, both options use the same input data and any necessary data manipulations are done automatically. See Figure 1 for a diagram of how the options available in different iterations of this CRM framework are related to each other.

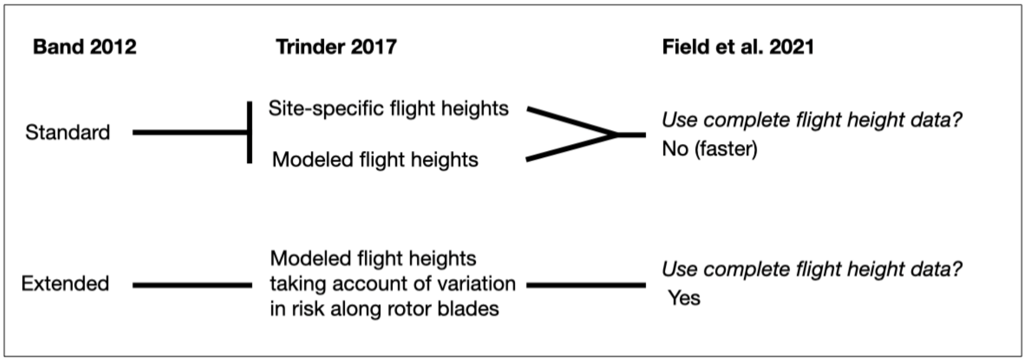


Figure 1. Comparison of methods for simulating collision risk across implementations of the Band (2012) CRM framework.

Because these options estimate collision risk using the same data input, we recommend selecting “Yes” for “Use complete flight height data” whenever the user is not severely limited by computation time. We do not recommend running both options for comparison because, in most cases, the option that uses complete flight height data will provide the more accurate estimate (see Appendix IV). Selecting “No” reverts to the “Standard” option in Band (2012), which either use a single estimate for the proportion of birds flying at risk height or obtains this estimate by summarizing the full flight height distribution, which is the method SCRAM uses (Figure 1). For SCRAM’s target species, we recommend selecting “Yes” to use the full flight height distribution to account for the variation in risk along the rotor blades. We also recommend this option for other species whenever time allows. Appendix V shows the underlying mathematical routines that differ between options and a test of their relative speed.

7) *Select the number of iterations*. Use the slider to specify how many iterations of the model will be run in order to propagate the influence of parameter uncertainty on the simulation results. A dialog box at the bottom of the application interface will give an estimated run time for the currently selected option. We recommend running at least 1000 iterations, but fewer can be specified as long as there are at least 100.
8) *Set a threshold for the maximum acceptable number of collisions.* If desired, specify a threshold for the maximum acceptable number of collisions (this number can be zero). The application will show this value alongside the results for reference and the probability of exceeding this value in a dialog box.
9) *Run CRM.* The button to run the CRM appears when the minimum conditions for running the model are met. For technical details on how SCRAM determines minimum conditions, see Appendix VI. A status bar will update progress through the specified species and turbine models. If this button is clicked in error, clicking “cancel” will stop the current run.
10) *Run sensitivity analyses (optional)*. Once the model has been run successfully, SCRAM provides the option to run a simple sensitivity analysis that determines the relative contribution of the input parameters to the uncertainty bounds of the results. We have provided the option as a general guide for determining which parameters are likely to lead to the biggest gains in our understanding of collision risk for the species and arrays of interest if more precise data were obtained. The analyses provide estimates of the

proportion of the variation in the results that is contributed by each parameter, for the 10 most influential parameters. For example, if the value for turbine avoidance rate is 0.12, approximately 12% of the width of the final uncertainty bounds is a result of uncertainty in our understanding of avoidance behavior. The analyses provided in SCRAM are approximate (see Borcard 2002 for details on the methods) due to computational constraints, so we recommend using this option as a rough guide for understanding parameter sensitivity or potentially planning additional sensitivity analyses. We do not recommend using running sensitivity analyses when the number of iterations is less than 1000. The results are saved to a .csv file that can be exported from the application using the "Download model runs" button.

11) *Generate summary report and/or download results for iterations.* Once the model has run successfully, which will be indicated in a dialog box at the bottom of the interface, the option to download the full results will appear ("Download model runs"). Downloading these files allows users to create their own visualizations from the model output. Clicking "Download model runs" will download multiple .csv files: the sensitivity analysis results (if run), the stochastic draws of all input parameters for each iteration, the number of monthly and annual collisions for each iteration, and an RData file with results for further use in R as needed.
12) *Check the CRM results.* Check the results of the model to see that they are sensible.

The model may be run again at this point by selecting new species, uploading new data, or choosing a new option (you can also refresh the browser if you prefer to start with a clean slate).

Tips for interpreting the results

For a more detailed discussion of the strengths and limitations of SCRAM and other CRMs, see Appendix VII. SCRAM provides several types of visualizations to aid with interpreting results and the option of downloading spreadsheets and RData files of the raw model output (see step 10 above). The application interface also shows a plot of the projected number of collisions per year. These annual projections include projections for months for which data were provided (highlighted in purple). All figures include uncertainty, either summarized as bars or shown as the spread of model iterations. In this CRM framework, uncertainty – i.e. variation in the results among iterations – is a result of the variance estimates provided for the input parameters. Increasing the number of iterations will give more accurate estimates for the model outputs, until the error associated with estimating outputs via stochastic simulation is arbitrarily small, which is around 10,000 iterations for this model.

The plot on the application interface shows the number of simulation iterations associated with the values on the x-axis, with wider distributions being a result of greater uncertainty in the number of collisions per year. We have also provided the option to summarize this uncertainty probabilistically by specifying a threshold for the acceptable number of collisions per year. This threshold can be any integer between zero and an arbitrarily large number. When provided, SCRAM will calculate the proportion of iterations that are larger than the specified threshold, which is an estimate of the probability of the number of collisions exceeding it. This value is a probability, rather than a known outcome, because there is uncertainty in the input data. For example, if all of the input data were known and provided as values with 100% certainty, as opposed to estimates with error bounds, a simple yes or no could be given for whether collisions exceed the threshold. Because input data will never be known with 100% certainty, however, SCRAM provides a probability of exceeding the specified threshold, the precision of which is scaled to account for the number of iterations – i.e. SCRAM will show more significant figures as the number of iterations is increased. This probability is also dependent on model assumptions. A plain language interpretation is "X is the probability of exceeding the specified threshold, taking into account the uncertainty of the input data and assuming that the model is a reasonable description for how collisions happen in reality."

For more detailed visualizations, including variation by month, download the automatically generated report. We have included tables for the input parameters so that all results are associated with the input data, as the results of CRMs can be sensitive to certain aspects of the availability and quality of the underlying data (see step 9 above and Appendix IV).

What software is involved?

Masden (2015) adapted the Band (2012) model for the programming language of the computing software R (R Core Team 2021). Trinder (2017), McGregor et al. (2018), and Field et al. (2021) have adapted this R code. McGregor et al. and Field et al. developed online user interfaces using the R package 'shiny' (Chang et al. 2021), which allows users to run computational tasks in R on a remote server.

References

Band, B. (2012). *Using a collision risk model to assess bird collision risks for offshore windfarms*. The Crown Estate as part of the Strategic Ornithological Support Services programme, project SOSS- 02.

Bengtsson, H. (2020). *R package "future": Unified parallel and distributed processing in R for everyone* (1.21.0). https://CRAN.R-project.org/package=future.

Borcard, D. (2002). Partial r2, contribution and fraction [a]. *Multiple and Partial Regression and Correlation*. http://biol09.biol.umontreal.ca/borcardd/partialr2.pdf.

Chang, W., Cheng, J., Allaire, J. J. et al. (2021). R package *"shiny": web application framework for R* (1.6.0). https://CRAN.R-project.org/package=shiny.

Cheng, J. (2020). *R package "promises": Abstractions for promise-based asynchronous programming* (1.1.1). https://CRAN.R-project.org/package=promises.

Judson, O. P. (1994). The rise of the individual-based model in ecology. *Trends in Ecology & Evolution*, *9*(1), 9–14.

Masden, E. (2015). *Developing an avian collision risk model to incorporate variability and uncertainty*. Scottish Marine and Freshwater Science Volume 6, Number 14.

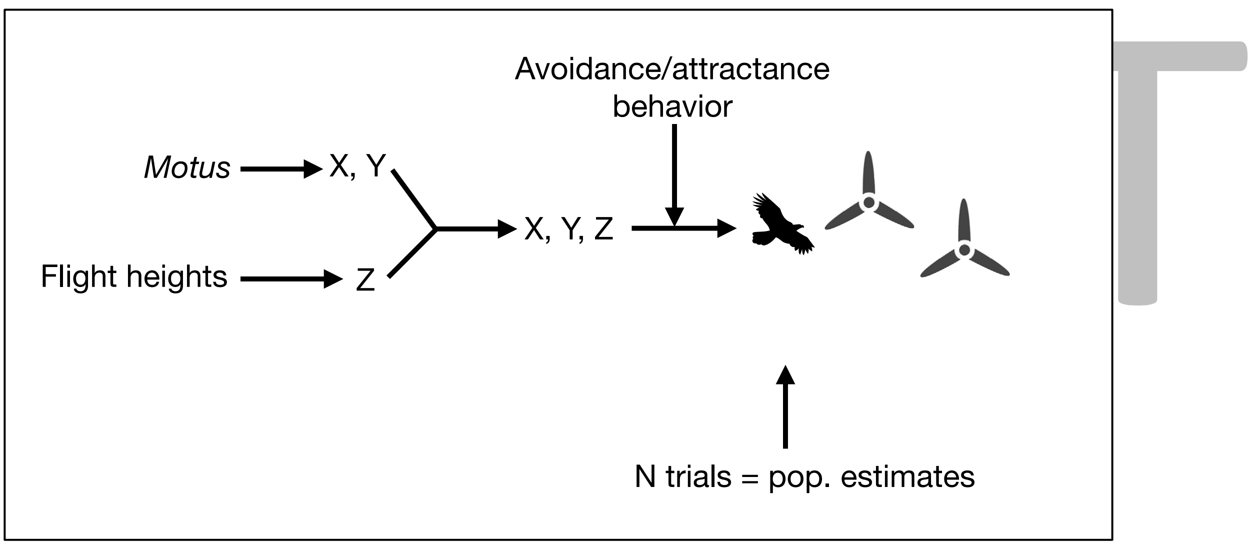
McGregor, R., King, S., Donovan, C., Caneco, B., & Webb, A. (2018). *A stochastic collision risk model for seabirds in flight.* Marine Scotland, Issue 1, Document number: HC0010-400-001.

Péron, G., Calabrese, J. M., Duriez, O., Fleming, C. H., García-Jiménez, R., Johnston, A., Lambertucci, S. A., Safi, K., & Shepard, E. L. C. (2020). The challenges of estimating the distribution of flight heights from telemetry or altimetry data. *Animal Biotelemetry*, *8*(1), 5.

R Development Core Team (2021). *R: A language and environment for statistical computing* (4.0.3). https://www.r-project.org.
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.

Appendix I. Differences between SCRAM and previous implementations of the Band (2012) model.

Our approach to developing SCRAM was to make full use of recent advancements in quantifying the potential impacts of offshore wind from Band (2012) and adaptions of the Band framework (Masden 2015, Trinder 2017, McGregor 2018). We aimed to advance the implementation of this framework in the western Atlantic by 1) contributing an improvement to the primary model script and 2) developing an online interface that best addresses the specific needs of users and stakeholders in the U.S. While there is significant overlap in the general philosophy and scripts between our version and previous iterations, there are several important differences. The most consequential change to the underlying model is that we have re-worked the data inputs to work primarily with movement data – as opposed to boat-based surveys – as these data are widely available through automated telemetry, such as the *Motus* program. In developing this new input, we also made it possible to use the Band model framework with missing values, allowing assessments to be conducted for areas or time periods with incomplete data availability.



Conceptual diagram of the primary model inputs for species data. Estimates of species movements from *Motus* data and species-specific flight height distributions determine the longitudinal (X), latitudinal (Y), and altitudinal (Z) locations of individuals over the course of a year, which in turn determines their risk of encountering rotor space. These estimated movements are combined with regional population estimates, corresponding to the sampling populations of the *Motus* projects, to estimate the total number of individuals likely to encounter rotor space for each month. Monthly collision risk is then estimated using the same assumptions as the other CRMs that are based on the Band framework (this component of the model is denoted by the square that contains turbines).

We have also revised several components of the primary model script, including an adjustment to how flight height distributions are integrated with risk along the rotor blade. We modified this component to treat flight heights as a statistical distribution, as opposed to point-wise sampling along the range of flight heights (see bullets below for more information). All of the changes to the underlying model code are tracked on GitHub.

The largest differences between our version and previous iterations relate to the delivery of the primary model script via the online interface. Our general philosophy was to make the interface as simple as possible, with most data input being accomplished using .csv files that the user can store locally, as opposed to data input on the interface itself. We also designed the tool to encourage linear advancement through the model specification process. We accomplished this by 1) having a defined, but not required, order of operations and 2) providing specific, evidence-based guidance to identify the most appropriate model option while discouraging the use of more than one option at a time. We have also created more points of dialog with the user and functionality that address the inconvenience of potentially long run times with this CRM framework. SCRAM provides the take-home results on the application interface, but most of model output is delivered via downloads of the raw results or an automatically generated report that contains visualizations and input data tables. A more comprehensive list of differences between our version and previous iterations is given below.

Major differences in primary computational script

- SCRAM integrates the flight height distributions with risk along the rotor blade using cell-wise instead of point-wise probabilities. The consequence of this change is that the first probability of the flight height distribution (labeled as 1 m) corresponds to the band that is 0 – 1 m above sea level.
- SCRAM's primary species input is *Motus* data and therefore it uses movement estimates rather than bird passage rates per unit time (referred to as "flux" in Band 2012).
- The primary computational script was revised to include a preamble that conducts a set of checks on the input data sources to ensure they are uploaded correctly.
- SCRAM allows the user to conduct an approximated global sensitivity analysis to quantify the contribution of input data to the uncertainty bounds of the results.
- SCRAM allows for missing values (specified as NA) in the input data. This is useful, for example, when movement data are not available for every month. Missing values are automatically propagated through the model and displayed in the results accordingly.
- SCRAM calculates total operation time as wind availability*(1 – down time) to avoid the fact the negative values can theoretically happen with the original formulation (wind availability – down time).
- SCRAM estimates rotor speed using the relationship between tip speed ratio (TSR), wind speed (S), and rotor diameter (r): w = (TSR*S)/r. w, which is in radians/s is converted to rpm.
- Fixed an error in the Riemann sum for rotor risk (used for the "extended" version of Band [2012]) that was causing a redundant loop

Major differences in online interface

- SCRAM's interface was built from the ground up, focusing on simplicity and encouraging a linear path through the tool.

- Only the most appropriate options in SCRAM are available to the user, depending on the input data and model specifications, to minimize the chance of running the model in a way the user did not intend.
- The majority of SCRAM's results are given in a downloadable report rather than on the application interface.
- SCRAM includes additional warnings to caution the user about applying the application to areas of with sparse data or low confidence.

Technical differences

- The history of changes since Trinder (2017) available on GitHub.
- We replaced functions that use 'dplyr' in favor of base R equivalents to minimize the number of packages needed to run the script.
- The primary computational script is run asynchronously, using the framework of making "promises" with the 'promises' and 'future' R packages (Bengtsson 2020, Cheng 2020), to allow multiple users simultaneously and the ability to cancel computational tasks.
- SCRAM allows the user to download inputs and outputs from every iteration of the model run.
- Simplified input so that a global avoidance is used instead of option-specific rates
- Tidal offset and nocturnal activity are no longer user-specified parameters

Appendix II

Table 1. Turbine and array characteristics. For the input dataset, each characteristic, and when appropriate its associated uncertainty, is specified in a column. The table should be 54 columns wide. Each row gives the specifications for a turbine array of interest, so the number of rows should be equal to the number of different arrays.

Turbine parameter name	
Run	The model run value
Num_Turbines	The number of turbines in the wind farm array
TurbineModel_MW	Megawatt rating of turbine model
Num_Blades	Number of blades for the turbine model
RotorRadius_m	Rotor radius (hub to blade tip; m)
RotorRadiusSD_m	Standard deviation of rotor radius
HubHeightAdd_m	Hub height (m)
HubHeightAddSD_m	Standard deviation of hub height
BladeWidth_m	Chord width of blade (m)
BladeWidthSD_m	Standard deviation of blade width
WindSpeed_mps	Wind speed rating of turbine model (m/s)
WindSpeedSD_mps	Standard deviation of wind speed rating
Pitch	Pitch angle of blades (degrees relative to rotor plane)
PitchSD	Standard deviation of pitch angle of blades
WFWidth_km	Wind farm width (km)
Latitude	Latitude (decimal degrees)
Longitude	Longitude (decimal degrees)
Prop_Upwind	Proportion (0 - 1) of birds flying in upwind direction
MonthOp	Wind availability (maximum amount of time turbines can be operational/month). One column for each month, e.g., JanOp, FebOp…
MonthOpMean	Mean time that turbines will not be operational (“down time”), assumed to be independent of “MonthOp” – i.e. total operation = MonthOp*(1 – MonthOpMean). One column for each month, e.g., JanOpMean, FebOpMean…
MonthOpSD	Standard deviation of mean operational time. One column for each month, e.g., JanOpSD, FebOpSD…

Table 2. Species characteristics. For the input dataset, each characteristic, and when appropriate its associated uncertainty, is specified in a column. The table should be 10 columns wide. Each row gives the values for a species of interest, so the number of rows should be equal to the number of species.

Species parameters	
Species	Species name for associated data.

Avoidance	Proportion of birds that avoid turbines
AvoidanceSD	Standard deviation of avoidance estimate
Body_Length	Body length of target species (m)
Body_LengthSD	Standard deviation of body length
Wingspan	Wingspan of target species (m)
WingspanSD	Standard deviation of target species
Flight_Speed	Flight speed of target species (m/sec)
Flight	Flight mode (“flapping” or “gliding”)

Table 3. Flight height data example. This dataset specifies the estimated flight height distribution for the species of interest. The flight height distribution gives the relative probabilities of an individual flying at each height across the range of possible heights, at 1 m intervals from 1 - 1000 m. Columns are samples from the uncertainty distributions of the relative probabilities, which can be bootstrap samples or draws from a posterior distribution. The table should contain 1000 rows (not including the header). The number of columns is flexible, but we recommend at least 100 to adequately represent the sampling distributions.

Species	**Height_m**	**BootId_1**	**BootId_2**	**BootId_3**	**BootId_4**	**…**	**BootId_1000**
Roseate_Tern	1	0.10	0.08	0.10	0.08		0.09
	2	0.09	0.08	0.09	0.07		0.08
	.						
	500	0.08	0.07	0.08	0.07		0.04

Table 4. Count data example. This dataset specifies regional population sizes associated with the movement dataset for each species of interest. Columns specify the mean and standard deviation for the estimated population size, which can vary by month. The table should be 25 columns wide. The number of rows should be equal to the number of species.

Species	**Jan**	**JanSD**	**Feb**	**FebSD**	**Mar**	**MarSD**	**…**	**Dec**	**DecSD**
Roseate_Tern	80	15	70	12	65	14		86	19
Common_Tern	1200	20	1400	40	900	30		950	25

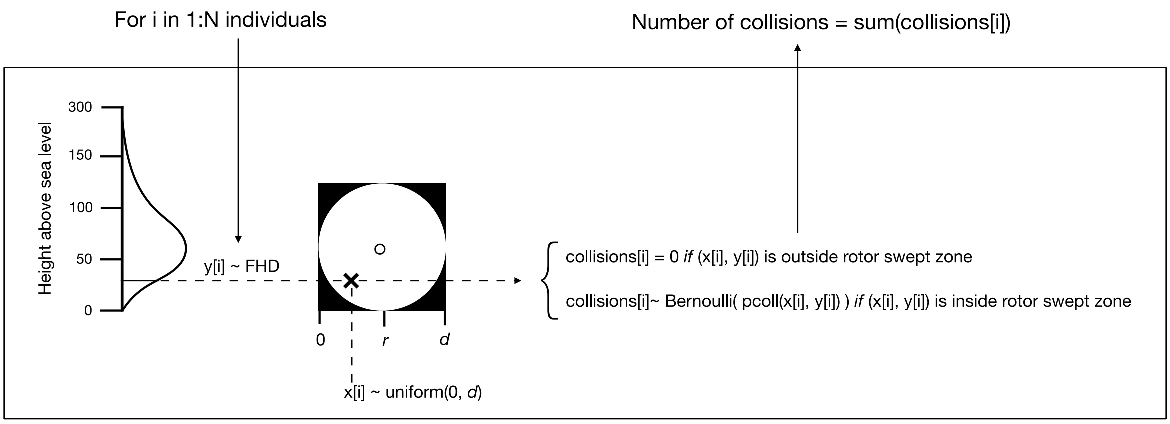
Table 5. Format of underlying movement data, which are “baked in” to the tool and do not need to be provided by the user. This dataset specifies the estimated probability that an individual from the target population will pass through the target area in each month. Each row is a sample from the uncertainty distributions of these estimates, which can be a bootstrap sample or a draw from a posterior distribution. The table should be 13 columns wide. The number of rows is flexible, but we recommend at least 100 to adequately represent the sampling distributions.

Species	**Jan**	**Fed**	**Mar**	**Apr**	**May**	**Jun**	**Jul**	**Aug**	**Sep**	**Oct**	**Nov**	**Dec**

Roseate_Tern	0.09	0.11	0.12	0.07	0.05	0.10	0.12	0.06	0.08	0.11	0.12	0.05
	0.08	0.12	0.09	0.05	0.03	0.14	0.13	0.04	0.10	0.11	0.12	0.04
	.	.	.	.	.	.	.	.	.	.	.	.
	0.07	0.13	0.11	0.06	0.03	0.12	0.12	0.07	0.09	0.10	0.11	0.04

Appendix III. Description of individual-based options.

All previous implementation of the Band (2012) CRM framework (Masden 2015 and Trinder 2017) estimate collision risk using double integration of the flight height distribution and risk along the rotor blade when the “Extended” options is selected (see Appendix V for the equations). Because the analytical solution to this double integral is difficult to obtain, the Band, Masden, and Trinder model approximates the integrals using a Riemann sum (similar to point sampling along a grid over the rotor area). The “standard” options (Band 2012; see Figure 1) also eliminate the need for one of the integrals by reducing the flight height distribution to a single estimate of flying at risk height. The advantage of Band’s (2012) integration approach is a simple theoretical specification for collision risk. The implementation of this component of the model, however, is relatively computationally expensive as it involves looping across the x and y dimensions of rotor area. The integration approach is also potentially limited in terms of model extensions, since the solution can quickly become intractable as complexity is added. For example, increasing evidence suggests that flight height distributions should be multi-state to account for time, seasonal, or weather-based variation in flight heights (e.g. Péron et al. 2020). Such time and space variant flight height distributions might be difficult to specify as an easily approximated integral. Given that any implementation of this modeling framework involves computation, we explored alternative options for estimating collision risk that would either provide more accurate estimates or potentially be easier to adapt for future iterations of this modeling framework. The first alternative, obtaining the analytical solution to the double integral, was prohibitively time intensive and therefore is only used as the “correct” answer against which to compare the quicker options (see Appendix V). The second alternative is a generalized version of the collision risk framework that reduces the processes that determine an individual’s risk into their basic statistical components and uses an individual-based approach to scale these processes to the population level. This “individual-based” option is conceptually simple and does not require integration, but within the Band, Masden, and Trinder framework can require long computation times when many individuals are potentially at risk. One advantage of this approach is that because the underlying statistical processes are specified explicitly, it is possible to estimate the potential influence of environmental or behavioral stochasticity on collision risk, in addition to the influence of parameter uncertainty (which the model already accounts for). A diagram of the individual-based approach is shown below.



If the previous components of the model suggest that individuals will enter the airspace of the rotors, the fate of each individual is tracked through the five steps represented in the diagram:
1) An individual's flight height is drawn from the flight height distribution.
2) An individual's position along the rotor's x axis is randomly drawn. The length of this axis is determined by the diameter, *d*, or radius, *r*.
3) These x, y coordinates determine whether the individual will pass above, below, or to the sides (the black spaces) of the rotor.
4) Alternatively, if the individual's x, y coordinates lie in the rotor swept zone, it faces a collision trial, represented by a Bernoulli distribution. The parameter of the Bernoulli distribution is the probability of a collision, *pcoll*, at the individual's location, which is determined by the equations that estimate relative risk across the rotor area from Band (2012).
5) This process is repeated for every individual that enters the airspace of the rotors, the sum of which is the estimated number of collisions.

While this option is not yet provided in the SCRAM web application, we have included it in the source code as a potentially useful foundation for future model extensions, which are likely to involve allowing for additional complexity over time and space. We also use these options as a point of comparison for simulations of the trade-off between accuracy and computation time (Appendix V). When computational time is not a constraint, the individual-based options will provide the most accurate estimates, although gains in accuracy might be marginal in light of the uncertainty from the input data (see Appendix V).

Appendix IV

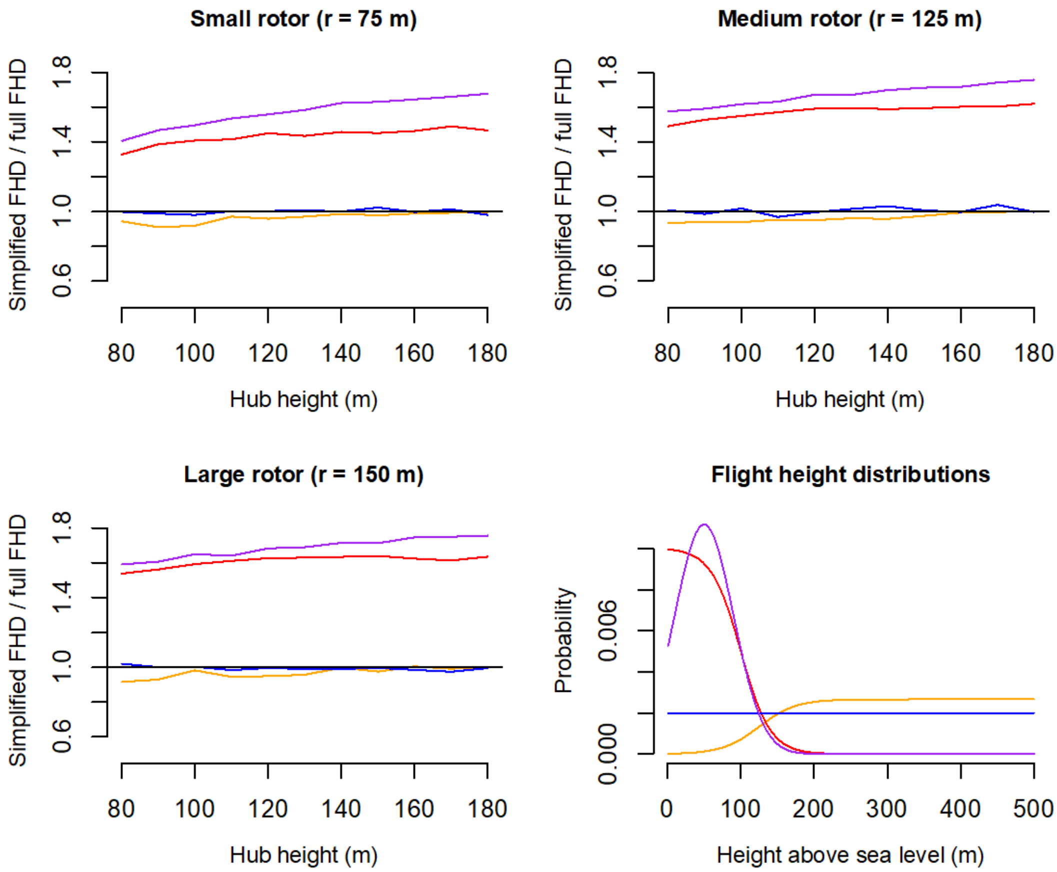
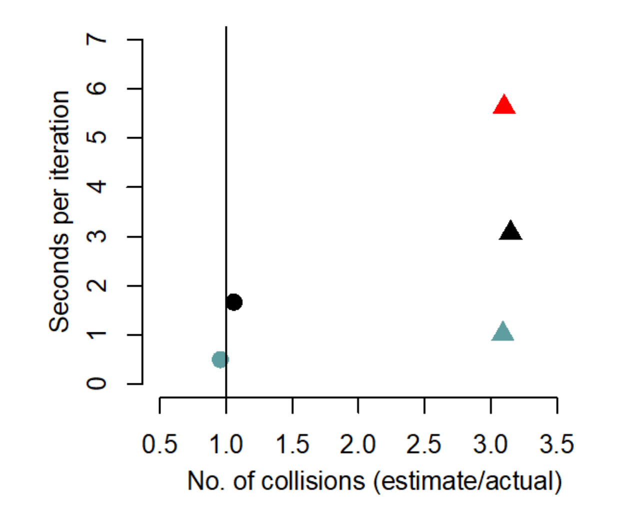


Figure 2. Demonstration of the potential influence of model option on the estimates of the number collisions/year, for a range of shapes for the flight height distribution. We ran the model, for Red Knot as a test case, with identical inputs using 1) the option that integrates the full flight height distribution (FHD) and 2) the option that summarizes the FHD to obtain a single value for the proportion of birds flying at risk height. We then obtained the results from this latter option (2) as a proportion of the former (1). The resulting variable gives an estimate of the difference between options in terms of what percentage the simpler option is off in either direction. For example, a value of one means the options gave the same answer, while 1.8 means the simplified FHD gave an answer that was 80% larger than the more realistic option. Both options used the same input data and identical flight height data so that the only difference was that the simpler option summarized the FHD instead of using the full distribution (see Appendix V for more information on the underlying calculations). The two options compared here are related to the “extended” and “standard” options of Band (2012) (“Yes” and “No” for “Use complete flight height data?”, respectively, in SCRAM; see Figure 1). We ran the simulation for a range of hub heights, four shapes for the FHD, and three rotor sizes (75 m, 125 m, and 150 m) to quantify the influence of these variables on the differences between model options. The colors of the lines relate to the four FHD shapes shown in the bottom-right plot.

We recommend using the full FHD when running SCRAM for the tool's four default species ("Yes" for "Use complete flight height data?"). When uploading data for additional species, we recommend first identifying the general shape of the FHD (e.g. skewed low or high, bell shaped, uniform, etc.). When data are not available, it might be possible to use data from a similar species of the same taxonomic group for the comparison to these simulation results. Johnston et al. (2014) provide FHDs for a wide range of species that might be impacted by offshore wind. Once a general shape for the FHD has been identified, we recommend finding the appropriate rotor size and hub height on the plots to determine whether this is likely bias associated with not using the complete flight height data.

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. *Journal of Applied Ecology*, *51*, 31–41.

Appendix V



Full integrated version

$$N * 2/pi \int_{-1}^{1}\int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} d(y)\, p(xy)\, dx\, dy$$

Integration (analytical)

▲ (red) $$N * \left(\int_{-1}^{1} d(y)\, dy\right)/2 * 2/pi \int_{-1}^{1}\int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} p(xy)\, dx\, dy$$

Integration (approximated)

● (blue) $$N * 2/pi\left(\sum_{y=-1}^{1}\sum_{x=-\sqrt{1-y^2}}^{\sqrt{1-y^2}} d(y)\, p(xy)\right)/xinc$$

▲ (blue) $$N * \left(\sum_{y=-1}^{1} d(y)\right)/2 * 2/pi\left(\sum_{y=-1}^{1}\sum_{x=-\sqrt{1-y^2}}^{\sqrt{1-y^2}} p(xy)\right)/xinc$$

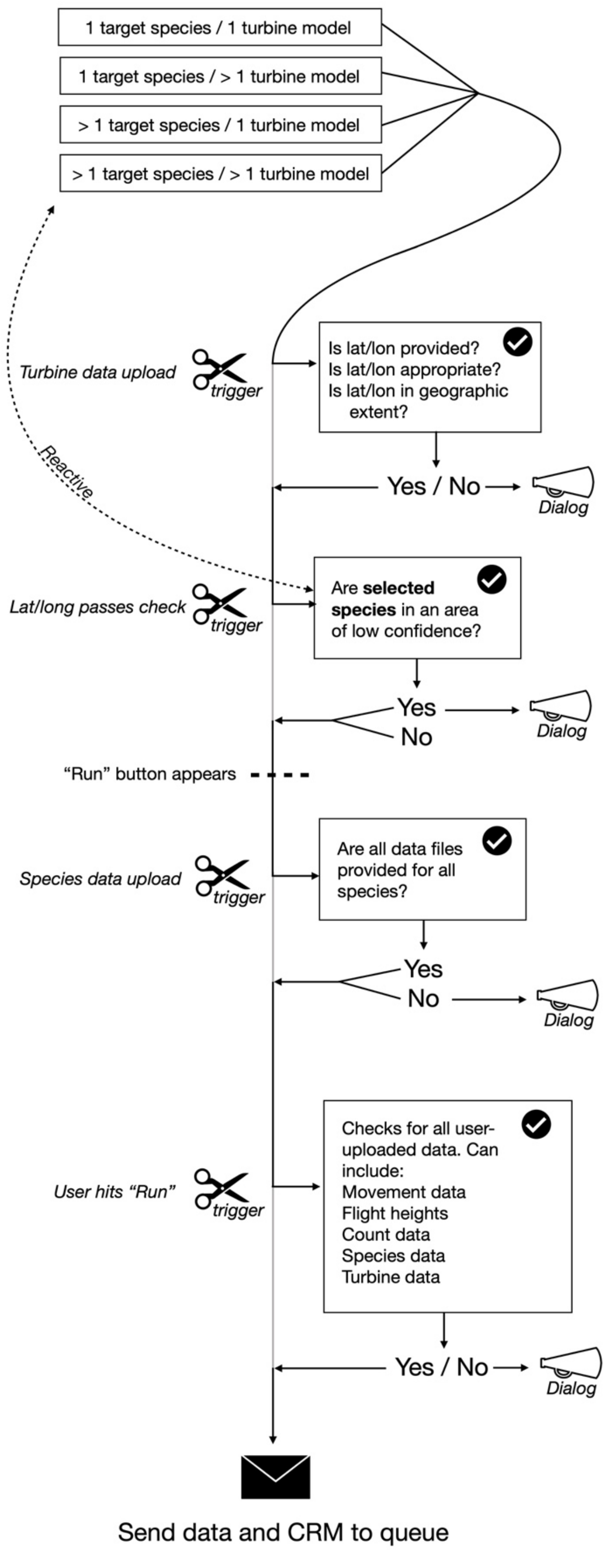
Individual-based

● $$\sum_{i=1}^{N} rmultinomial(d) * rBernoulli(p(x_i, y_i))$$

▲ $$\sum_{y=\min(rotor)}^{\max(rotor)} d(y) * \sum_{i=1}^{N} rBernoulli(p(x_i, y_i))$$

Speed vs. accuracy for five options for estimating collision risk. The five options' estimates of the number of collisions over the course of one year, assuming no avoidance, are compared to the full double integral of Band (2012) (eqs. 9 and 10), which does not use approximations or simplifying assumptions. The options are categorized by their approach to estimating the full integral: blue is an approximation of the integral ("Yes" for "Use complete flight height data?" in SCRAM); red uses an analytical solution for the integral for rotor risk (not available in SCRAM); and black is an individual-based population simulation that uses a probabilistic framework instead of integration (not available in SCRAM). Versions of the options that estimate the proportion of birds flying at risk height and the risk over the rotor disk are shown as triangles ("No" for "Use complete flight height data?"). Versions that estimate these probabilities in a combined fashion, either by double integration or in a probabilistic framework are shown as circles ("Yes" for "Use complete flight height data?"). For each option, a general form of the equation that estimates the number of collisions is shown for reference.

Appendix VI



The decision tree that conducts data checks and determines the minimum conditions required to run SCRAM’s collision risk model. At predetermined triggers (denoted by a scissors icon), a series of checks are performed, the results of which determine whether the user is allowed to proceed. After each check, a pop-up dialog box (denoted by a megaphone icon) will appear to warn the user if a certain condition is met. For example, if the user-specified data passes a check for whether latitude and longitude are appropriate, a trigger will activate a check for whether the coordinates are in an area of low confidence for species data. If the check results in a “Yes”, a dialog box will appear warning that caution should be used when interpreting the results. The user must “OK” out of the dialog box and is then allowed to proceed.

Appendix VII

A starter guide for understanding the strengths and limitations of collision risk models

Christopher Field[1], Pam Loring[2], Brian Gerber[1]

University of Rhode Island[1]

U.S. Fish and Wildlife Service[2]

The aim of this guide is to clarify a few key points about the collision risk models to help potential users interpret their results and better understand how they can use emerging tools, like the recently developed web application SCRAM, to inform their work.

The aim of collision risk modeling

Collision risk models (CRMs) generate projections for the likely number of bird collisions at wind farms, which can inform planning, mitigation, and assessments of the cumulative impacts of increased mortality for at-risk species. The primary benefit of CRMs is that they generate this evidence at the earliest stages of wind farm planning, when there is the greatest potential to influence the design of turbine arrays to minimize risk. During this critical time in the life span of a wind farm, it is not possible to collect data on collisions *per se*. Data collection is therefore limited to the factors that will likely influence collisions, including passage rates and flight heights. Projections models are one way to synthesize these data, which are likely to be essential components of pre- and post-construction monitoring, to provide estimates of the quantity we are often most interested in – collisions. Using CRMs to project the number of collisions also makes it possible to identify turbine and array characteristics that are most likely to influence collision risk. The greatest utility of CRMs is during the planning, leasing, and construction phases of wind farms. Over time, the projections from CRMs can be validated, and eventually replaced, by direct monitoring of collisions during operation (Figure 1).

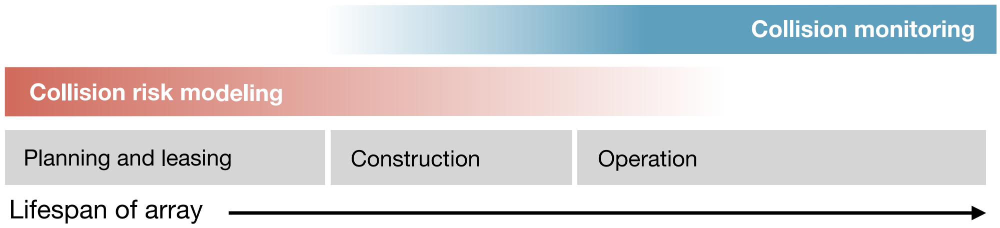


Figure 1. A conceptual diagram showing relative utility of collision risk modeling vs. direct monitoring of collisions over the lifespan of an offshore wind farm. Collision risk modeling, which is necessary pre-construction, becomes less important as collision monitoring becomes possible.

Intended audience

The intended audience for SCRAM, an adaptation of Band (2012), is anyone with an interest in understanding collision risk from wind for decision-making, planning, policy, or environmental assessments. This audience includes conservation practitioners, state and federal agencies, non-governmental organizations, and industry. The web application was developed for users who do not have previous experience with statistical or computational modeling, but as with previous iterations of the Band model (Masden 2015 and Trinder 2017), the code for both the CRM and web application are open source.

What causes CRMs to be wrong?

The value of CRMs for planning depends on how successful they are at synthesizing the available data and generating reliable estimates of collision risk. When considering the performance of CRMs, we think it is helpful to distinguish between bias caused by misspecification of the CRM vs. bias in the underlying data. Bias and uncertainty in data for offshore areas are a general roadblock to understanding the impacts of wind, and CRMs are not more sensitive to these data limitations than other approaches for understanding potential impacts, such as verbal models. The reliability of a collision risk modeling framework is, instead, determined by 1) how it handles uncertain data and 2) whether the assumptions underlying the translation of these data sources to collision risk are appropriate.

Incorporating uncertain data is straightforward, as current implementations of the Band (2012) framework for collision risk modeling are stochastic – i.e. the bounds of the collision risk estimates accurately reflect the uncertainty of the input data. This framework even allows a user to, if appropriate, specify complete uncertainty for a parameter with no reliable data input (e.g. behavioral avoidance rates for some species). Most of the assumptions of CRMs – i.e. the processes that a CRM captures through its model specification – are straightforward or based on a strong body of evidence (e.g. the importance of the distribution of flight heights; Johnston et al. 2014). An important feature of the Band (2012) CRM framework is that the assumptions of the model are all transparent, straightforward, and based on first principles, making misspecification of the CRM unlikely. For example, at one step of the model the number of birds at risk is narrowed by the proportion of birds flying at risk height, which is simple (and uncontroversial) multiplication. The rotor risk component of the CRM, which models the process of birds flying through a spinning turbine blades, is based on simple, verifiable geometry. In fact, modeling these geometric relationships is the only means of making inferences about risk that incorporate the influence of blade length, shape, and speed. Other processes, such as correcting for behavioral changes around large arrays, are reasonable, but would benefit from validation either through additional research or direct validation of CRM projections when data are available. For some of these cases, such as large array corrections, the underlying assumptions are only likely to be consequential under a narrow set of circumstances (Band 2012).

When CRMs generate projections that do not accurately reflect reality, it is usually the result of bias in the underlying data, which can be caused by biases in collection or analysis. For example, several inputs to CRMs, both historically (flight height distributions) and with the SCRAM implementation (movement tracks), use estimates from statistical models. The aim of these statistical analyses is to find the best balance between uncertainty and bias. While uncertainty is fully accounted for in a stochastic CRM, any bias that remains falls under the category of “known unknowns”, highlighting the importance of identifying potential bias whenever possible. The potential issues associated with biased data are not limited to their use in CRMs, however. For example, if the estimated flight height distributions are skewed, the potentially skewed estimates of collision risk are a symptom of a general problem arising from our biased understanding of flight heights, and would affect any inference about risk including simpler or verbal models.

A related source of bias is micro-scale deviances from the input data caused by a spatial or temporal mismatch in the scale of the array of interest and the underlying data. For example, with SCRAM we are facilitating the use of CRMs by providing all necessary inputs to generate projections for any location in the Atlantic Outer Continental Shelf for the target species. The underlying statistical models are aiming to capture the large-scale processes that drive variation across this planning area, and therefore might not fully capture micro-scale variation. SCRAM does, however, allow the user to supersede the built-in data if higher resolution data exist for a particular site.

The strengths of CRMs

Identifying the strengths and limitations of CRMs and applying them in the appropriate contexts is key to using them successfully. Inference in the Band (2012) framework is focused on individual arrays, which makes it well-suited for providing information on planning at this scale, including turbine models and characteristics, spacing, and operation time. Questions that the Band model is well-suite to address, for example, are: What is the best estimate of Red Knot “take” for a 200 MW array of GE Haliade X turbines that is placed 20 km off the coast of Massachusetts? By how much does re-powering an existing array with the latest technology change risk?

The Band model can be run as many times as needed to estimate risk for multiple arrays or different time periods. It does not, however, contain any methods for explicitly quantifying cumulative risk over space and time, making it less suited, in its current form, for questions about scenarios of total power capacity across the Altantic OCS or questions about changing conditions such as bird population trends and range shifts.

Data limitations specific to the current version of SCRAM

SCRAM is an evolving tool that will be updated as additional data and methods become available. SCRAM's CRM currently uses static flight height distributions, as opposed to distributions that vary over space and time, which are more realistic (Péron 2020). While we feel that treating flight heights as a statistical distribution, as we have done, implicitly accounts for some of this potential variation, more research is needed to determine whether unexplained variation is likely to influence collision risks (e.g. if flight heights increase substantially with increasing distance from land). One particular challenge to estimating passage rates from movement data is determining the overall number of individuals that could potentially encounter arrays. While we are using the latest regional population estimates for the target species, there are currently limitations in our knowledge of how representative movement data are of these populations. As with other data sources, users can supersede the built-in passage rate estimates with site-level data, allowing formal comparisons and tests of sensitivity to the input data.

References

Band, B. (2012). *Using a collision risk model to assess bird collision risks for offshore windfarms*. The Crown Estate as part of the Strategic Ornithological Support Services programme, project SOSS- 02.

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. *Journal of Applied Ecology*, *51*, 31–41.

Masden, E. (2015). *Developing an avian collision risk model to incorporate variability and uncertainty*. Scottish Marine and Freshwater Science Volume 6, Number 14.

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.

Péron, G., Calabrese, J. M., Duriez, O., Fleming, C. H., García-Jiménez, R., Johnston, A., Lambertucci, S. A., Safi, K., & Shepard, E. L. C. (2020). The challenges of estimating the distribution of flight heights from telemetry or altimetry data. Animal Biotelemetry, 8(1), 5.