HABITAT MODELLING USING SATELLITE TRACKING DATA:

A WORKSHOP TO ADVANCE THIS PROCESS

Convened by BirdLife International and IMAR-CMA (University of Coimbra)

21 – 23 September 2011

University of Coimbra (Portugal)

BirdLife INTERNATIONAL
Together for birds and people

LENFEST OCEAN PROGRAM

UNIVERSIDADE DE COIMBRA

Citation: BirdLife International (2011). Habitat modelling using satellite tracking data: a workshop to advance this process. Results from a workshop held 21 – 23 September 2011 at University of Coimbra, Portugal. BirdLife International report. Cambridge, UK.

Organizers: Ben Lascelles (ben.lascelles@birdlife.org), Iván Ramírez (ivan.ramirez@birdlife.org), Vitor Paiva (Local contact vitorpaiva@ci.uc.pt), Jaime Ramos, José Xavier, Phil Taylor, Mark Miller

Introduction

Identification of "hotspots of activity" (e.g. as BirdLife marine Important Bird Areas and/or Convention on Biological Diversity Ecologically or Biologically Significant marine Areas) for wide ranging, pelagic species (such as seabirds) is likely to rely heavily on tracking data. In recent years major advances have been made in how best to interpret the data, using a variety of analytical techniques, to draw the most robust conclusions about the most important areas.

To assist this process, in 2009, BirdLife International and CNRS, Chize, France convened a workshop to investigate the use of satellite tracking data for identifying marine IBAs. BirdLife has since developed a methodology for undertaking this type of analysis. However, approaches rely heavily on the extent of data and can only identify sites for populations represented by them. For these reasons, the uses of additional analytical techniques are being investigated. Habitat modelling has been identified as a key tool for refining these sites, and currently offers the best opportunity for predicting the location of hotspots in unsurveyed areas. Habitat modelling is thought to be particularly useful for:

- Better understanding the extent of suitable habitat and (potential) distribution of the population at present, past and futures (spatio-temporal predictions)
- Understanding animal movement (during breeding, migration) and identify key relationships between species and their environment.
- Inferring potential sites away from sampled areas (spatial predictions).
- Helping refine site boundaries and placement
- Potentially providing estimates of abundance within a species distribution.
- Understanding the habitat use by seabirds within EEZ's or other jurisdictional boundaries

Similar questions are also being raised in the scientific research community where telemetry studies have moved away from simple description of animal distributions, and toward empirically explaining and predicting distributions under future climate scenarios and for other populations.

A number of habitat modelling techniques are now published and there has been widespread application using conventional data types (such as at-sea survey data). However, their application for more complicated data types, such as telemetry data, is less well advanced. To this end, BirdLife International and IMAR, University of Coimbra convened a workshop to bring together scientists familiar with seabird ecology, telemetry data and habitat modelling techniques to discuss approaches and methodologies to advance the scientific best practice techniques for habitat modelling, and identify the approaches which could be readily applied by BirdLife in its marine IBA work.

Aims

With such issues in mind, the workshop was organised to bring together scientists and researchers working with tracking data to discuss methodologies and approaches to using the data in habitat modelling. The aim was for workshop participants to make recommendations about how they can best be applied to the identification of "hotspots of activity", particularly for seabirds, but also drawing on experience from analysis of, and applicable to, other mobile pelagic taxa.

Specific aims were:-

- To develop *'optimum'* (i.e. the most rigorous and comprehensive) methodologies for modelling habitat preference using tracking data,
- To develop *'pragmatic'* approaches for use in designation of marine IBAs and EBSAs, considering the necessary balance between reliability of results and ease of application and taking particular account of logistical and sample size restrictions.
- To develop data standards regarding input of tracking data and environmental variables to habitat modelling approaches
- To compile relevant information in a document/report describing the available approaches, the circumstances when they are most suitable, and some of the limitations in their use.
- The proceedings from the workshop will be captured in a 'layman's report' for use at the science/policy interface, as part of BirdLife's marine IBA toolkit.

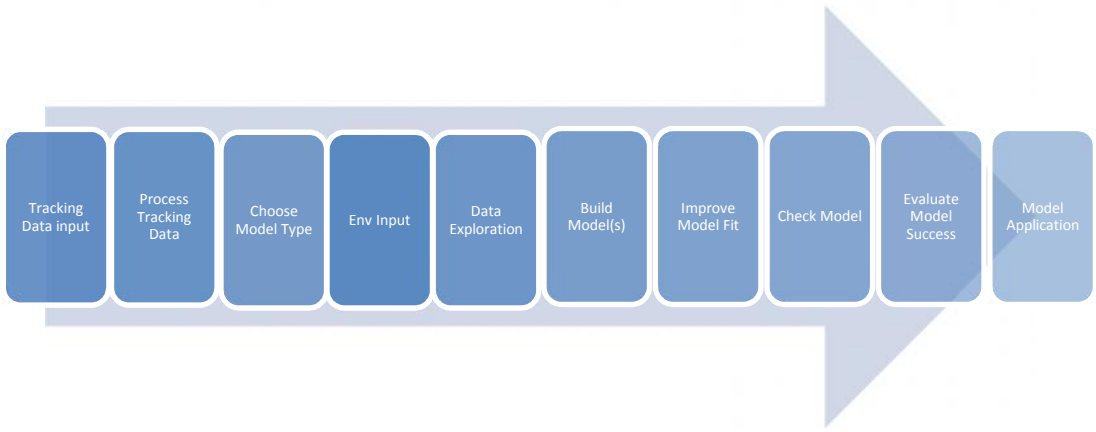

Figure 1: Flow chart showing various stages of a habitat modelling process. At each stage decisions and interpretation of data are required if resulting models are to be successful and meaningful.

The habitat modelling process may be broken into stages, as per the graphic in Figure 1 above. During the workshop the various options available under each stage were presented, and the benefits and limitations of each choice discussed. The aim was to place existing experiences within this framework to see which combinations are already well understood, and identify gaps that may be explored in the future.

Summary of the workshop

The first half-day of the workshop was a public open session, and was attended by workshop participants as well as students and staff from the University of Coimbra. Around 100 people attended this session.

This session provided an introduction to the work of BirdLife's Global Seabird Programme, the history and progress made in identifying marine IBAs within the BirdLife Partnership and highlighted the importance of ensuring that habitat modelling approaches can be easily applied and shared throughout the BirdLife Partnership.

Additional presentations highlighted the utility of habitat modelling in understanding and predicting seabird distributions, as well as highlighting the importance of ensuring approaches and their outputs were relevant and understandable to policy.

A summary of studies showed how modelling can improve our understanding of seabird ecology and be useful for predicting future changes. For example long term studies have shown that habitat modelling can be used to predict the implications of climate change (La Bohec et al, 2008).

The importance of relating distributions and analyses such as habitat modelling to feeding ecology and actual biological processes highlighted that it is important to keep such ecology in mind and not focus only on the technical approach.

The rest of the workshop was limited to invited participants and included a combination of talks that discussed various modelling techniques in greater detail, discussion sessions on environmental variables and break out groups to develop new ideas.

The final day was put aside as a practical session to trial some of the ideas developed.

Summary of existing experiences

The combined experience of workshop attendees was collated into table 1 below. GPS, PTT and GLS are the devices for which there is most experience, other devices may be appropriate but are becoming outdated and less frequently used (eg compass loggers and VHF) making them less practical for wide ranging species.

Data type	*Preprocessing*	*Response variable*	*Model type*	*Software used*	*Attendee*	*Reference*
Geolocator	State-space modelled, SST corrected	Presence/absence (Binary)	GAMM	R & WinBugs	ALH	
Geolocator	State-space modelled, SST corrected	Density of Individuals	GAMM	R & WinBugs	ALH	
Geolocator	SST corrected	Presence/absence (absences from CRWs)	GAM	R	RZ	Zydelis et al, 2011
Geolocator	SST corrected	Diving/non-diving	Mixed effects ANOVA	Matlab	SS	Shaffer et al, 2009
GPS	None	Feeding/non-feeding (calculated from speed)	Linear mixed model	Matlab	LT	
PTT	Speed filtered	Time spent per cell	GLMM	R	ML	Louzao et al, 2011
PTT	Speed filtered	ARS behaviour/non-ARS behaviour	GLMM	R	ML	Louzao et al, 2011
GPS	Degraded to frequency of PTT	Feeding/non-feeding (from stomach sensor)	GLMM	R	ML	Louzao et al, 2011
Geolocator	Filtered	50%UD/95%UD (Binary)	GAM	R	LT	
GPS	None	Diving (Presence only)	ENFA	Biomapper	HS	Skov et al, 2008
PTT	Speed filtered	Residency time	GLMM	R	CP	Peron et al, 2010

Table1: Workshop attendee experience with habitat modelling of telemetry data. GAM – Generalised Additive Model, GLM – Generalised Linear Model, GAMM – Generalised Additive Mixed Model, ENFA - Environmental Niche Factor Analysis, SST – Sea Surface Temperature, ANOVA – Analysis of Variance

Further studies that make use of habitat modelling that were discussed included work with Boosted Regression Trees, and 'mboost. Additional published studies not represented at the workshop are summarised in table 2.

Data type	*Preprocessing*	*Response variable*	*Model type*	*Software used*	*Attendee*	*Reference*
PTT	Speed filtered	Presence/absence (psuedo absences)	GAMM	R	--	Wakefield et al, 2011
PTT	Speed filtered	Presence/absence (psuedo absences)	GAMM	R	--	Aarts et al, 2008
PTT	Speed filtered	First Passage Time value	Linear mixed model	Matlab	--	Kappes et al, 2010
PTT	Resampled to every two days	Presence only	Maxent	Maxent Software	--	Edren et al, 2010

Table 2: Additional published experience with habitat modelling of telemetry data.

During the course of the workshop several presentations were given that introduced some of the modelling techniques in greater detail. Some of the key differences and approaches are highlighted below.

Ramunas Zydelis – GAMs to predict distributions from geolocator data. (Zydelis et al, 2011)

- Binomial GAMs were fitted to post-breeding movements of Black-footed (*Phoebastria nigripes*) and Laysan (*Phoebastria immutabilis*) Albatross (telemetered with geolocator archival tags).
- Locations were improved using Sea Surface Temperature matching algorithms.
- GAMs require absence data so Correlated Random Walks (CRWs) were used to calculate pseudo-absence locations based on the movement parameters, and within the bounding box, of the tracked birds.

- To account for locational error, environmental data were attributed to bird and pseudo-absence locations using a kernel sampling approach.
- GAMs were improved using a backwards stepwise approach looking at AIC values.

Maite Louzao – Hierarchical approaches to modelling Albatross distribution (Louzao et al, 2011)

- Three scales of animal distribution/movements were modelled, percentage of time spent within an area (normal distribution), whether the animals were foraging or passing (binomial distribution), whether the birds were feeding or not (binomial distribution). Each was used as a separate response variable in a separate model.
- The grid size used to calculate response variable (0.25º) was defined by the maximum resolution of the available environmental data and tracking device error (mainly PTT).
- GLMMs were used and were favoured over GAMMs because they allow AIC robust values to be calculated in a standard and well established statistical framework and hence it was possible to use a multi-model inference approach.
- Spatial autocorrelation was tested in "best" model residuals and it was found to be non-significant.
- Environmental covariates were excluded when correlation between environmental variables was high.

Henrik Skov – Presence only modelling (Skov et al, 2008)

- Dive locations were modelled using Environmental Niche Factor Analyses (ENFA), this is a presence only model so no absences were needed.
- Environmental data was normalised using box-cox algorithms
- Further environmental layers were calculated such as ocean fronts.
- ENFA can cope with collinearity so all data layers may be used.

Autumn-Lynn Harrison – Modelling ocean region occupancy

- GAMMs were fit to the likelihood of an animal being within a biogeographic region.
- All data was filtered with a state space model Kalman filter so that it was standardised across data types.
- Species which were harvested (e.g. the tuna) caused large 'tag bias', with many short trips but few completing a full migration. Weights, representing the number of trips at large during any given day, were applied to the data to avoid this.
- The response variable was the biogeographic region occupied (from Longhurst classifications) and the explanatory variables included day of the year.
- Large scale predictions can be made adding ecological meaning to species migrations.

Leigh Torres – Boosted Regression Trees (BRT)

- BRT applies regression tree analyses to describe and predict the associations between species and their environment.
- Unlike most other regression techniques, BRT does not fit a single 'best' model to the data, but rather fits an ensemble of many simple classification or regression tree models that partition observations into groups with similar values. The boosting algorithm iteratively adds trees to the

final model in a stage-wise procedure that emphasizes those observations poorly fit by the previous trees.
- Advantages to regression trees are that this modelling method can include continuous and categorical data as predictors, can handle missing data in predictor variables, are immune to outliers in data, can deal with irrelevant and correlated predictor variables, and can automatically fit interactions between predictors.
- BRTs can generate predictive maps of species distribution based on the fitted associations.
- BRTs avoid over fitting by using cross-validation to progressively grow models while testing predictive accuracy on withheld portions of the data.
- Boosted Regression Trees (BRTs) are a useful tool for modelling, and account for many of the issues inherent in modelling telemetry data.
- Regression trees are good at selecting relevant variables and boosting overcomes inaccuracies in any single tree.
- Cross validation is done **within** the model build and avoids over-fitting as the model grows.
- They can handle interactions between variables (thereby also tackling issues of collinearity) and can handle both continuous and categorical data.

Workshop Recommendations

As highlighted in Figure 1 the analytical process can be split into various stages and these were discussed in turn at the workshop. The paragraphs below indicate the main points of the discussions for each of these stages. Text in bold indicate recommendations.

Habitat Modelling Inputs - Response variables

- **Optimally, when different types of data exist (i.e. geolocator, PTT and GPS) tracking data would be State-space model filtered to standardise. For large databases such as BirdLife's, however, this process would take over a year to complete. (ALH)**
- **PTT data should be filtered; commonly used filters are the Freitas filter and speed filters (Freitas et al, 2008; BirdLife International, 2004; McConnell et al, 1992) (PT).**
- **There can be great benefit to applying SST correction because of issues with geolocating during equinox periods when daylength is equal across all latitudes (SS).**
- **However, care should be taken when applying SST correction when water masses are not stratified. When mixing occurs, the approach is not appropriate (RP).**
- Many response variables may be used in modelling, however, the predictions from the model will only relate to the input behaviour. It is important, therefore, that responses are relevant to the desired prediction. In terms of IBAs, it would be more interesting to determine high occurrence areas for rare species. In the case of wide-ranging species, it would be more interesting to identify ecologically important areas such as where they spend more time or seek for preys since it is not possible to protect their overall distribution range
- To obtain a representative overall distribution of a population, it would be necessary to test the representativeness of the tracking dataset via discovery curvers. (PT, MM).
- Raw point locations can be used as presences in models but may require absences to be generated, depending if a presence-only model is used (e.g. ENFA, Maxent).
- **When absences are created they should be ecologically relevant, i.e. areas accessible to the animals but actively not chosen. (PT)**
- Correlated Random walks are a commonly used way to establish pseudo-absences, they may be calculated using r scripts available online (eg. using 'circular' package)
- When creating pseudo absences the ratio of presences to absences can affect the model. 1:1 is commonly agreed as best in model building (Wood, 2005; Zuur et al, 2009). However, 5:1 (Block

et al. 2011); 10:1 (LT) and 50:1 (RZ) have all been used successfully in modelling telemetry data in the past.

- **To establish the correct ratio Zydelis et al built a number of models with incrementally higher ratios of absences to presences and investigated how this impacted the AUC of the model, the ratio at which AUC values stabilised was used. This was agreed as useful, although pseudo-absence generation still a contentious point with much discussion in the literature (Stokland et al, 2011; Lobo et al, 2010) (SO, RZ)**
- In some studies the average proportion of time spent by a tracked individual within a given area has been used as a response; however, this assumes all trips, regardless of the duration of time spent at sea, are equal, a number of other approaches accounting for the relative 'importance' of trips can be used. (CP, ML, RP)
- **Presence only models (such as ENFA and Maxent) treat 'background data' (i.e. the study area) as pseudo-absences. It is recommended, therefore, that similar rules to those defined above for creating absences, are applied here to setting the 'background' and the environmental input.**
- Models using pseudo absences perform similarly to Maxent (presence only) models. (RZ)
- **If we are modelling continuous data (e.g. time spent per cell) data may be zero-inflated. Pragmatically we can get around this skew using a hierarchical (or Hurdle) model. However, optimally we should model the data in a zero-inflated framework. (SO)**

Habitat Modelling Inputs - Explanatory variables

- All modelling requires some *a priori* decisions about which environmental variables to consider within the analyses. Assessing the variables that are important to our species and what environmental data products are available should help ensure all relevant variables are included.
- Remote sensing is the pragmatic choice for environmental variables, the optimal needs to consider water column variables as well, particularly for diving species
- **Variables considered in this type of modelling should be either;**
    - **Physical features which facilitate or constrain movement or;**
    - **Physical or biological features that enhance productivity. (SS)**

When it is available, prey distribution data should also be included. However, this is rarely the case so we rely on proxies.

- As users of environmental data we should be aware of their origins and that while some data are primary, many are derived products with their own uncertainty or with different temporal relevance. Using these data types can add uncertainty to the model and this should be accounted for. (SS)
- **Some derived products, such as primary productivity (VGPM), are products of many other variables, these products should not be used alongside the variables used to create them.**
- Oceanographic data can be easily accessed via the 'xtractomatic' tool developed by NOAA, found here (*coastwatch.pfel.noaa.gov/xtracto) t*his now has Matlab and R implementations. (SS, ML)
- **Variable interactions in linear and additive models should only be used when there is ecological justification for combining them, i.e. when two variables characterise an oceanographic feature. (SS)**
- By categorising explanatory oceanographic variables best explaining a distribution as static or dynamic, it may also be possible to characterise species distributions as static or dynamic – such categorisation could be useful in MPA designation and suggested management (Louzao et al, 2011) (AH, ML).

- Different approaches (variables) may be needed for identifying dynamic vs static sites. Larger sites may be needed to bound dynamic features within spatially appropriate buffers. Habitat modelling may also help to in designing mobile protected areas, such as those of Turtle Watch (Hawaii) and Polovina et al 2004 (turtles distribution modelled around sea surface height and chlorophyll).
- It is likely that oceanographic fronts (including those that are beneath the surface and so cannot be identified using remotely sensed data) explain much of marine predators distributions, we would therefore benefit from working with oceanographers in identifying these or with using oceanographic models as explanatory variables (such as CARS and ROMS). (HS, LT)
- **Some environmental variables should be considered retrospectively. Louzao et al, 2009 Found that high chlorophyll-a were affecting the movements of birds up to 5 months later, this is likely due to the lag time as the energy moves up the food chain. (ML, JX, VP)**

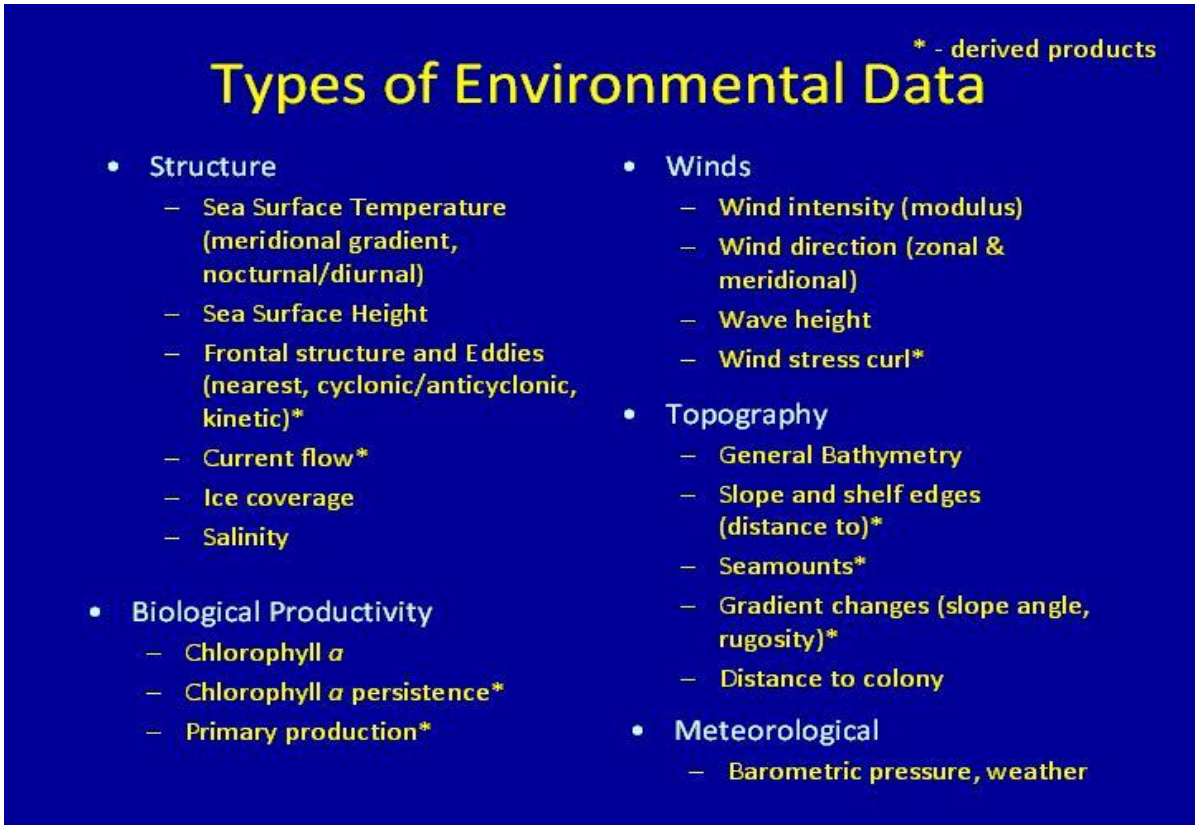

Fig: list of environmental variables discussed at the workshop that may help explain seabird distributions and act as explanatory variables in habitat models (adapted from SS presentation).

Data Exploration

- Ecologists often violate statistical assumptions when applying models to data, this results in type 1 and type 2 errors which can result in ecologically inappropriate answers.
- **Zuur et al (2009) present a framework for data exploration prior to model fitting, this framework is recommended. (VP, PT)**
- **Correlation between explanatory variables should be investigated prior to model building. Spearman rank correlation coefficient should be calculated between each pair of variables and**

where there is a strong correlation (thresholds may vary, but 0.6 is frequently used) variables should be considered highly correlated and only one should be used in model building. (VP)

- One option for deciding which correlated variable to drop is to fit two models, each with one of the correlated variables, and use the variable whose model has the lowest AIC value. (CP, SO)
- **Collinearity is not an issues in all model types (ENFA, MaxEnt, RandomForest and BRTs can incorporate highly correlated data without violating model assumptions, for example). The capabilities of the model should be understood before dropping variables. (LT)**
- **Test for homogeneity of variance and independence of response variables, these will influence the model type. (VP)**

Model Building

Using more conventional data types, such as vessel based line transects, it is possible to run a number of different models on the same dataset and compare their utility and predictive power directly.

Steffen Oppel – Comparing modelling techniques. (Oppel et al, in press)

- GLM, GAM, Boosted Regression Trees, Random Forests, Maxent and an ensemble model of all of the above were fit to the same dataset of Balearic Shearwater observations from vessel based line surveys using customised R code (also alternatively available in R package 'BIOMOD', except for Maxent which has not been implemented as of November 2011).
- To avoid issues of zero inflated data a hierarchical model was built, first modelling presence/absence, then modelling density where animal presence was recorded.
- Models were ranked based on their AUC values (75% of data was used to build the model, 25% was used in this test).
- Models with similar AUC values actually resulted in drastically different spatial distributions. It is therefore important to investigate prediction surfaces as well as statistical values when evaluating models.
- Ensemble models were determined by weighting each of the component models based on their AUC score.
- Density predictions were very poor in independent test data

Model fitting

- Following data preparation and exploration models can be fitted using a number of software packages, these are well documented in the relevant literature but a short, non-exhaustive summary is given here.
    - GAMs and GLMs fit through 'stats', 'gam', 'mgcv', packages in R
    - GAMMs and GLMMs fit through 'mgcv', 'lmer', 'MASS' packages in R
    - ENFA may be fit through 'Biomapper' standalone software or in the R packages 'adehabitat' and 'ade4'.
    - Maxent may be fit through either the 'Maxent' standalone software or the 'maxent' or 'dismo' packages in R
    - BRT may be fit through 'gbm' package in R (plus 'ROCR package'), further code available in Elith et al, 2008, suppl material; or via the TreeNet software (Salford Systems)
    - RandomForest can be fit with the 'randomForest' or 'party' packages in R, or via the RandomForest software (Salford Systems)

- Biomapper is a GIS kit - and statistical tools designed to build habitat suitability (HS) models and maps. It is centered on the Ecological Niche Factor Analysis
- R package 'BIOMOD' (which also exists as a free standalone software) is a computer platform for ensemble forecasting of species distributions, enabling the treatment of a range of methodological uncertainties in models and the examination of species-environment relationships.

When fitting models the following additional considerations are recommended.

- When fitting models to autocorrelated data, adding an autoregressive term (as per Peron et al, 2010; Zuur et al, 2009) which may be calculated using R packages 'lmer' and 'armaFit' function ('fSeries' package), significantly improves the model fit. (CP, ML)
- **When homogeneity of variance and independence of response variables are violated, mixed effect models can be used to overcome those violations. (VP)**
- **When multiple trips from individuals are used, the individual's identity should be used as a random effect. (VP)**
- **Random effects cannot be added to algorithmic models such as boosted regression trees or RandomForest, however, the data can be split along individuals for model evaluation, thus achieving a similar outcome as linear random effects models (Buston & Elith, 2011) (SO)**

<u>Model evaluation</u>

- Cross Validation
- AUC values are frequently used for testing model fit and predictive power. However, actual predictions may be different for similar AUCs, it is therefore important to investigate calibration and prediction surfaces as well as statistical tests when comparing models (Phillips & Elith, 2010). (SO)
- Statistical comparison of predictive performance (tested using cross validation) should be used for all models, particularly those for which AIC values cannot be calculated (such as GAMMs). **(SO, ML)**

Practical Session

Following talks and discussion a practical session was set up to test some of the discussed approaches on common datasets. Three datasets were made available to the workshop, and Table 3 below, shows which approaches were tested;

- PTT data of Laysan Albatross (*Phoebastria immutabilis*) movements during the incubation period (provided by Scott Shaffer)
- GPS data of Cory's Shearwater (*Calonectris diomedea*) movements during breeding (provided by Vitor Paiva)
- Geolocator data of White-chinned Petrel (*Procellaria aequinoctialis*)distribution during non-breeding (provided by Richard Phillips and BAS)

Dataset	***Presences***	***Absences***	***Model(s)***	***Reference***	***Findings***
Laysan PTT	ARS	Non-ARS	BIOMOD	S. Oppel	From BIOMOD: Random Forest (RF) best AUC (0.8), BRT also good, GLM worst. Also ran RF in separate R package and got *same* result as BIOMOD (Validation)
Corys GPS	ARS/Time per cell?	Non-ARS	Boosted Algorithm	M. Louzao	Boosted Algorithm can't explain relative variable importance but good for describing variable relationships
Corys GPS	ARS	Non-ARS	GLMM	C. Peron	Model able to detect and remove collinearity between variables
Corys GPS	ARS	Non-ARS	ENFA	H. Skov	Approach appears to work well. Good predictive map, Variables; bathymetry, SLA and SST explain 96.4% of total variance. V. high AUC (0.95) possibly due to low spatial resolution (1 deg)
Laysan PTT	ARS	Non-ARS	BRT	L. Torres	Appeared to work well, variable relationships look good. Suspiciously high AUC values.
Laysan PTT	Points	Random pseudo-a	BRT	L. Torres	To be run…
WC Petrel GLS	Points	Random pseudo-a	BRT	L.Torres	To be run…

Table 3: Modelling approaches trialled during the workshop practical session

To pragmatically calculate response variable the workshop practical session modelled foraging vs non-foraging (based on velocity) for GPS data, ARS vs non-ARS from PTT data, and presences vs pseudo-absences for GLS data (following Louzao et al, 2011; Zydelis et al, 2011)

Steffen Oppel – BIOMOD

- Movements of Laysan Albatross from French Frigate Shoals during incubation were modelled
- ARS vs non-ARS was used as a response variable to model the probability of a bird foraging in an area.
- All available explanatory variables were input in untransformed form.
- No factors were included.
- The data was split to use 36% as a testing dataset and 64% for model training.
- Initially the split was done based on individuals (i.e. some individuals were used to train the model, others to test it). However, this resulted in very poor predictive performance of the model

(which may indicate one individuals behaviour cannot be used to estimate another's). Therefore splitting randomly within individuals' data was preferred.

- BIOMOD fit:
  GLM with quadratic functions for all variables and using a stepwise AIC fit.
  GAM with a 'smooth spline' of 3
  Random Forests
  GBM boosted regression trees, with a maximum of 3000 trees.
  Artificial Neural Networks with two fold cross-validation
  Multivariate adaptive regression splines (MARS)
- BIOMOD proved pragmatic and took only 11 minutes to run 5 cross-validations of all 6 models.
- The results indicated that boosted regression trees and random forests provided the best predictions.
- All techniques found distance from land to be an influential variable.
- The two best models used all variables to some degree.
- GAM, GLM, MARS and ANN removed some variables (such as Wind) as not significant.
- The 6 different models provide very different predictions of ARS behaviour.
- By creating an ensemble model, these differences can be smoothed.

- BIOMOD proved pragmatic and useful for testing the variability between model predictions. Boosted Regression Trees and Random Forests performed best out of all models.

Henrik Skov – ENFA

- Movements of Cory's Shearwater from Selvagems Island were modelled.
- ARS were modelled to identify probability of foraging.
- ARS locations were counted within 1 degree grid cells so that there was a common scale across presences and all environmental data.
- 6 environmental explanatory variables were extracted for an area around NW Africa.
- Because ENFA is insensitive to collinearity, all data layers were used.
- All explanatory variables were Box-Cox transformed to normalise.
- The model was fit using the BIOMAPPER software.
- Bathymetry and primary productivity were shown to be the most significant explanatory variables.
- The model scored an AUC value of 0.95 showing it was a good fit.

ENFA was easily applied and provided useful results. It now needs to be tested on different datasets.

Leigh Torres – BRT

- Movements of Laysan Albatross from French Frigate Shoals during incubation were modelled
- All available explanatory variables were input in untransformed form (SST. CHLA, SLA, VGPM, WIND, WAVE, BATHY, SLOPE, DIST_LAND)
- No factors were included.

The data was split to use 36% as a testing dataset and 64% for model training

- Initially, ARS vs non-ARS was used as a response variable to model the probability of a bird foraging in an area. However, this approach did not work well because of poor classification of points based on ARS. This approach may work well with BRT, but was not tested adequately due to time limitations.
- Pseudo-absences were generated within a minimum convex polygon of tracking data. Twice as many presence points (n=2355) were generated as pseudo-absence points (n=4710).
- The BRT model was run using a variable learning rate of 0.0025.
- The BRT model produced a % deviance of 0.506, AUC of 0.927, and a correlation coefficient of 0.758.
- The contributions of each predictor to the model were:

	var	rel.inf
1	BATHY	19.334446
2	DIST_LAND	19.173227
3	WAVE	14.756038
4	VGPM	12.934347
5	SST	11.120692
6	WIND	6.758337
7	SLA	6.344606
8	SLOPE	4.853876
9	CHLA	4.724432

- The model was validated with the withheld test data and produced an AUC of 0.99.

This AUC does seem suspiciously high, so perhaps some overfitting of the model occurred. However, the prediction plots, functional plots, and model results all appear valid. Moreover, validation data was independent. More testing is needed of the BRT method applied to tracking data, but clearly there is some potential here.

Conclusions

The workshop succeeded in introducing to the group a wide range of habitat modelling techniques that have been used to model and explain seabird distributions. Extended discussions were held on developing data standards regarding input of tracking data and environmental variables to habitat modelling approaches and these are captured in this report.

Due to time constraints in the practical session it was not possible to finalise all models and compare their fit and predictions during the course of the workshop. However all of the examples undertaken during the practical session did manage to fit at least one model, which shows that all tested approaches were reasonably easy to apply and therefore were pragmatic. It should be noted that due to the differing needs of each modelling approach, a range of model inputs were tested which means that result are not completely comparable.

The workshop also concluded on approaches to develop a variety of model types (Figure 2). These approaches provide a framework for future work, and further avenues of investigation which BirdLife (and willing workshop attendees) will take forward and use for a comparison of models.

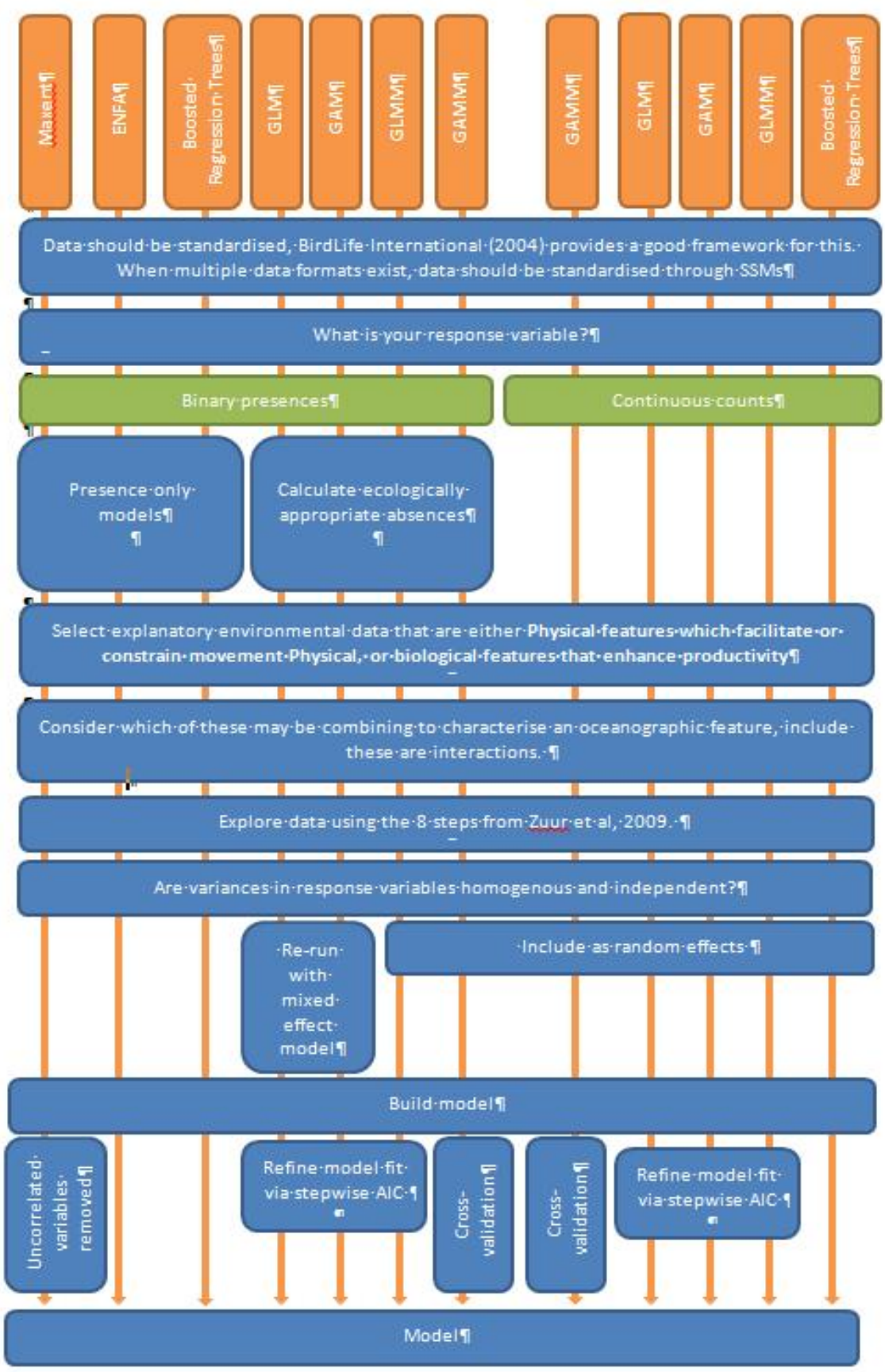

Figure 2: proposed future framework for developing a variety of model types

	Name	Email	Institution
RP	Richard Phillips	raphil@bas.ac.uk	British Antarctic Survey (BAS)
HW	Henri Weimerskirch	henriw@cebc.cnrs.fr	Centre national de la recherche scientifique CNRS, Chize
SS	Scott Shaffer	scott.shaffer@sjsu.edu	San Jose State University/Tagging of Pacific Predators (TOPP)
AH	April Hedd	ahedd@mun.ca	Memorial University of Newfoundland
MLC	Matthieu Le Corre	lecorre@univ-reunion.fr	University of Reunion/ECOMAR
BL	Ben Lascelles	ben.lascelles@birdlife.org	BirdLife International - Cambridge
PT	Phil Taylor	phil.taylor@birdlife.org	BirdLife International - Cambridge
MM	Mark Miller	mark.miller@birdlife.org	BirdLife International - Cambridge
IR	Iván Ramírez	ivan.ramirez@birdlife.org	BirdLife International - Brussels
ML	Maite Louzao	maite.louzao@gmail.com	Spanish Institute of Oceanography
ALH	Autumn-Lynn Harrison	harrison@biology.ucsc.edu	Univ. of California Santa Cruz/Tagging of Pacific Predators (TOPP)
AM	Ana Meirinho	ana.meirinho@spea.pt	Sociedade portuguesa para o estudo das aves (SPEA)
VP	Vitor Paiva	vitorpaiva@gmail.com	Institute of Marine Research (IMAR/CMA), University of Coimbra
JR	Jaime Ramos	jramos@ci.uc.pt	Institute of Marine Research (IMAR/CMA), University of Coimbra
JX	José Xavier	jxavier@zoo.uc.pt	Institute of Marine Research (IMAR/CMA), University of Coimbra
RZ	Ramunas Zydelis	rzy@dhi.dk	DHI
LT	Leigh Torres	l.torres@niwa.co.nz	National Institute of Water and Atmospheric Research (NIWA)
SO	Steffen Oppel	Steffen.Oppel@rspb.org.uk	Royal Society for Protection of Birds (RSPB)
HS	Henrik Skov	hsk@dhigroup.com	DHI Water and Environment
CP	Clara Peron	Clara.PERON@cefe.cnrs.fr	Centre national de la recherche scientifique CNRS, Chize
KD	Karine Delord	delord@cebc.cnrs.fr	Centre national de la recherche scientifique CNRS, Chize
CS	Carlos Silva	carlos.silva@spea.pt	Sociedade portuguesa para o estudo das aves (SPEA)
NO	Nuno Oliveira	nuno.olivera@spea.pt	Sociedade portuguesa para o estudo das aves (SPEA)
MB	Mark Bolton	Mark.Bolton@rspb.org.uk	Royal Society for Protection of Birds (RSPB)
AC	Albert Cama	acama@seo.org	SEO/BirdLife

References

Buston, P.M., Elith, J., 2011. Determinants of reproductive success in dominant pairs of clownfish: a boosted regression tree analysis. Journal of Animal Ecology 80, 528-538.)
Stokland, J.N., Halvorsen, R., Støa, B., 2011. Species distribution modelling - Effect of design and sample size of pseudo-absence observations. Ecological Modelling 222, 1800-1809.

Lobo, J.M., Jiménez-Valverde, A., Hortal, J., 2010. The uncertain nature of absences and their importance in species distribution modelling. Ecography 33, 103-114.

Jimenez-Valverde, A., Lobo, J.M., Hortal, J., 2009. The effect of prevalence and its interaction with sample size on the reliability of species distribution models. Community Ecology 10, 196-205.

Lobo, J.M., Tognelli, M.F., 2011. Exploring the effects of quantity and location of pseudo-absences and sampling biases on the performance of distribution models with limited point occurrence data. Journal for Nature Conservation 19, 1-7.

Phillips, S.J., Elith, J., 2010. POC plots: calibrating species distribution models with presence-only data. Ecology 91, 2476-2484.

Shaffer, S.A., Weimerskirch, H., Scott, D., Pinaud, D., Thompson, D.R., Sagar, P.M., Moller, H., Taylor, G.A., Foley, D.G., Tremblay, Y., and Costa, D.P. (2009) Spatio-temporal habitat use by breeding sooty shearwaters (*Puffinus griseus*). Marine Ecology Progress Series 391, 209-220.

Kappes, M.A., Shaffer, S.A., Tremblay, Y., Foley, D.G., Palacios, D.M., Robinson, P.W., Bograd, S.J., and Costa, D.P. (2010) Hawaiian albatrosses track interannual variability of marine habitats in the North Pacific. Progress in Oceanography 86, 246-260.

Le Bohec C., Durant J.M., Gauthier-Clerc M., Stenseth N., Park Y.H., Pradel R., Grémillet D., Gendner J.-P., & Le Maho Y. (2008) King Penguin population threatened by the Southern Ocean Warming, Proceedings of National Academy of Sciences USA, 105, pp. 2493-2497.

Zydelis, R. et al. (2011). Dynamic habitat models: Using telemetry data to project fisheries bycatch. Proc. R. Soc. B 278, 3191_3200

Skov H., Humphreys E., Garthe S., Geitner K., Gremillet D., Hamer K.C., Hennicke J., Parner H., Wanless S. (2008). Application of habitat suitability modelling to tracking data of marine animals as a means of analysing their feeding habitats. ecological modelling 212 (2008) 504–512

Wakefield E., Phillips R.A, Trathan P.N., Arata J., Gales R., Huin N., Robertson G., Waugh S.M., Weimerskirch H., and Matthiopoulos J. (2011). Habitat preference, accessibility, and competition limit the global distribution of breeding Black-browed Albatrosses. Ecological Monographs, 81(1) 141–167

Aarts G., MacKenzie M., McConnell B., Fedak M. and Matthiopoulos J.(2008). Estimating space-use and habitat preference from wildlife telemetry data. Ecography 31: 140_160, 2008

Freitas C., Kovacs K.M., Lydersen C., Ims R.A. (2008). A novel method for quantifying habitat selection and predicting habitat use. *Journal of Applied Ecology,* 45*,* 1213–1220

BirdLife International (2004). Tracking ocean wanderers: the global distribution of albatrosses and petrels. Results from the

Global Procellariiform Tracking Workshop, 1–5 September, 2003, Gordon's Bay, South Africa. Cambridge, UK: BirdLife International.

McConnell, B. J. et al. 1992. Satellite tracking of gray seals (Halichoerus grypus). _ J. Zool. 226: 271_282.Louzao et al, 2009

Zuur A.F., Leno E.N., Elphick C.S. (2009). A protocol for data exploration to avoid common statistical problems. Methods in Ecology and Evolution Volume 1, Issue 1, pages 3–14

Turtle Watch (Hawaii) - http://seaturtlestatus.org/sites/swot/files/061810_SWOT4_p36_TurtleWatch.pdf

Polovina JJ, Balazs GH, Howell EA, Parker DM, Seki MP, Dutton PH (2004) Forage and migration habitat of loggerhead (*Caretta caretta*) and olive ridley (*Lepidochelys olivacea*) sea turtles in the central North Pacific Ocean. Fish Oceanography 13:36–51

Oppel, S., Meirinho, A., Ramirez, I., Gardner, B., O'Connell, A., Miller, P., Louzao, M. *In press.* Comparison of five modelling techniques to predict the spatial distribution and abundance of seabirds. Biological Conservation.

Péron, C., Charbonnier Y., Louzao, M., Delord, K., Phillips, R.A., Weimerskirch, H. 2010. Annual variations in oceanographic preferences of White-chinned Petrels (*Procellaria aequinoctialis*) breeding at Kerguelen Islands. Marine Ecology Progress Series 416: 267–284.

Louzao, M., Pinaud, D., Péron, C., Delord, K., Wiegand, T., Weimerskirch, H. 2011. Conserving pelagic habitats: seascape modelling of an oceanic top predator. Journal of Applied Ecology 48: 121–132.

Louzao. M., Bécares, J., Rodríguez, B., Hyrenbach, K.D., Ruiz, A., Arcos, J.M. 2009. Combining vessel-based surveys and tracking data to identify key marine areas for seabirds. Marine Ecology Progress Series 391: 183–197.