SCAR Retrospective Analysis of Antarctic Tracking Data.
A Multi-disciplinary Workshop held be the Expert Group of Birds and Marine Mammals

Belgian Science Policy Office (BELSPO)
Avenue Louise 231 Louizalaan
B-1050 Brussels
http://www.belspo.be/belspo/organisation/contact_en.stm

18 -22 May 2016

Attendees: Mark Hindell (Chair), Yan Ropert-Coudert, Phil Trathan, Ben Raymond, Bruno Danis, Horst Bornemann, Luis Huckstadt, Ian Jonsen, Anton van de Putte, Nabil Youdjou, Antonio Agüera

SCAR EG-BAMM and EG-ABI sponsored a five-day workshop on the Retrospective Analysis of Antarctic Tracking Data (RAATD). The overarching goals of the RAATD project are to undertake a multi-species assessment of habitat use of Antarctic top predators in the Southern Ocean based on existing animal tracking data to identify *Areas of Ecological Significance* (AES), which are regions that are important for foraging to a range of predators and which have high diversity and abundance of lower trophic levels. The project will provide (i) a greater understanding of fundamental ecosystem processes in the Southern Ocean (ii) facilitate future projections of predator distributions under varying climate regimes and (iii) provide input into spatial management planning decisions for management authorities such as CCAMLR. The synopsis of multi-predator tracking data will also expose potential gaps of data coverage in regions or seasons that are important but under-represented, either as a result of a low regional research presence or a low ecological significance. This will provide an important input for directing future studies

The meeting was hosted by the Belgian Science Policy Office in Brussels. The participants included database specialists and spatial ecologists and the overall objectives of the meeting were:

1. To review the analytical approaches to identify multi-species, Areas of Ecological Significance from tracking data in the Southern Ocean
2. To decide on the analytical approach, or approaches to be used going forward.
3. To identify and prioritise the outputs from RAATD
4. To identify who will have carriage of those outputs
5. To review the available datasets, and to identify any data gaps
6. Discuss options for the housing of the legacy data set (those data that providers are happy to contribute to an open database for use by the Antarctic science community).
7. To undertake preliminary data quality control, filtering/smoothing and habitat models

The meeting opened with short presentations from Phil Trathan on the outcomes of the recent Scotia Sea penguin tracking distribution workshop in Cambridge, as well from Ian Jonsen who talked about the Tagging of Pacific Predators (TOPP) program. Ben Raymond, Mark Hindell and Luis Huckstadt presented descriptions of distribution and habitat modelling approaches that they have used recently.

The meeting initially considered the outputs for the analyses, as having some idea of these was thought to be important when deciding on analytical approaches. Several papers were identified as end –products. These included (but not exclusively) (i) an overview for a high impact journal, (ii) a detailed description of the AES and their environmental determinants (iii) quantifying changes in species and AES distributions using past and future scenarios of climate and fishing (iv) Quantifying the community compositions of AES. It was also recognised that there could be many species specific papers which would take advantage of the compilation of the global data sets, but these would not be a central responsibility of RAATD. The first overview paper could specifically examine overlaps of AES and human activities in the form of fishing, tourist and science activities (primarily shipping). The

workshop felt that the first paper should be completed by the end of 2016. (ACTION ITEM: Phil T. to investigate obtaining tourist ship track data. M. Hindell to investigate fishing effort data from CCAMLR).

Another important output from RAATD is the legacy dataset (*i.e.* the final compiled tracking data from all the data contributors) that will be made publically available for future analyses. The meeting participants stressed that only those datasets that were specifically released for this purpose by their data providers should be included. There was interest in housing the database in biodiversity.aq as it could then contribute *inter alia* to the online version of Atlas. This is particularly important as the original Atlas did not contain tracking data and therefore greatly under-represents a number of bird and mammal species. The previous RAATD workshop held in Strasburg in December 2013 established Darwin Core fields that will make the data suitable for several other large Ocean databases. The data could also be published as a data paper in a journal such as a Nature Scientific Data as a companion paper to the overview article.

The meeting then considered several high level questions that need to be resolved prior to running detailed analyses. These were:

1. What overall approach should be adopted? The two alternatives were to either firstly identify multi species ESA and then model them (as per Hindell et al 2011) or alternatively model each species individually and then combine them to identify ESA's (as per Raymond et al 2015). The consensus was that there would be too few data with sufficient overlap to identify enough ESAs for the first approach. Therefore we would develop distribution models for the individual species, use them to estimate distribution across the entire spatial domain, and then combine these predicted distributions to identify the ESAs. It was acknowledged that this approach was contingent on the species specific models being sufficiently accurate to produce reliable estimates, and that the uncertainty associated with the estimates needed to be quantified.

2. How will the data be aggregated? In particular the workshop discussed whether there should be global *vs.* region modelling domains, as it was recognised that for some regions, such as the West Antarctic Peninsula (WAP), may have contrasting factors influencing species distributions. It was resolved to start with a global assessment and then to investigate whether regional assessments were necessary to improve model performance and predictive capacity.

   The workshop also considered how to deal with inter-annual variability within the data set. It was acknowledged that while factors that influence species distribution will vary among years, there too few data to explicitly model this, and so year effects would not be included in the models.

Regarding intra-annual aggregations the workshop decided to develop models for both the summer period (when most species are breeding and therefore have very different distributions) and the winter period (when most species are less constrained by breeding requirements. The summer period will be 6 months from October to March and the winter period April to August.

The workshop also considered whether to model important intrinsic biological aspects within a species such as age and sex. As there is relatively little difference among genders with respect to foraging range it was felt that the sexes could be combined for the analyses, with the exception of elephant seals where males use very different geographic regions to females. The workshop considered there to be few data to separate animals by age.

3. How do we deal with data quality control? The first issue was how to best and most efficiently filter the data to deal with anomalous locations. The three candidate approaches were (i) The R package CRAWL which uses a rapid Kalman filter, (ii) the BSAM state space approach and (iii) the newly developed SGAT approach. The workshop recognised that each approach had its own strengths and weaknesses and that we needed to explore these options further before setting on a final approach (or approaches). Kalman filters could be quickly run on the data during the workshop, which would provide a preliminary data set that could be used for mapping and early modelling. SGAT could also be tried on a subset of the data during the workshop for comparison with the Kalman filter results BSAM is relatively slow and could not be run within the timeframe of the workshop, but Ian Jonsen would run a subset of the data so that each of the three approaches could be compared and assessed. (ACTION ITEM: Mark H., Ian J. and Ben R. to investigate filtering options and decide on the most appropriate).

   The workshop also discussed if we needed to adopt a consistent filtering approach across data types. The data contain locations derived from System Argos, light-level geolocation (GLS) and from GPS, each of which have different levels of accuracy and different levels of uncertainty. It was felt that while Argos and some GLS tracks needed filtering using one of the approaches outlined above, GPS tracks could be dealt with by using a simple speed filter. Those GLS tracks which had been processed using SGAT would not require additional filters as this was already done during the location estimation.

4. What modelling approaches should be adopted? The workshop recognised two broad groups of distribution models (i) “Selectivity” approaches which contrast areas where animals did go (i.e. their tracks) with areas that they did not go, even though these were equally accessible to them. There are a range of methods for deriving the unused locations (e.g. randomly sampling throughout the spatial domain, or generating sets of random tracks). (ii) “Usage” approaches which use data along a track to identify areas where the animal spent time (Area Restricted Search) or passed through quickly

(Transit). The ARS and transit areas can be identified dichotomously using State Space Models, or quantified as a continuous measure based on the time that animals spend in a grid cell. Again the different approaches have their own strengths and weaknesses, and there was no clear preference of any one approach. The workshop agreed to adopt a model ensemble approach and that the final publications would present the results of multiple models and examine where they agreed and where they differed. At least one selectivity model (Ben R. pseudo tracks) and one usage model (the time spent in a cell) are to be used but others such as MAXENT and SSM may also be tried. (ACTION ITEM: Mark H., Ian J. and Ben R. to run a variety of model types).

The workshop then split into two working groups, one devoted to compiling the datasets provided by V. Andrews-Goff into standardised templates, conducting initial quality control, improving the metadata files as well as identifying additional data sets that might be added (the Data Acquisition Group). The other group was devoted to initial data analysis including running filters, identifying trips, generating distribution maps and developing preliminary models (the Modelling Group).

Data Compilation and Acquisition Group.
The group established a common template file to be used for each species, largely inspired from the Darwin Core Template defined in the 2013 Strasbourg meeting. The template has columns for all 21 variables that will be informative in the future analyses, even if some are not known now and need to be populated a later date after consultation with data providers. The templates were also sufficiently generic to be used for all species and for all data types (Argos, GLS and GPS). The columns names are descriptive and detectable by systems using Darwin Core fields. It was decide to produce one file for each species, thereby combining data from all device types.

The final variable names were:
vernacular_name, scientific_name, individual_id, sex, breeding_stage, deployment_site, deployment_decimal_latitude, deployment_decimal_longitude, device_type, device_id, year, month, day, time, time_zone, decimal_latitude, decimal_longitude, location_quality, processing_method_gls, latitude_uncertainty_metres, longitude_uncertainty_metres

The process of bringing data into the template files identified a number of issues with some datasets, (*e.g.* missing data, incorrect data, *etc*). In most cases these issues were resolved by contacting V. Andrews-Goff and getting the original datasets. Once compiled these template files were added to a Dropbox folder so they could be accessed and used by the Data Analysis group. Only Yan and Anton can add files to this folder in order to maximise control over the data that are posted there. The compilation of these data sets was a major undertaking and occupied most of the group's time during the meeting.

Several additional datasets were identified that were important to be included in the data set. These were:

- Antarctic fur seals: data from Crozet and Kerguelen are not included. Winter data from the Antarctic Peninsula, South Georgia and Marion Is also not included, but are waiting for data holder permission?
- Crabeater seals : Complete
- Humpback whales: Could request data for South Orkney from Alex Zerbini
- SES: Complete
- Weddell Seals: 8 deployments on Ross Sea from 2014 can be added
- Adelie penguin: Ross Sea data have now been sourced and are in process. Permission from Silvia Olmastroni is being sought. David Ainley has given permission but there are troubles accessing the data. There are no data from the WAP yet, but there are several that might be included from Jefferson Hinke, Bill Fraser and others. (ACTION ITEM: Phil T to ask after these data at the CCAMLR MPA workshop in Buenos Aires)
- Black-browed albatross: Ask Javier Arata for his Diego Ramirez data.
- Sooty albatross *spp*: They were not in the original list of species for RAATD, but they do make an important contribution.
- Emperor penguins: There may be more data from the Ross Sea collected by Gerald Kooyman. (Action ITEM: Luis to ask Gerry for these data)
- Grey headed albatross: Complete. Marion Island data were added during the workshop.
- King penguins: Barbara Weinecke and Klemens Putz were added during the workshop. Data from Crozet were also added. Kerguelen and South Georgia data could also be added.
- Macaroni penguins: Kerguelen and BAS data could be included. Phil T noted that these are currently being used for a post-doc project and may not be available. However, Marion Island data were added during the workshop, doubling the total amount of data for this species.
- Wanderer albatross: complete after adding Marion Island data during the workshop.
- White-chinned petrels or any other type of petrels with good coverage would be good to add. Yan RC talked with Peter Ryan to know if they are happy with making the petrels data available, but it seems that these data will be analysed prior to be made available for RAATD.
- Short-tailed and sooty shearwaters could be added
- Snow petrel data from East Antarctic could also be added, although the samples are relatively small – less than 50

The workshop decided that one person (Yan RC) should be in charge of maintaining the final files. Virginia Andrews-Goff can prepare new and updated files for each species received, send them to Yan RC who will include them into the final files that will be stored in the Dropbox folder Anton vdP created. Only Anton vdP and Yan RC will be able to edit them.

Table 1. The number of individual tracks for each species in each of the three data types (GPS, PTT (Argos) and GLS) after completion of all the available data. Only individuals with more than 20 locations are included here. (NB this is still provisional as there are outstanding issues of quality control that need to be resolved.)

Species	GPS	PTT	GLS	Total
Adelie penguin	166	137	0	303
Antarctic fur seal	0	354	0	354
Black browed albatross	81	246	2	329
Crabeater seal	0	103	0	103
Dark-mantled sooty albatross	23	0	0	23
Emperor penguin	0	139	0	139
Grey headed albatross	102	8	1	111
Humpback whale	0	46	0	46
King penguin	9	63	0	72
Light mantled sooty albatross	10	19	3	32
Macaroni penguin	75	82	0	157
Southern elephant seal	0	276	0	276
Wandering albatross	111	139	3	253
Weddell seal	0	160	0	160
Total	577	1772	9	2358

Data Analysis Group.

This group had several tasks.

1. Develop R scripts to do bulk runs of the CRAWL Kalman filter. This was straightforward for Argos data, but needed to be modified for GLS data, which have a different location error structure that CRAWL cannot handle. Ian J. developed a script (available on the Dropbox) which incorporated GLS errors, where appropriate. The script also ran an initial speed filter to remove the worst locations, and also dropped tracks that had fewer than 20 observed locations. A major challenge was that CRAWL could not recognise when an animal circumnavigated the globe, so this needed to be accounted for in the new script. All of the Argos and PTT tracks were run through this new script and generally produced sensible tracks. There were nonetheless still issues with data gaps (due to duty cycling) and circumpolar travel that resulted in spurious tracks. While unresolved at present the outputs were sufficient to provide preliminary distribution maps (see appendix 2)
2. Develop scripts to detect and number separate trips with a track. Different approaches are required for Argos and GPS data. Mark H. and Phil T. developed R scripts for Argos and GPS data respectively. One problem is that in some species individuals may not return to the deployment site at the end of a trip. This means that while the scripts work in most cases there is still a need to visually check each track. This is an area that will require more attention before final datasets can be made available for modelling.

3. Compare CRAWL and SGAT filter performance. Ben R. investigated the use of SGAT to filter both Argos and already processed GLS tracks. SGAT worked for Argos tracks, but could not simply be modified to deal with pre-processed GLS tracks. This will be further explored after the meeting. Also, Ian J. is working on a new implementation of the *BSAM* location filtering model that should be as fast as CRAWL. The new version will be run in either ADMB or TMB via R (to be determined).
4. Generate maps of the filtered data. Preliminary distribution maps for each species were made by Ian J. using the Kalman filtered data. "Heat" maps illustrating the mean time spent in $50km^2$ grid cells were also generated for some species by Mark H., but these were complicated by the circumnavigation of the globe by several of the albatross species and will require additional work after the workshop.
5. Conduct primary distribution model for one species. Ben R. was able to run a demonstration of his pseudo-track modelling approach on the filtered crabeater seal data.

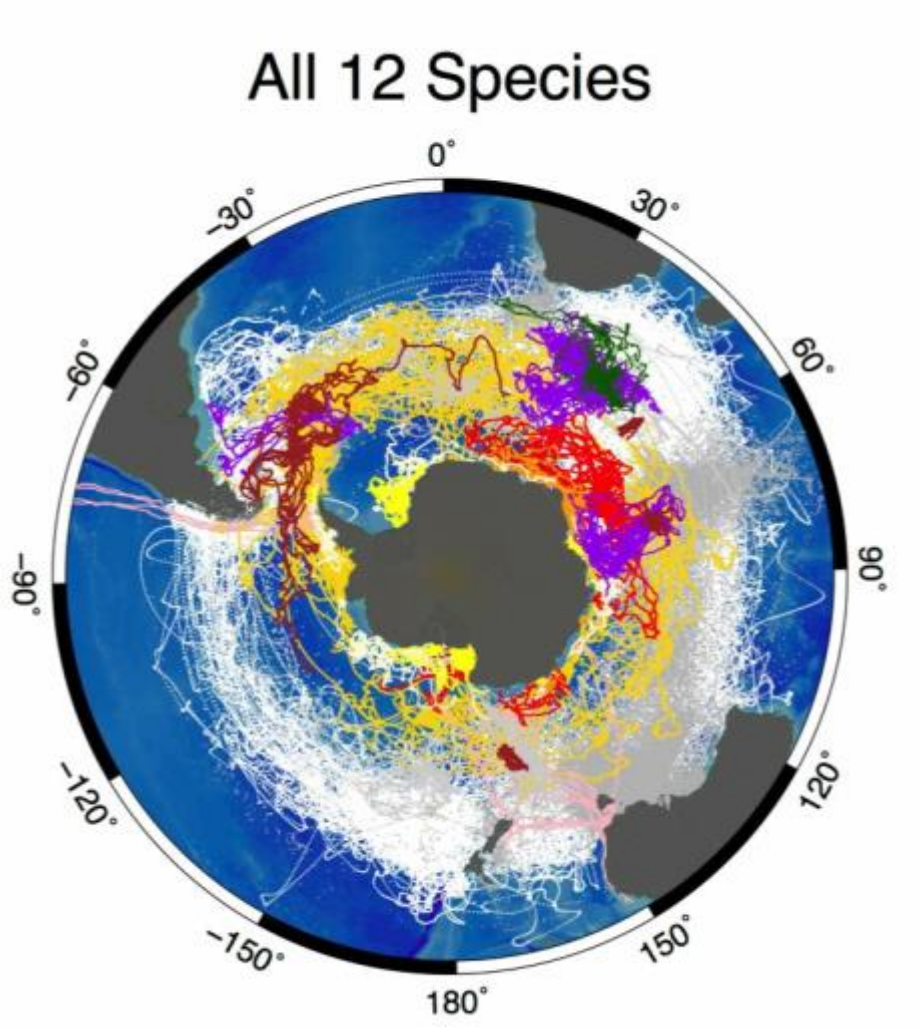
Conclusion.

The workshop was very successful and achieved all of its stated objectives. The RAATD project now has a very large, standardised tracking data set from 14 species which can be used for preliminary modelling and characterisation of AES. We have also identified additional data sets that could be incorporated if the data holders are willing. Initial QC has begun and some problems identified and overcome, but this still remains a big task for the project in the near future. Initial filters have now been run all the data which, while not yet finalised, provide data that can be used to develop models. Trip identification scripts have been developed which can automatically detect and number trips within a track for most individuals.

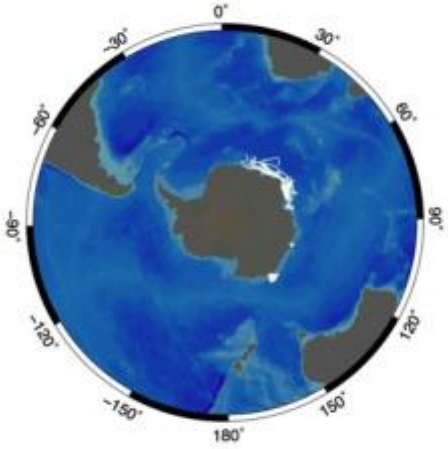
The workshop agreed that 12 months will be required to deal with the outstanding tasks of (i) final filtering of the data, (ii) automated trip allocation, (iii) developing the distribution model ensemble and integration of the these models to indicate AES. The primary carriage of this body of work will be Ben R., Ian J. Mark H. and Luis H. along with other statistical experts not present at the workshop.

The next workshop was suggested to be held at AWI, Germany in May 2016. Horst B. volunteered to seek funding. This workshop will (i) consider the results of the initial modelling efforts, (ii) consider the biological interpretation of those models, (iii) discuss the next analytical steps (e.g. what metrics to use to describe AES, such as species richness, prey consumption rates, etc.), and (iv) develop preliminary outlines for the publications.

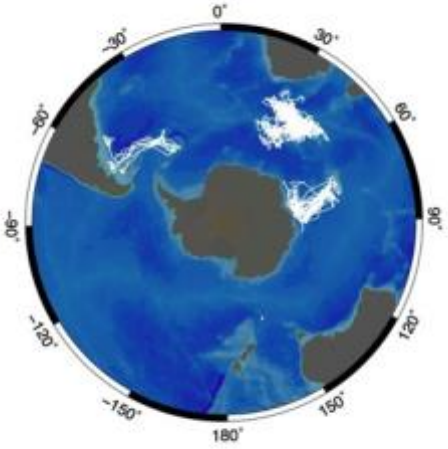
Appendix 1: Preliminary distribution maps of the filtered tracks for 12 species of Southern Ocean predators.



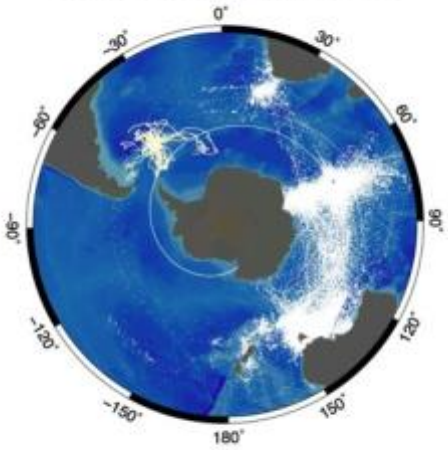
Adelie Penguin



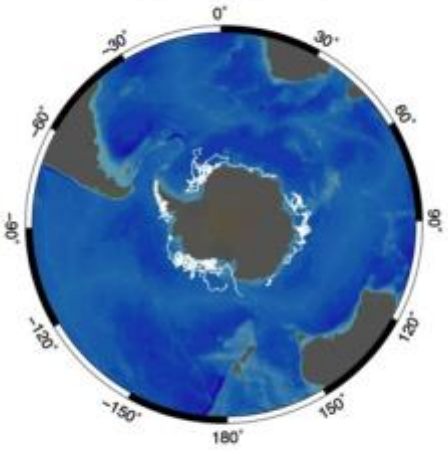
Antarctic Fur Seal



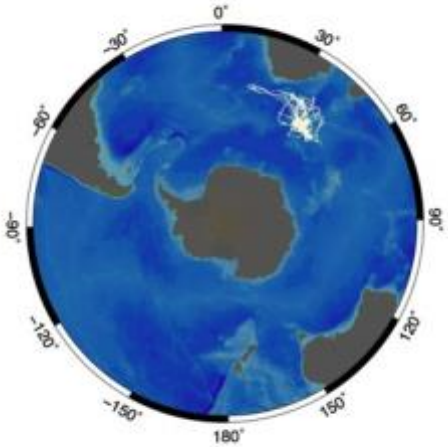
Black−browed Albatross



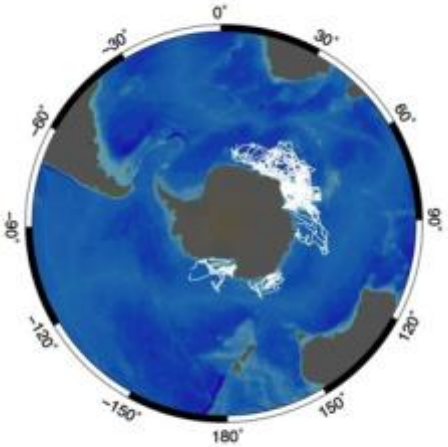
Crabeater Seal



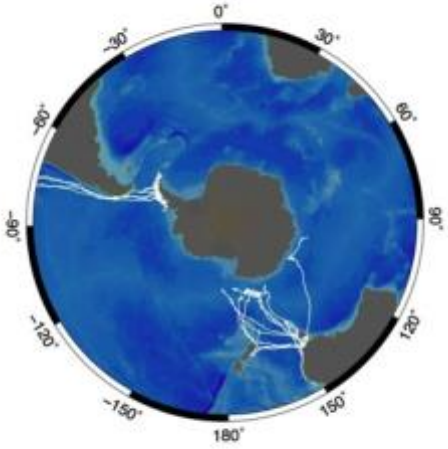
Sooty Albatross



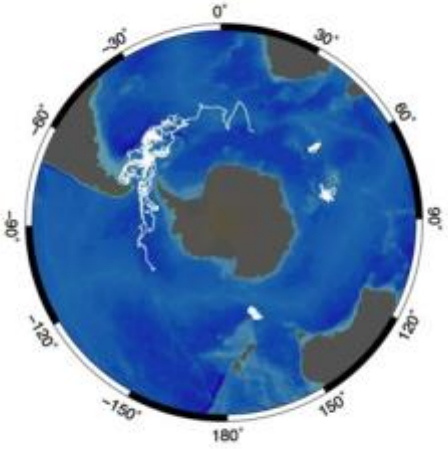
Emperor Penguin



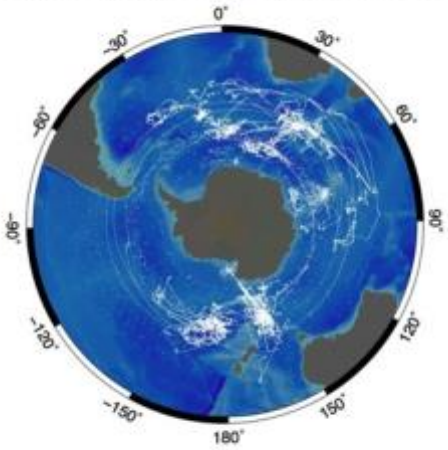
Humpback Whale



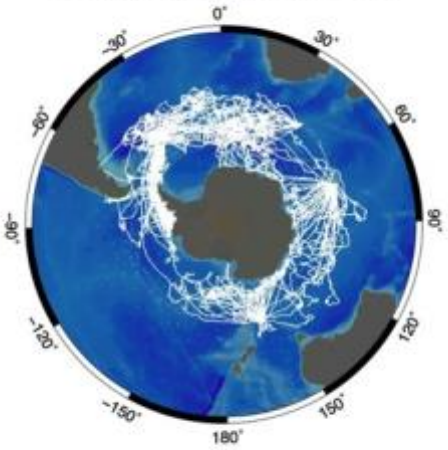
King Penguin



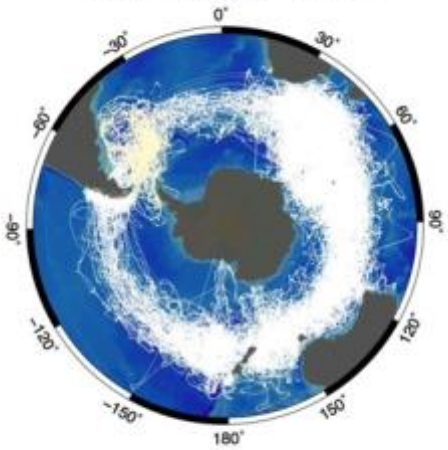
Light−mantled Sooty Albatross



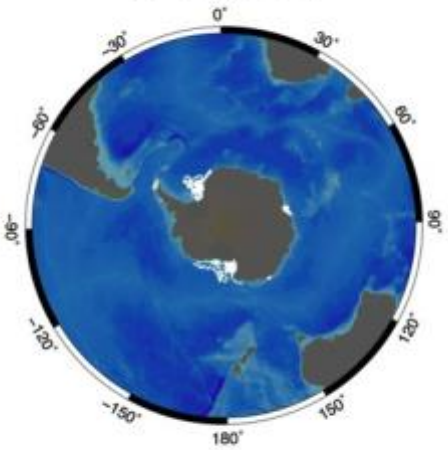
Southern Elephant Seal



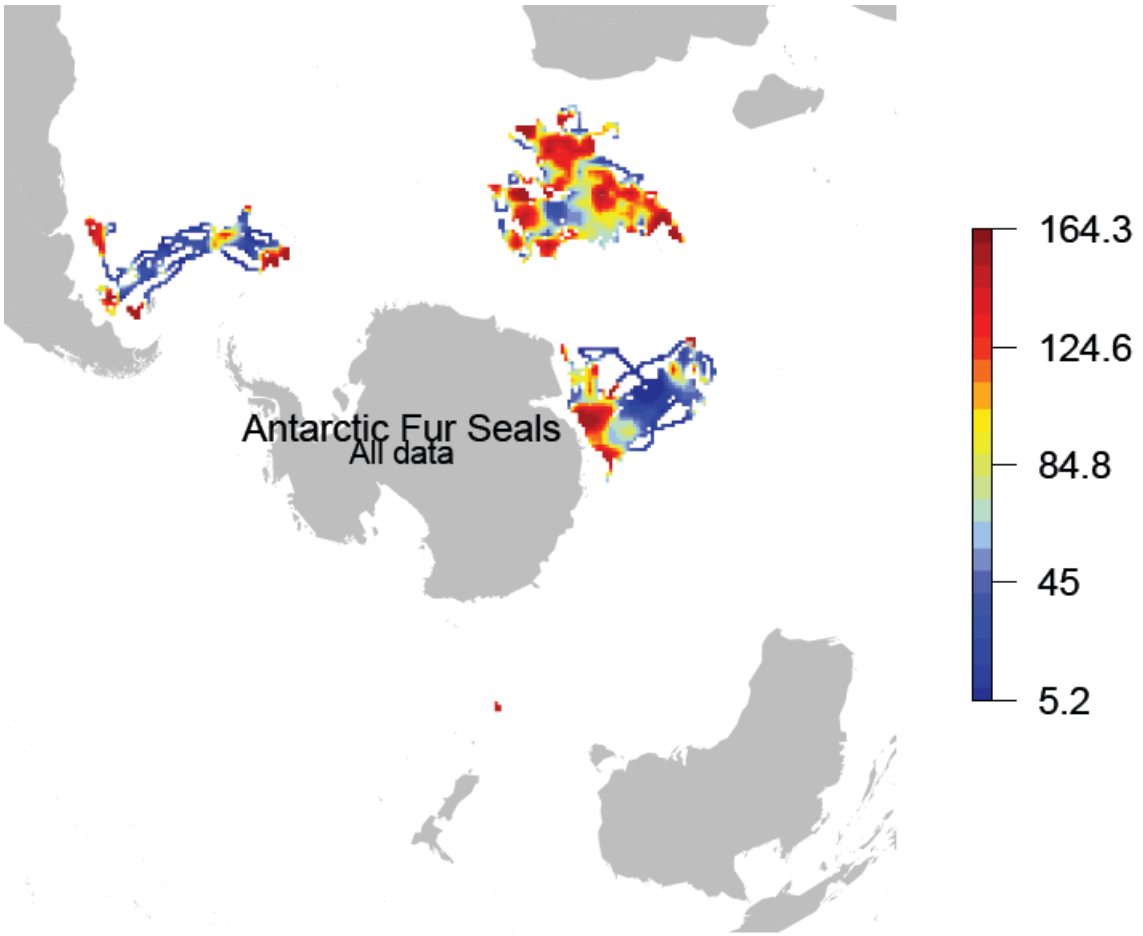
Wandering Albatross



Weddell Seal



Appendix 2: Example utilisation map for Antarctic fur seals. The maps show the mean time that individuals spend in 50km^2 grid cells.



Appendix 3: Agenda of the meeting

Agenda:

Monday May 18, 0900-1200. Presentations on a potential framework (Mark) and modelling approaches (Ben/Phil), Ian, Luis, Mark.
Monday, May 18, 1300-1700. Discussion on planned outputs
Tuesday, May 19, 0900:1200. SCAR Biogeographic Atlas, RAATD and the legacy dataset
Tuesday, May 19, 1300-1700. Discussion on technical issues (data processing, metrics, modelling approaches)
Wednesday, May 20, 0900:1200. Investigation of data and meta data
Wednesday, May 20, 1300:1700. Investigation of data and meta data
Thursday, May 21, 0900:1200. Preliminary data analysis
Thursday, May 21, 1300:1200. Preliminary data analysis
Friday, May 22, 0900:1200. Preliminary data analysis
Friday, May 22, 1300:1700. Future work planning and allocation of tasks and responsibilities