GPDD: Initial Summary

Liam Furman

Marine Resource Assessment and Management Group (MARAM)
Department of Mathematics and Applied Mathematics
University of Cape Town
Rondebosch 7701

February 2010

Background:
It has been recognised by the RMP sub-committee of the IWC that it is necessary to consider the effects of environmental variability in performing a meta-analysis of growth rates in baleen whale populations, which is to be used in determining a plausible range for MSYR in RMP trials. Furthermore, it was agreed that it is important to investigate existing studies in order to better estimate parameters describing the impacts of environment variability. Although studies on whale populations have been identified, information is still limited. Therefore, datasets from other taxa, particularly large mammals, are to be considered.

It is hoped that the GPDD (Global Population Dynamics Database) may provide the required data. This appears to be "the largest collection of animal and plant population data in the world", which has been developed in collaboration by the NERC Centre for Population Biology at Imperial College, London; the National Center for Ecological Analysis and Synthesis, Santa Barbara, and the University of Tennessee. In brief, it contains population data from a variety of sources and includes related taxonomic data, habitat data and metadata.

The first stage in this process is to examine the GPDD to determine whether it contains the required type, amount and quality of data.

Marginals:
To get useful insights into the contents of the GPDD we look at marginals across single variables which are restricted to datasets involving mammals. Note that there are 977 mammal datasets of a total of 5238 in the database overall.
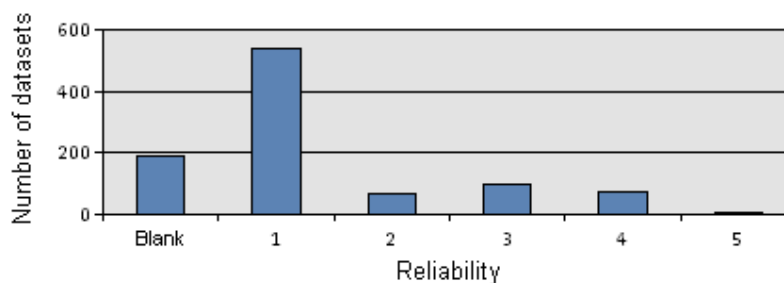

Figure 1: Distribution of the reliability of the datasets. Reliability is a subjective measure of data quality, which is based on criteria such as type of environment sampled, the species in question, the area of the sampling site, and the sampling method. It is intended as a rough guide only.
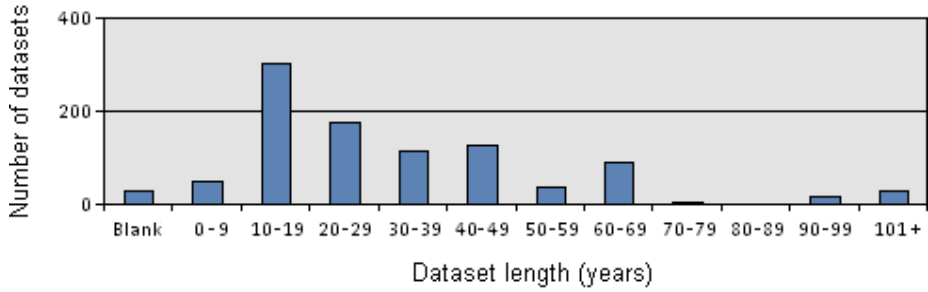
Figure 2: Distribution of the approximate lengths of the datasets, in years. The length of a dataset is calculated by dividing the number of data points, excluding any missing values, by the number of samples per year. For example, a kangaroo rat dataset has fifty-two data points with seven missing values and a sampling frequency of four samples per year. The length of the dataset will be calculated as approximately eleven years. The eleven rather than the fifty-two is included in the figure.
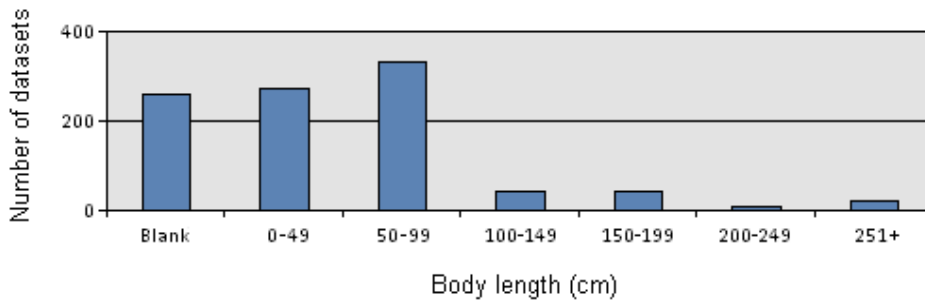

Figure 3: Distribution of the average body length of the target species of the datasets, in cm.
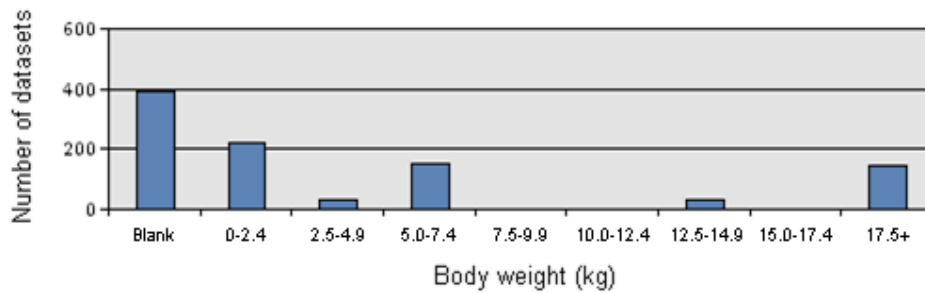

Figure 4: Distribution of the average body weight of the target species of the datasets, in kg.
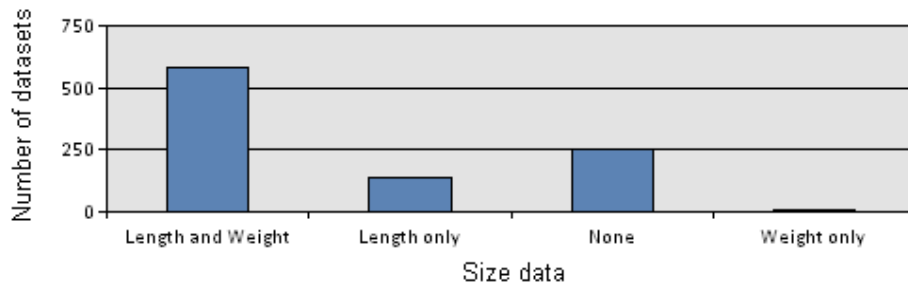

Figure 5: Distribution of size data availability for the target species of the datasets.
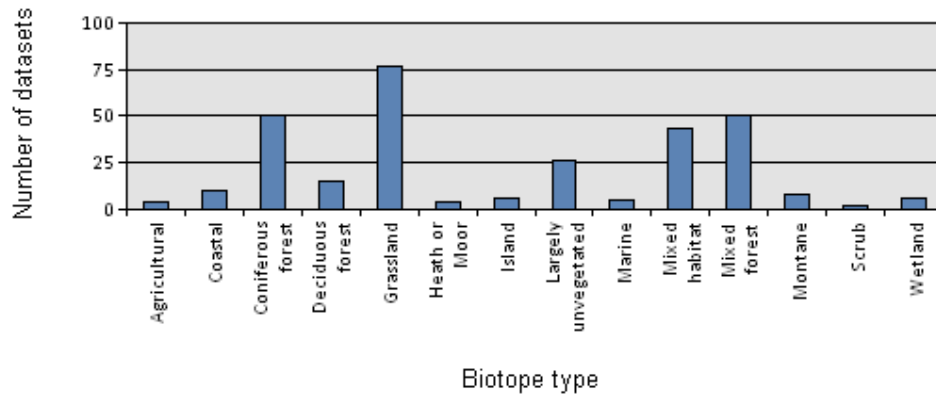
Figure 6: Distribution of biotopes of target species of the datasets. The *Unspecified* biotope category has been excluded for clarity. It contains 671 datasets.
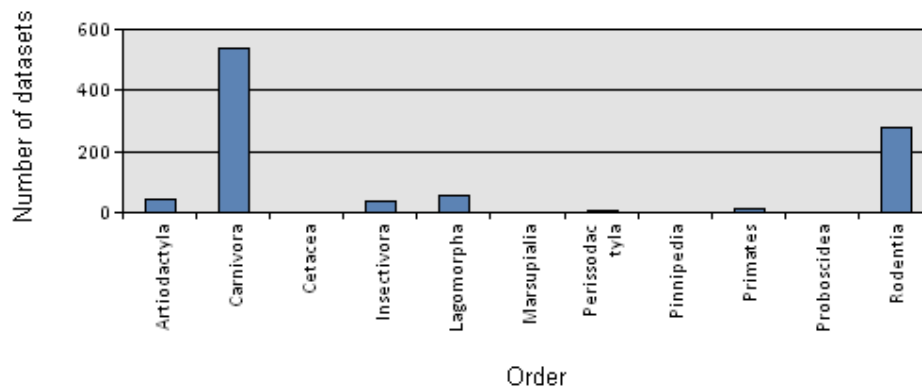

Figure 7: Distribution of taxonomic orders of target populations of the datasets.

Defining potentially useful datasets:

The marginals give us an overview of the GPDD that can be used to define constraints that isolate useful datasets. There are two conflicting requirements. One the one hand, we require high quality datasets for populations that are sufficiently biologically similar to baleen whales. On the other hand, a sufficient number of datasets are required to reduce variability in parameter estimates. Probably somewhere in the region of fifty usable datasets might be appropriate.

Fig. 1 suggests that datasets with a reliability value of three and greater should be accepted. Only eighty datasets have a reliability value of four or greater, so setting the restriction to this level would be unlikely to produce fifty usable datasets once further constraints are imposed. A dataset with a reliability value of two is probably too unreliable to be useful.

To isolate datasets for populations of large mammals, we appeal to average body length and body weight data. Fig. 5 indicates that by using size data, one quarter (253) of the mammalian datasets are immediately excluded as they have no such data. This suggests that it may be useful to find an alternative method of isolating biologically similar mammals. If size data are used, Fig. 5 shows that it is worthwhile to place restrictions on body length, but not to place additional alternative restrictions on body weight, because 717 datasets have length data and additionally using weight data would only include seven extra datasets. Fig 3. indicates that datasets for populations with an average body

length of greater than fifty centimetres should be accepted. Although it is debatable whether these should be regarded as large mammals, it is the only possibility if we hope to produce fifty usable datasets.

Dataset length, biotope and taxonomic order are to be unrestricted for now. Although dataset length is obviously important in determining the usefulness of a dataset, restrictions will only be decided once the marginals with the reliability and body length constraints in place have been considered. Fig. 6 shows that biotope is too evenly distributed to be readily constrained and it may not even be biologically useful if we are simply interested in large mammals. Taxonomic order, although currently unrestrained, could be a good method for isolating useful datasets. Not only is it possibly more reliable than body length in selecting populations that are biologically similar to whales, but all datasets have taxonomic classification data and therefore none are excluded by default, as is the case when using body length data.

In summary, for the moment we restrict further investigation to datasets for populations with an average body length of fifty centimetres or greater that have a reliability value of three or greater.

Results:
The restrictions result in only thirty-three useful datasets. This is below the target number of datasets and it is worth bearing in mind that restrictions still need to be placed on dataset length. It appears that the only way the GPDD could be useful in its current state is to isolate usable datasets on some basis other than average population body size. Recently, the GPDD has been revised and it is possible that new datasets as well as missing size data has been added, although the revised version has not yet been received. Marginals using only the useful datasets have been included. Table 1 lists the twenty-two useful datasets that contain at least eleven years of population data.
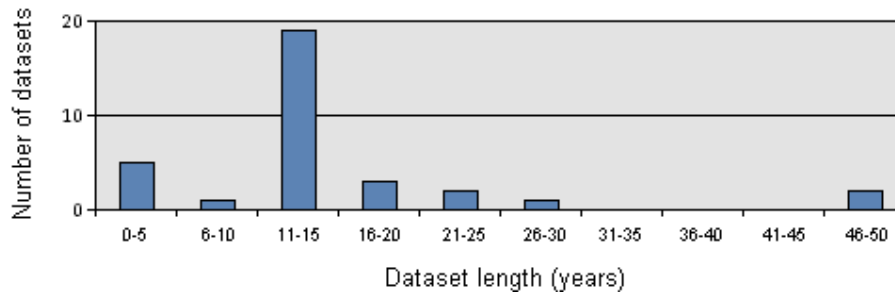

Figure 8: Distribution of the approximate lengths of the useful datasets, in years. The length of a dataset is calculated by dividing the number of data points (excluding any missing values) by the number of samples per year.
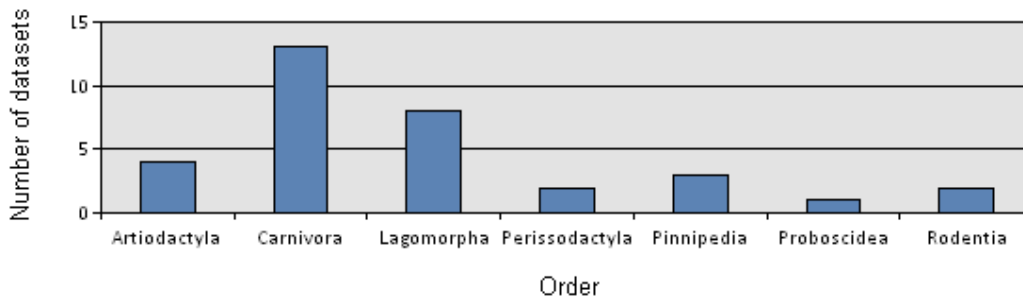

Figure 9: Distribution of taxonomic orders of target populations of the useful datasets.

4

Species	Dataset length (years)	Reliability	Body Length (cm)	Data type	Sampling units	Location
Grey wolf	48	4	50*	Estimate	Individuals	Bialowieza, Poland
Grey wolf	48	4	50*	Estimate	Individuals	Bialowieza, Belarus
Grey seal	26	3	274	Sample	Individuals	Gulfs of Finland and Riga, Finland
Mountain hare	24	3	64	Sample	Individuals killed	Glenrinnes Estate, Scotland
Lion	20	5	2603*	Count - resident females > 3yrs	Individuals	Serengeti, Tanzania
Snowshoe hare	18	3	50	Sample	Hare pellets	Pocahontas County, USA
Mountain hare	17	3	64	Sample	Individuals killed	Dufftown, Scotland
Elephant seal	15	4	650	Sample	Individuals	Macquarie Island, Antarctica
Grey wolf	14	4	50*	Estimate	Individuals	Bialowieza, Poland
Mountain hare	14	4	64	Sample	Index of individuals per 100 trap sessions	Dufftown, Scotland
Snowshoe hare	13	3	50	Density	Individuals	Cloquet, USA
Snowshoe hare	13	3	50	Sample	Individuals	Oba, Canada
European red fox	13	4	67	Density	Individuals	Lapland, Finland
European red fox	13	3	67	Sample	Tracks	Lapland, Finland
European red fox	13	3	67	Sample	Tracks	Lapland, Finland
European red fox	13	3	67	Sample	Tracks	Lapland, Finland
Brown bear	12	4	180	Count	Adults	Yellowstone National Park, USA
European red fox	12	3	67	Sample	Tracks	Lapland, Finland
Snowshoe hare	11	3	50	Density	Individuals	Lake Alexander, USA
Grey seal	11	3	274	Sample	Calves	Farne Islands, England
Greater kudu	11	4	215	Density	Individuals	Kruger National Park, South Africa
Greater kudu	11	4	215	Density	Individuals	Kruger National Park, South Africa

Table 1: A list of mammalian datasets that have: a reliability score of at least three, at least eleven years of data, and populations with an average body length of at least fifty centimeters. Entries are sorted in descending order of dataset length. Values with stars (*) are outside their expected range and seem to be errors in the GPDD.