**Estimating body density of cetaceans: the manual**

Tomoko Narazaki ([naratomoz@gmail.com](mailto:naratomoz@gmail.com); IGOR Pro, R)

Lucia Martina Martin Lopez ([luciamartinaml@gmail.com](mailto:luciamartinaml@gmail.com); Matlab)

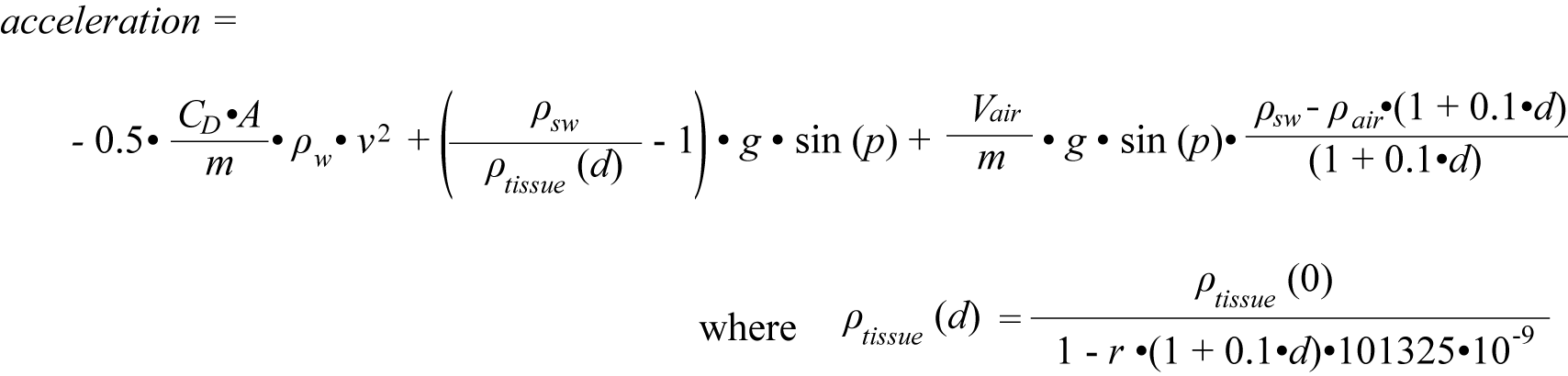
Saana Isojunno ([si66@st-andrews.ac.uk](mailto:si66@st-andrews.ac.uk); R, jags)

12 June 2016

This document describes the sequence of operations needed to estimate body density of whales from tag data using a hydrodynamic glide model. For this analysis, you need either IGOR Pro (WaveMetrics, https://www.wavemetrics.com) or Matlab (MathWorks, http://www.mathworks.com/products/matlab/) for tag data processing and R Studio (https://www.rstudio.com) for Bayesian estimation.

***Part 2: Estimating unknown parameters using R***

Based on hydrodynamic glide model, acceleration during sub-glides can be expressed as below.



*CD*: drag coefficient *ρsw*: density of the surrounding seawater (kg m-3)

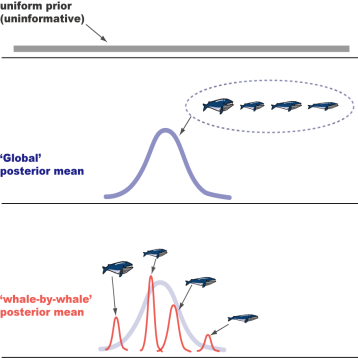
*A*: relevant surface area (m2) *ρtissue*: density of the non-gas component of the whale body (kg m-3)

*m*: mass of the whale (kg) *Vair*: volume of air inspired at the surface (ml)

*g*: gravity *r*: compressibility for animal tissue

*p*: animal pitch (radians) *ρair*: density of air (kg m-3)

*d*: glide depth (m)

Here, four unknown parameters in the equations (drag term CDA/m, body density ρtissue (0), diving gas volume Vair/m, tissue compressibility r) are estimated using Bayesian Gibbs sampling in freely available software JAGS within R.

In this analysis, a total of 12 models were evaluated to explore variability in body density, the drag term and diving gas volume. We use hierarchical model structure in which each individual (body density, drag term) or dive-by-dive (diving gas volume) estimate is considered to be a sample from a global or ‘population’ distribution with an estimated global mean, with an estimated variance across the dives and/or individuals. After running all the models, best model is selected based on the deviance information criterion (DIC).

10) Install JAGS and R2jags

Both JAGS and R2jags are required in this analysis. JAGS (Just Another Gibbs Sampler) is a tool for analysis of Bayesian hierarchical models using Markov Chain Monte Carlo (MCMC) simulation. See <http://mcmc-jags.sourceforge.net> for the details. R2jags is an R package used to call JAGS from R (See <https://cran.r-project.org/web/packages/R2jags/R2jags.pdf> for the details).

* JAGS
  + Download files at <https://sourceforge.net/projects/mcmc-jags/files/JAGS/>.
  + Follow the instructions to install.
* R2jags
  + Open R (or R-studio)
  + Type **install.packages("R2jags")** and hit enter. To check if it is properly installed, type **library(R2jags).**
* rjags
  + You might need to install rjags to install R2jags. In this case, type **install.packages(“rjags”)** to install.
* Lattice
  + This r package is required for the trace history plots (Step 16)

11) Read glide summary tables in R.

* Open **make\_all\_whales\_data.R**.
* Set working directory
* **dataDir** 🡪 where the glide summary tables are stored
* **modelDir** 🡪where the model files etc. are stored
* **whales** 🡪 List of whaleID. Should be the same as the names in your glide summary tables (*whaleID*.csv)Check the variable names specified in the R script match with the variables in your .csv file
* Run the codes to make **all\_whales.Rd** that includes glide data from all whales in a single data frame.

12) Set priors.

To estimate four unknown parameters using Bayesian framework, we need to set a specific prior distribution for each parameter. Here, we use a uniform (non-informative) prior for compressibility, body density and diving gas volume. The uniform prior specifies a maximum plausible range of values where the true parameter should be found, but does not inform the model what the most likely value is. For example, the prior ranges used for Northern bottlenose whales (*Hyperoodon ampullatus*) were:

**compr ~ dunif (0.3, 0.7) # x 10-9 Pa-1**

**Vair ~ dunif (5, 50) # ml kg-3**

**body.density.g ~ dunif(800, 1200) # kg m-3**

For the combined drag term (CDA/m), we specified an informative prior based upon a calculated draft coefficient (Miller et al., 2004). . In the case of Northern bottlenose whales, for example, body length ranges from 5.8 to 9.8 m, surface area (mean 23.0 m2, range 12-36 m2) and mass (mean: 6816 kg; range: 3027-12739 kg). Drag coefficient was estimated to be roughly 0.0030 based on previous research of some cetaceans. Using these values, the prior is set to be a normal distribution with mean around the calculated CDA/m of 10.0x10-6 m2 kg-1.The variance of the prior describes uncertainty about our prior knowledge about the mean, and in this case we calculated a standard deviation of 2.0x10-6 m2 kg-1 from the likely range of values for both the drag coefficient and the whale’s surface area/mass.

**Table:** CD values reported in previous research

|  |  |
| --- | --- |
| **Species** | **CD** |
| Killer whale | 0.0029 (Fish 1998) |
| Fin whale | 0.0026 (Bose and Lien, 1989) |
| Sperm whale | 0.0031 (Miller et al. 2004) |

**CdAM.g ~ dnorm(10, 0.25)T(5,20) # x10-6 m2 kg-1, dnorm(mean, precision)**

**# so, if SD = 2, precision = 1/22**

Similar to the uniform priors, here the normal prior is truncated to a maximum range of plausible values (5-20 x10-6 m2 kg-1).

Once you have decided what the prior distributions should be, you can specify them in the models by following steps:

* Open text file named **model(*number*).txt**.
* For each text file, change the prior appropriately and save.
  + For uniform priors, specify parameter ~ dunif(a,b) where a and b is the lower and upper bounds of the distribution
  + For normal priors, specify parameter ~ dnorm(mu,tau), where mu is the mean and tau is the precision parameter (1/variance)
  + To truncate a continuous prior (e.g. Normal, Gamma), use T(a,b) where a and b are constants that define a range where the true parameter value is known to lie a-priori. To truncate from left or right alone, you can use T(a,) or T(b,) respectively.
  + For other prior specifications, see “Distribution aliases” in the latest jags user manual http://www.stats.ox.ac.uk/~nicholls/MScMCMC15/jags\_user\_manual.pdf

13) Set initial values.

* Open **make\_inits.R**
* The sp.inits function generates a list of initial values, one for each chain. In this case, we will run three different chains with fixed initial values for each chain.
* Initial values must be within the prior range and should be set randomly to represent a wide range of values across the different chains.
* Set the initial values according to your prior range, save and close the script (estimate\_all\_models.R will call the function as long as you have placed the make\_inits.R file in your workspace)**.**

14) Estimate models.

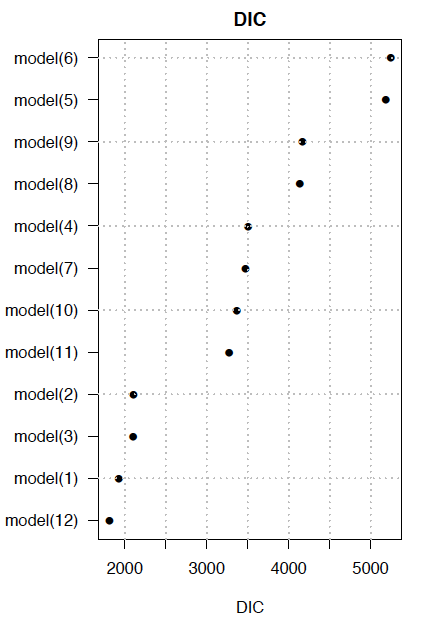
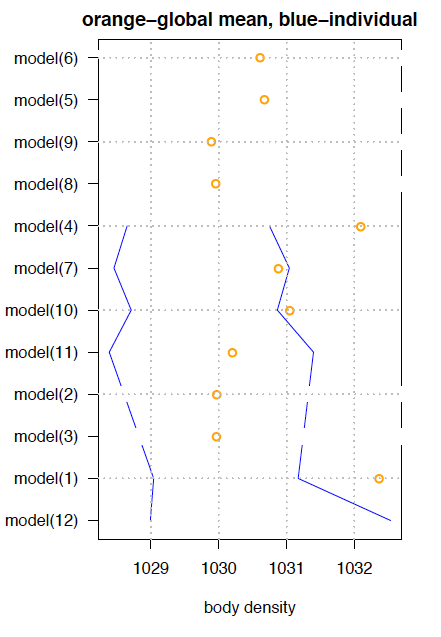
* Open **estimate\_all\_models.R**
* Make sure **setwd, dataDir, modelDir** are OK.
* Set **filterBool** 🡪 criteria for selecting ‘good’ glides for analysis.

Default is mean pitch > 30 degree, maximum dive depth > 100 m, variation of roll > 0.9, during descent and ascent only, no missing data in acceleration, swim speed, seawater density. Change the settings depending on your data.

* Set **fitName** 🡪 you can change this according to your filterBool criteria.
* The filtered data is stored in your workspace as an .Rd file (**data\_*fitName*.Rd**) which contains the filtered dataframe, a true/false vector (**filtBool**) identifying which data are included from the original data, and the list of whales and dives that were retained after filtering (**fit.whales** and **fit.dives**, respectively). The numeric names of **fit.whale** and **fit.dives** correspond to sequential identifiers for each whale (**whale.id**=1,2,3..NW) and dive (**dive.id**=1,2,3..ND) that are passed to the jags model. No missing values (e.g. 1,2,4,5) are allowed for either **whale.id** or **dive.id** as complete sequential series are required to index the hierarchical model structure correctly in jags. Therefore, it is important to keep track which whales and dives correspond to the numeric identifiers, especially when plotting model results with legends etc.
* Each model structure is summarized in a data frame **modTab**. It indicates whether dive-by-dive or individual variability is included for each term (Vair, CdAM or body density) in each model so that initial values and monitored parameters can be correctly specified in the jags function.
* Run the script to estimate models
* Because the sp.inits function will provide more initial values than some of the models will require, warning message “Unused initial value in chain” will appear. You can ignore this warning message.
* Result of each model is stored as a Rd file named **model(*number*)\_*fitName*.Rd** in the work directory.

15) Model selection

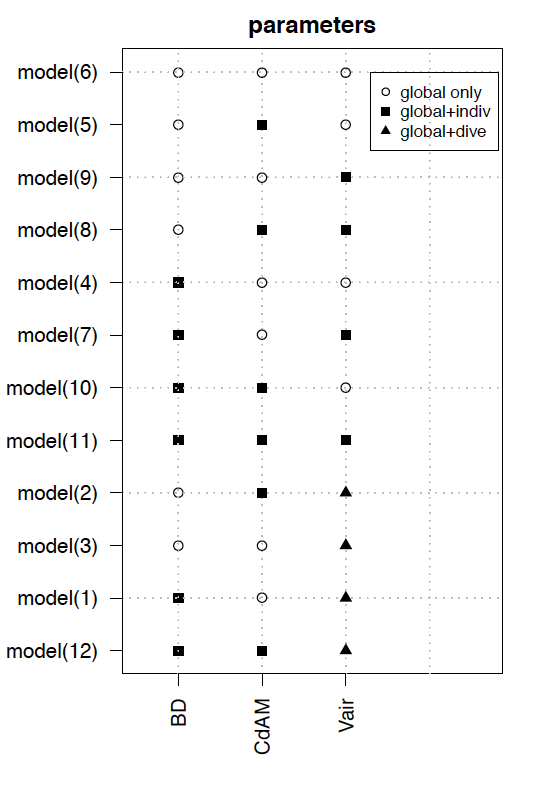
* Once you have estimated all the models, open **compare\_models\_DIC.R**
* Make sure **setwd, dataDir, modelDir** are OK.
* Make sure **fitName** is correct
* Run the code and you will get the following files in your work directory.
  + **compare\_models.pdf** 
    - Deviance, DIC etc. of each model
    - Parameter estimates (posterior means) of each model (to check how much estimates change across different models)

**○** global mean estimate

**—** whale-by-whale estimates

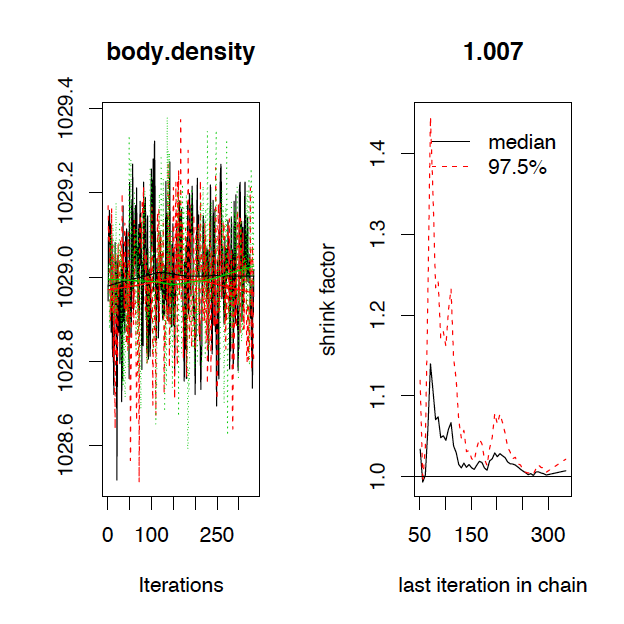
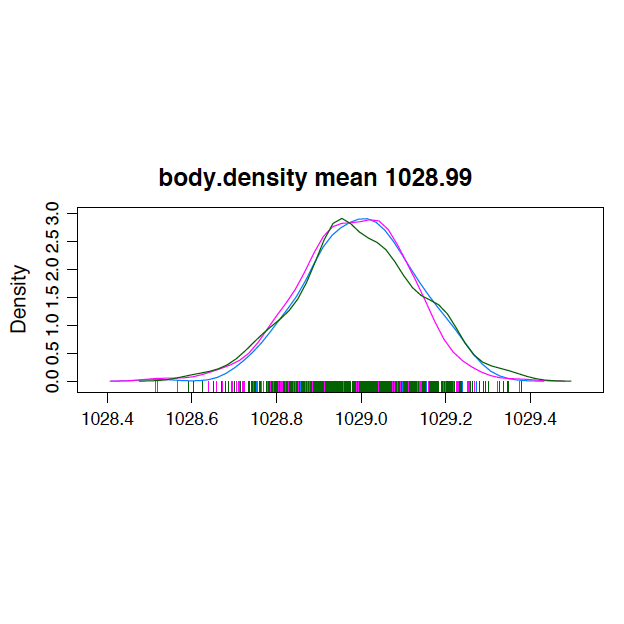
* + **all\_model\_estimates.csv**
    - a summary table of DIC, deviance and parameter estimates for each model
  + model\_parameters.pdf
    - shows combination of global, individual-specific, dive-specific parameters used in each model.



* Select the best model based on the lowest DIC.

16) Check the selected model.

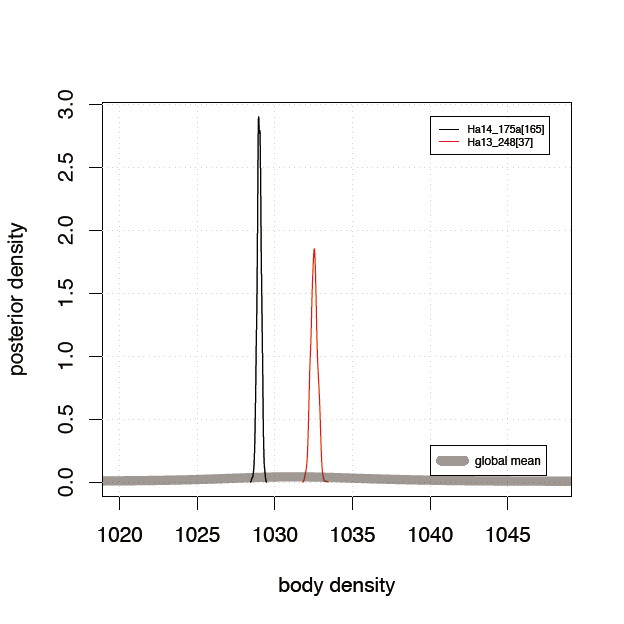
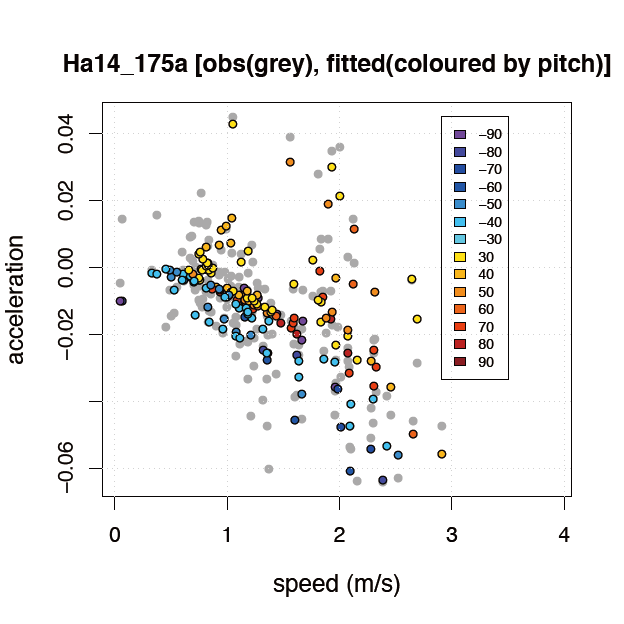
* Open **best\_model\_tracehistories.R**
* Make sure **setwd, dataDir, modelDir** are OK.
* Make sure **fitName** is correct.
* Set **modelName** to the best model (e.g. **modelName<-c(“model(12)”)**)
* Run the codes and you will get **plot\_best\_model(*number*)\_tracehistories.pdf**

* Check the trace histories and posterior density plots to assess convergence of each parameter.

17) Make result plots of the selected model.

* Open **best\_model\_plots.R**
* Make sure **setwd, dataDir, modelDir** are OK.
* Make sure **fitName** is correct.
* Set **modelName** to the best model (e.g. **modelName<-c(“model(12)”)**)
* For the plots that include priors, ensure correct priors and x-axis values are specified. These are commented (#) within the script.
* Run the script to save **plot\_best\_model(*number*).pdf** that includes the posterior density of each parameter, observed and estimated acceleration vs swim speed etc.

18) Make a summary table for the selected model.

* Open **make\_summary\_table.R**
* Make sure **setwd, dataDir, modelDir** are OK.
* Make sure **fitName** is correct.
* Set **modelName** to the best model (e.g. **modelName<-c(“model(12)”)**)
* Run the script to extract parameter estimates for body density (kg/m3), Cd\*A/m (x10-6 m2/kg), Vair/m (mL/kg3) and compressibility (x10-9 /Pa). Separate tables are given for global, individual and dive-by-dive estimates:
  + **model(*number*)\_*fitName*\_Estimates\_global.csv**

|  |  |
| --- | --- |
| mean | Global posterior mean |
| L95 | Lower 95% CRI) |
| U95 | Upper 95% CRI |
| range95 | 95% CRI range (U95-L95) |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

* + **model(*number*)\_*fitName*\_Estimates\_indiv.csv** (created only when the selected model includes individual-specific parameters)

|  |  |
| --- | --- |
| ID | Whale ID |
| numGlides | Number of sub-glides used in the model |
| BD.mean | Mean whale-by-whale body density (kg/m3) |
| BD.L95 | Lower 95% CRI whale-by-whale body density (kg/m3) |
| BD.U95 | Upper 95% CRI whale-by-whale body density (kg/m3) |
| BD.range95- | 95% CRI range of whale-by-whale body density (kg/m3) |
| CdAM.mean | Mean whale-by-whale Cd\*A/m (x10-6 m2/kg) |
| CdAM.L95 | Lower 95% CRI whale-by-whale Cd\*A/m (x10-6 m2/kg) |
| CdAM.U95 | Upper 95% CRI whale-by-whale Cd\*A/m (x10-6 m2/kg) |
| CdAM.range959 | 95% CRI range of whale-by-whale Cd\*A/m (x10-6 m2/kg) |
| Vair.meanVair | Mean whale-by-whale Vair/m (mL/kg3) |
| Vair.L95Vair | Lower 95% CRI whale-by-whale Vair/m (mL/kg3) |
| Vair.U95Vair | Upper 95% CRI whale-by-whale Vair/m (mL/kg3) |
| Vair.range95 | 95% CRI range of whale-by-whale Vair/m (mL/kg3) |

* + **model(*number*)\_*fitName*\_Estimates\_dive.pdf** (created only when the selected model includes dive-specific Vair).

|  |  |
| --- | --- |
| dive.id | Dive ID supplied to jags |
| ID | Whale ID |
| numGlides | Number of sub-glides in the dive |
| dive.number | Dive ID within each tag record before filtering |
| dive.all | Dive ID across all tag records before filtering |
| dive.max.depth | Maximum dive depth (m) |
| dive.duration | Dive duration (s) |
| Vair.mean | Mean dive-by-dive Vair/m (mL/kg3) |
| Vair.L95 | Lower 95% CRI dive-by-dive Vair/m (mL/kg3) |
| Vair.U95 | Upper 95% CRI dive-by-dive Vair/m (mL/kg3) |
| Vair.range95 | 95% CRI range of dive-by-dive Vair/m (mL/kg3) |