Supplementary material: ICES Journal of Marine Science, 70

Evidence for density-dependent changes in body condition and pregnancy rate of North Atlantic fin whales over four decades of varying environmental conditions

Rob Williams, Gisli A. Vikingsson, Astthor Gislason, Christina Lockyer, Leslie New, Len Thomas, and Philip S. Hammond

Analysis, expanded methods, and results

Choosing a body condition variable

Exploratory analyses showed that the only two measurements for which there are substantial quantities of data are m4 and v4 (and m4 is measured around 10% more than v4). So, we can ignore the others as they were not measured nearly so often. The two measures are highly correlated (Figure S1), as one would expect, so we chose m4, because sample size is larger.

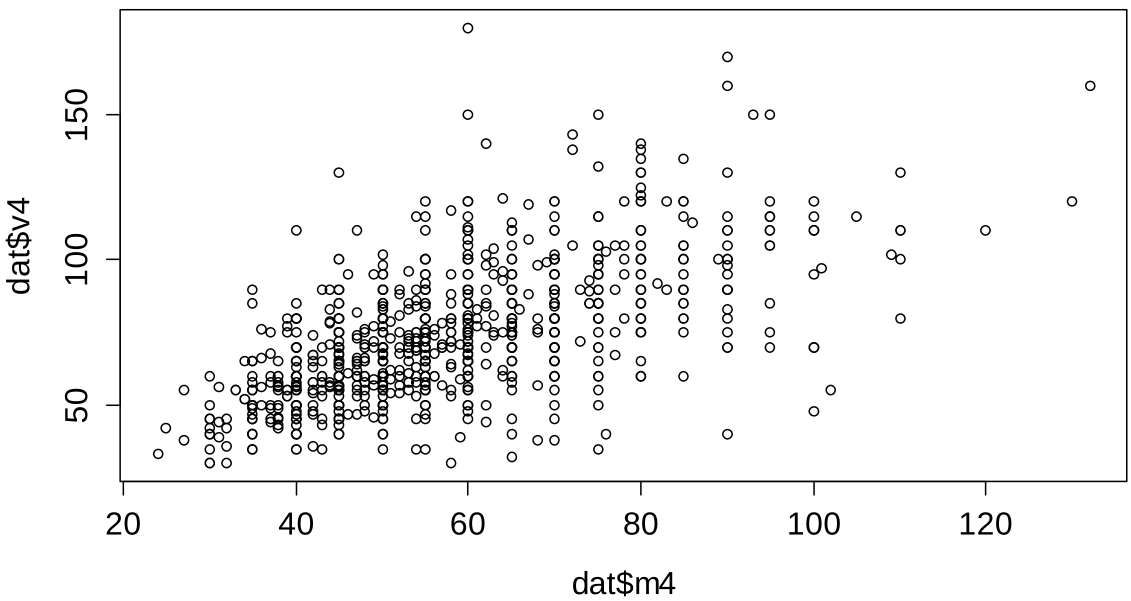

Figure S1. Correlation between v4 and m4, which has a Pearson r^2 of 0.6.

Summary of raw data

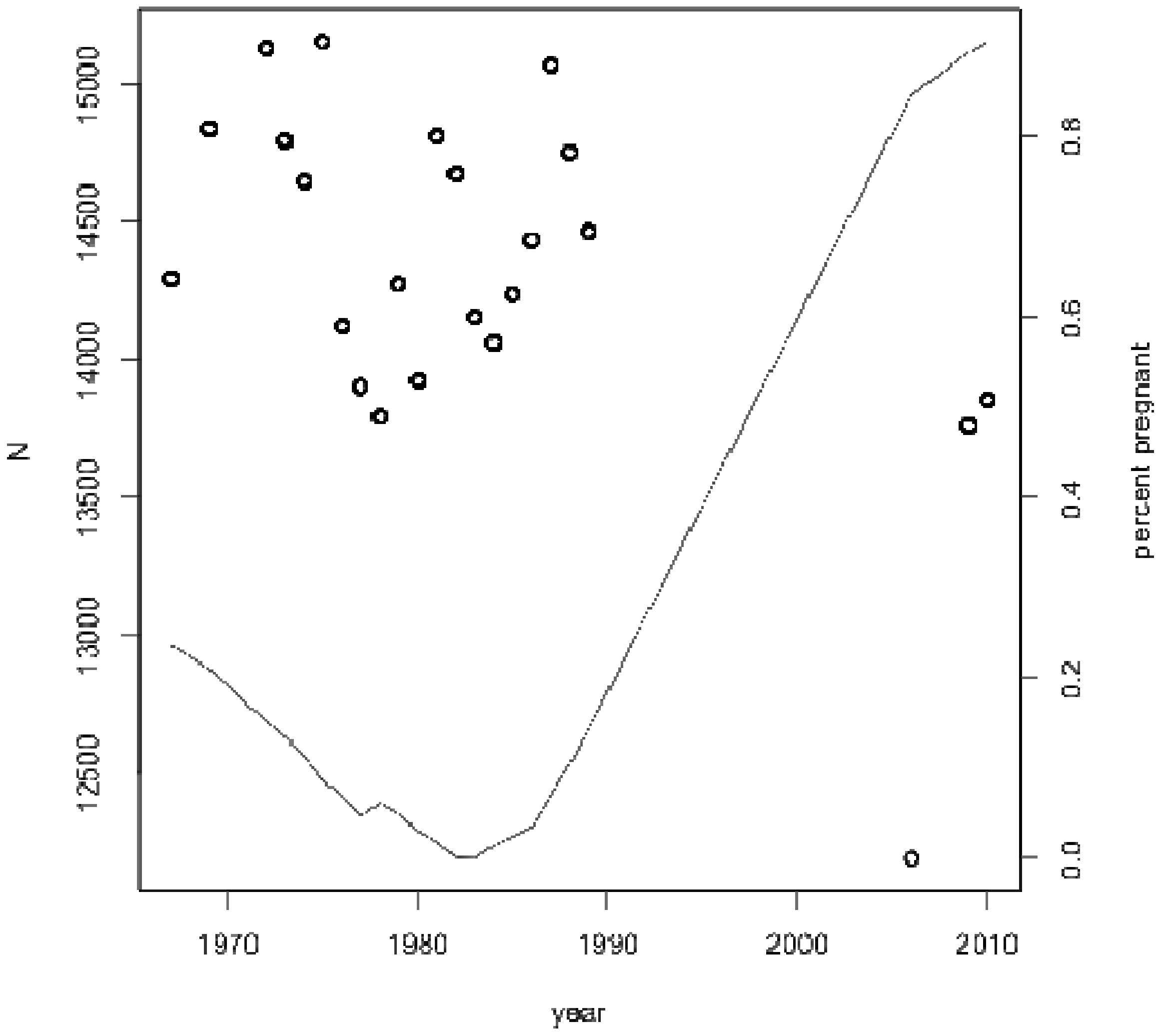

Figure S2. A plot of the key variables in the model, namely modelled abundance in each year, and percent of females that were pregnant in that year. Each year is represented once in this figure.

Expanded information on m4 analysis, including plankton covariate selection

There were 889 measurements of m4. Month 10 (October) had only 3 records in it, so we removed these (886 records). We converted the date that the whale was sampled to day of year. Initial modelling indicated that log(m4) fit much better than m4 (i.e. less patterning of residuals), so we used a log link.

We now want the option of including per capita abundance of the three prey variables: Siglunes (PC_Sig), Selvogsbanki ("PC_Sel"), or the CPR (PC_NeuphS) euphausiid data. This meant deleting any years where all 3 were not measured (no PC_Sel in 1967 and 1969; no PC_NeuphS in 1987–1989), so sample size was reduced to 646 data points.

This creates 53 candidate models. In all cases, we set $k = 4$ to prevent under-smoothing, following recommended practice (Wood, 2006 #32551).

Model	d.f.	AIC	Delta AIC	Model formula
mod8a	9.67	–220.52	0	mod8a<-gam(lm4~s(yday,k=4)+type+s(LE)+s(PC_Sel,k=4),data=dat4,gamma=1.4)
mod6a	11.27	–219.71	0.81	mod6a<-gam(lm4~type+s(yday,k=4)+s(LE,by=type)+s(PC_Sel,k=4),data=dat4,gamma=1.4)
mod12	7.93	–218.14	2.38	mod12<-gam(lm4~s(yday,k=4)+type+s(LE),data=dat4,gamma=1.4)
mod7a	11.90	–217.99	2.53	mod7a<-gam(lm4~type+s(yday,k=4)+s(LE)+s(PC_Sel,k=4,by=type),data=dat4,gamma=1.4)
mod4a	13.42	–217.35	3.17	mod4a<-gam(lm4~type+s(yday,k=4)+s(LE,by=type)+s(PC_Sel,k=4,by=type),data=dat4,gamma=1.4)
mod8c	8.89	–216.65	3.87	mod8c<-gam(lm4~s(yday,k=4)+type+s(LE)+s(PC_NeuphS,k=4),data=dat4,gamma=1.4)
mod8b	8.96	–216.41	4.11	mod8b<-gam(lm4~s(yday,k=4)+type+s(LE)+s(PC_Sig,k=4),data=dat4,gamma=1.4)
mod6c	10.46	–215.91	4.61	mod6c<-gam(lm4~type+s(yday,k=4)+s(LE,by=type)+s(PC_NeuphS,k=4),data=dat4,gamma=1.4)
mod6b	10.51	–215.66	4.86	mod6b<-gam(lm4~type+s(yday,k=4)+s(LE,by=type)+s(PC_Sig,k=4),data=dat4,gamma=1.4)
mod5a	13.64	–215.58	4.94	mod5a<-gam(lm4~type+s(yday,k=4,by=type)+s(LE)+s(PC_Sel,k=4),data=dat4,gamma=1.4)
mod7b	10.93	–214.77	5.75	mod7b<-gam(lm4~type+s(yday,k=4)+s(LE)+s(PC_Sig,k=4,by=type),data=dat4,gamma=1.4)
mod2a	15.29	–214.63	5.89	mod2a<-gam(lm4~type+s(yday,k=4,by=type)+s(LE,by=type)+s(PC_Sel,k=4),data=dat4,gamma=1.4)
mod3a	16.24	–214.36	6.16	mod3a<-gam(lm4~type+s(yday,k=4,by=type)+s(LE)+s(PC_Sel,k=4,by=type),data=dat4,gamma=1.4)
mod4b	12.55	–214.20	6.32	mod4b<-gam(lm4~type+s(yday,k=4)+s(LE,by=type)+s(PC_Sig,k=4,by=type),data=dat4,gamma=1.4)
mod1a	17.98	–213.91	6.61	mod1a<-gam(lm4~type+s(yday,k=4,by=type)+s(LE,by=type)+s(PC_Sel,k=4,by=type),data=dat4,gamma=1.4)
mod7c	10.82	–213.91	6.61	mod7c<-gam(lm4~type+s(yday,k=4)+s(LE)+s(PC_NeuphS,k=4,by=type),data=dat4,gamma=1.4)
mod4c	12.45	–213.05	7.47	mod4c<-gam(lm4~type+s(yday,k=4)+s(LE,by=type)+s(PC_NeuphS,k=4,by=type),data=dat4,gamma=1.4)
mod5c	12.89	–211.06	9.46	mod5c<-gam(lm4~type+s(yday,k=4,by=type)+s(LE)+s(PC_NeuphS,k=4),data=dat4,gamma=1.4)
mod5b	13.01	–210.94	9.58	mod5b<-gam(lm4~type+s(yday,k=4,by=type)+s(LE)+s(PC_Sig,k=4),data=dat4,gamma=1.4)
mod1b	19.86	–210.62	9.90	mod1b<-gam(lm4~type+s(yday,k=4,by=type)+s(LE,by=type)+s(PC_Sig,k=4,by=type),data=dat4,gamma=1.4)
mod3b	16.70	–210.53	9.99	mod3b<-gam(lm4~type+s(yday,k=4,by=type)+s(LE)+s(PC_Sig,k=4,by=type),data=dat4,gamma=1.4)
mod2c	14.51	–210.19	10.33	mod2c<-gam(lm4~type+s(yday,k=4,by=type)+s(LE,by=type)+s(PC_NeuphS,k=4),data=dat4,gamma=1.4)
mod2b	14.59	–210.05	10.47	mod2b<-gam(lm4~type+s(yday,k=4,by=type)+s(LE,by=type)+s(PC_Sig,k=4),data=dat4,gamma=1.4)
mod3c	14.79	–207.83	12.69	mod3c<-gam(lm4~type+s(yday,k=4,by=type)+s(LE)+s(PC_NeuphS,k=4,by=type),data=dat4,gamma=1.4)
mod1c	16.44	–206.87	13.65	mod1c<-gam(lm4~type+s(yday,k=4,by=type)+s(LE,by=type)+s(PC_NeuphS,k=4,by=type),data=dat4,gamma=1.4)
mod18	6.52	–192.26	28.26	mod18<-gam(lm4~s(yday,k=4)+type,data=dat4,gamma=1.4)
mod11a	8.07	–192.07	28.45	mod11a<-gam(lm4~s(yday,k=4)+type+s(PC_Sel,k=4),data=dat4,gamma=1.4)
mod11b	7.55	–190.67	29.85	mod11b<-gam(lm4~s(yday,k=4)+type+s(PC_Sig,k=4),data=dat4,gamma=1.4)
mod11c	7.51	–190.39	30.13	mod11c<-gam(lm4~s(yday,k=4)+type+s(PC_NeuphS,k=4),data=dat4,gamma=1.4)

mod9b	8.20	–181.49	39.03	mod9b<-gam(lm4~type+s(LE)+s(PC_Sig,k=4),data=dat4,gamma=1.4)
mod9a	7.20	–180.45	40.08	mod9a<-gam(lm4~type+s(LE)+s(PC_Sel,k=4),data=dat4,gamma=1.4)
mod16	5.39	–178.84	41.68	mod16<-gam(lm4~type+s(LE),data=dat4,gamma=1.4)
mod9c	6.38	–176.99	43.53	mod9c<-gam(lm4~type+s(LE)+s(PC_NeuphS,k=4),data=dat4,gamma=1.4)
mod14b	6.74	–159.69	60.83	mod14b<-gam(lm4~type+s(PC_Sig,k=4),data=dat4,gamma=1.4)
mod20	4.00	–157.40	63.12	mod20<-gam(lm4~type,data=dat4,gamma=1.4)
mod14a	5.69	–157.33	63.19	mod14a<-gam(lm4~type+s(PC_Sel,k=4),data=dat4,gamma=1.4)
mod14c	5.00	–155.43	65.09	mod14c<-gam(lm4~type+s(PC_NeuphS,k=4),data=dat4,gamma=1.4)
mod10a	7.32	–142.33	78.19	mod10a<-gam(lm4~s(yday,k=4)+s(LE)+s(PC_Sel,k=4),data=dat4,gamma=1.4)
mod17	5.40	–134.24	86.28	mod17<-gam(lm4~s(yday,k=4)+s(LE),data=dat4,gamma=1.4)
mod10b	6.36	–132.91	87.61	mod10b<-gam(lm4~s(yday,k=4)+s(LE)+s(PC_Sig,k=4),data=dat4,gamma=1.4)
mod10c	6.38	–132.40	88.12	mod10c<-gam(lm4~s(yday,k=4)+s(LE)+s(PC_NeuphS,k=4),data=dat4,gamma=1.4)
mod13a	5.15	–113.55	106.97	mod13a<-gam(lm4~s(LE)+s(PC_Sel,k=4),data=dat4,gamma=1.4)
mod13b	5.78	–111.37	109.15	mod13b<-gam(lm4~s(LE)+s(PC_Sig,k=4),data=dat4,gamma=1.4)
mod21	3.00	–105.39	115.13	mod21<-gam(lm4~s(LE),data=dat4,gamma=1.4)
mod13c	4.00	–103.51	117.01	mod13c<-gam(lm4~s(LE)+s(PC_NeuphS,k=4),data=dat4,gamma=1.4)
mod15a	6.14	–77.05	143.47	mod15a<-gam(lm4~s(yday,k=4)+s(PC_Sel,k=4),data=dat4,gamma=1.4)
mod15b	7.11	–76.03	144.49	mod15b<-gam(lm4~s(yday,k=4)+s(PC_Sig,k=4),data=dat4,gamma=1.4)
mod19	4.42	–74.47	146.05	mod19<-gam(lm4~s(yday,k=4),data=dat4,gamma=1.4)
mod15c	5.42	–72.44	148.08	mod15c<-gam(lm4~s(yday,k=4)+s(PC_NeuphS,k=4),data=dat4,gamma=1.4)
mod22b	4.82	–58.41	162.11	mod22b<-gam(lm4~s(PC_Sig,k=4),data=dat4,gamma=1.4)
mod22a	3.92	–55.13	165.39	mod22a<-gam(lm4~s(PC_Sel,k=4),data=dat4,gamma=1.4)
mod23	2.00	–51.89	168.63	mod23<-gam(lm4~1,data=dat4,gamma=1.4)
mod22c	3.00	–49.88	170.65	mod22c<-gam(lm4~s(PC_NeuphS,k=4),data=dat4,gamma=1.4)

The preferred model was:

mod8a<gam(lm4~s(yday,k=4)+type+s(LE)+s(PC_Sel,k=4), data=dat4,gamma=1.4)

The model diagnostics showed some undue influence of a few datapoints for prey abundance. Rather than delete outliers, we conditioned on this selected model (i.e. using PC_Sel), but using all possible data. That is, now that we have decided to use Selvogsbanki as the plankton covariate, we can use additional data from whales that were sampled during years when we did have data from Selvogsbanki, but did not have data from Siglunes or CPR. This brought our sample size up to 771 whales.

We redid the analysis of the top 5 models (down to DeltaAICc of 3.2)

mod8a<-gam(lm4~s(yday,k=4)+type+s(LE)+s(PC_Sel,k=4),data=dat4,gamma=1.4)
mod6a<-gam(lm4~type+s(yday,k=4)+s(LE,by=type)+s(PC_Sel,k=4),data=dat4,gamma=1.4)
mod12<-gam(lm4~s(yday,k=4)+type+s(LE),data=dat4,gamma=1.4)
mod7a<-gam(lm4~type+s(yday,k=4)+s(LE)+s(PC_Sel,k=4,by=type),data=dat4,gamma=1.4)
mod4a<-gam(lm4~type+s(yday,k=4)+s(LE,by=type)+s(PC_Sel,k=4,by=type),data=dat4,gamma=1.4)

and found that the AICc of the top 3 models were:

Model	**d.f.**	**AICc**	**DeltaAIC**
mod8a	9.884063	–219.8712	0.000000
mod6a	11.733774	–216.9364	2.934847
mod7a	13.545602	–214.1666	5.704689

Again, model 8a was preferred.

The selected model was of the form:

Family: gaussian
Link function: identity

Formula: lm4 ~ s(yday, k = 4) + type + s(LE) + s(PC_Sel, k = 4)

Parametric coefficients:

	Estimate	**s.e.**	***t*-value**	**Pr(>|*t*|)**
(Intercept)	4.016519	0.015405	60.721	<2e–16 ***
typeFP	0.166323	0.019265	8.633	<2e–16 ***
typeM	0.009395	0.021717	0.433	0.665

Significance codes: 0, ***; 0.001, **; 0.01, *.

Approximate significance of smooth terms:

	e d.f.	**Reference d.f.**	***f***	***p*-value**
s(yday)	2.471	2.799	14.16	2.60e–08 ***
s(LE)	1.148	1.282	25.33	8.99e–08 ***
s(PC_Sel)	2.265	2.623	12.72	4.07e–07 ***

Significance codes: 0, ***; 0.001, **; 0.01, *.

r^2 (adj) = 0.239, deviance explained = 24.7%, GCV score = 0.04431, scale est. = 0.043392, n = 771

The model provided good fit to the data. There were no patterns in the residuals (Figure S3).

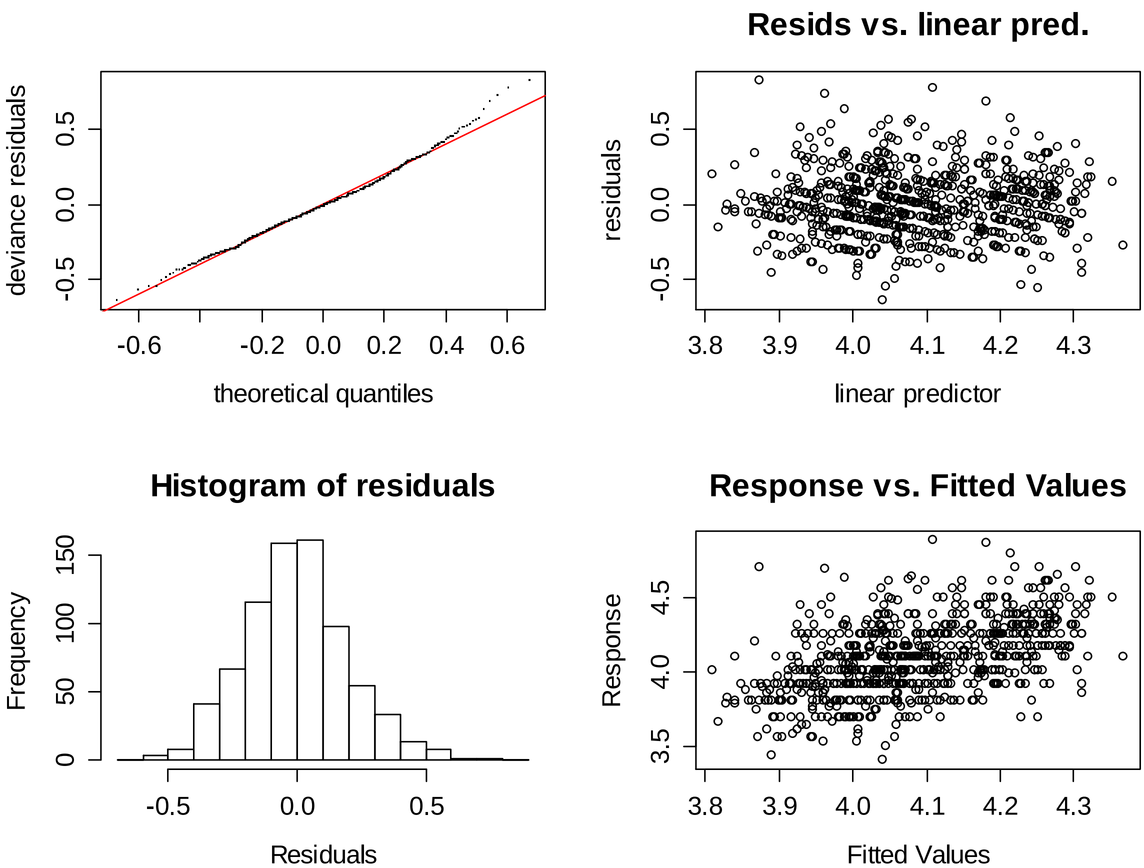

Figure S3. Diagnostic plots from the selected model (8a) showing variability in medial blubber thickness (m4).

The selected model of blubber thickness (m4) is shown in Figure S4.

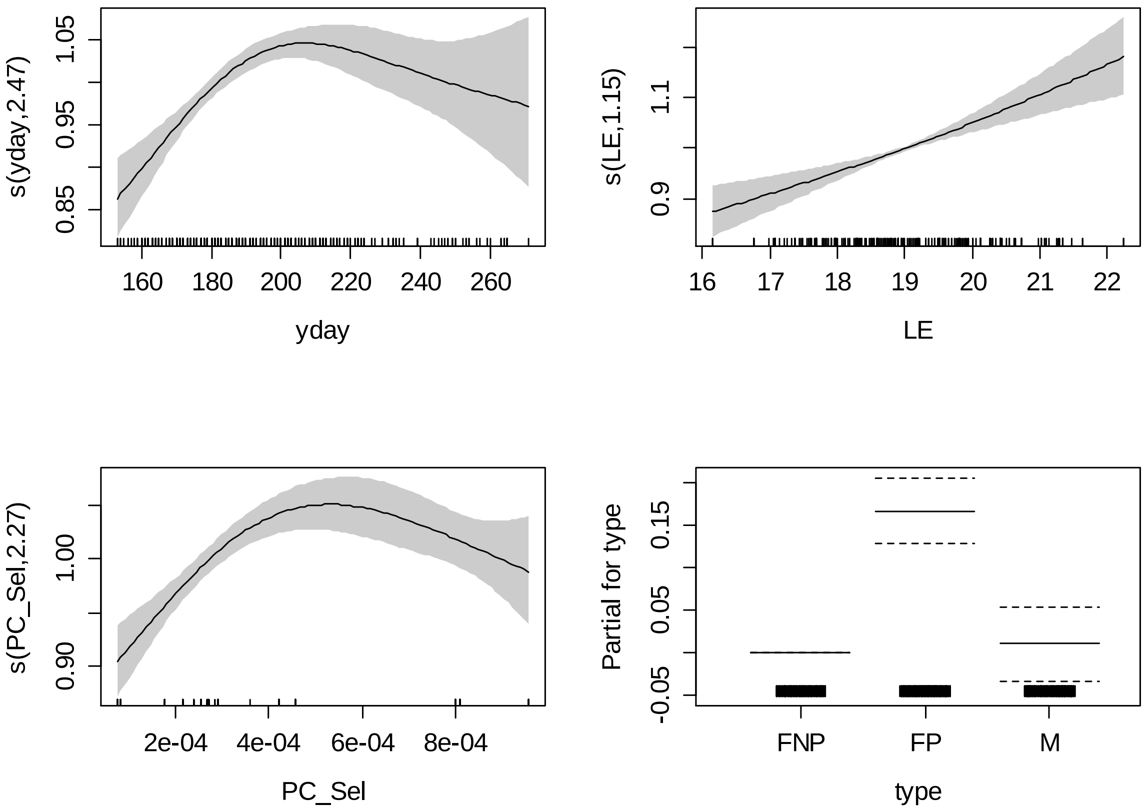

Figure S4. The selected model showing variability in medial blubber thickness (m4) as termwise/partial functions (i.e. conditional on all other terms being in the model) of day of year (yday), length (LE, in meters), per capita prey availability (PC Sel) and reproductive type [male (M), pregnant (FP), and non-pregnant females (FNP)]. Vertical ticks on the x-axis (rugplot) shows distribution of raw data. Grey shading indicate 95% confidence intervals on the termwise relationship.

The asymptotic relationship between blubber thickness (m4, in mm) is shown in Figure S5.

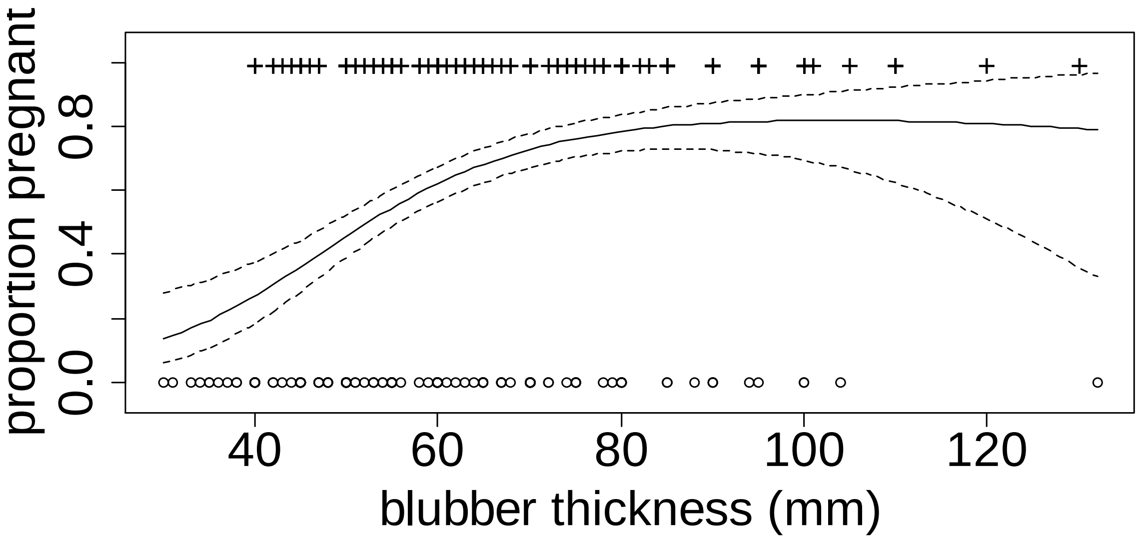

Figure S5. The selected model showing variability in pregnancy (proportion pregnant) as a smooth function of medial blubber thickness (m4, in mm). Crosses and circles show data values for pregnant and non-pregnant females, respectively. Grey shading indicate 95% confidence intervals.

The selected pregnancy model was of the form:

Family: binomial
Link function: logit

Formula: pregnant ~ s(m4)

Parametric coefficients:

	Estimate	**s.e.**	***z*-value**	**Pr(>|*z*|)**
(Intercept)	0.52531	0.09794	5.364	8.15e–08 ***

Significance codes: 0, ***; 0.001, **; 0.01, *.

Approximate significance of smooth terms:

	e d.f.	**Reference d.f.**	χ^2	***p*-value**
s(m4)	2.297	2.945	51.62	4.41e–11 ***

Significance codes: 0, ***; 0.001, **; 0.01, *.

r^2 (adj) = 0.116, deviance explained = 9.17%, UBRE score = 0.22581, scale est. = 1, n = 506

Checking that another blubber metric (v4) provides similar results
As a check on our model-selection approach, we only looked at PC_Sel, based on the previous analysis. There are 870 of these, 755 after keeping only those with PC_Sel values. Model selection is similar, with model 8 being selected. Note that it was necessary to set k=7 to prevent an overly wiggly (and nonsensical) fit. Given this, model selection actually chose a linear term for PC_Sel.

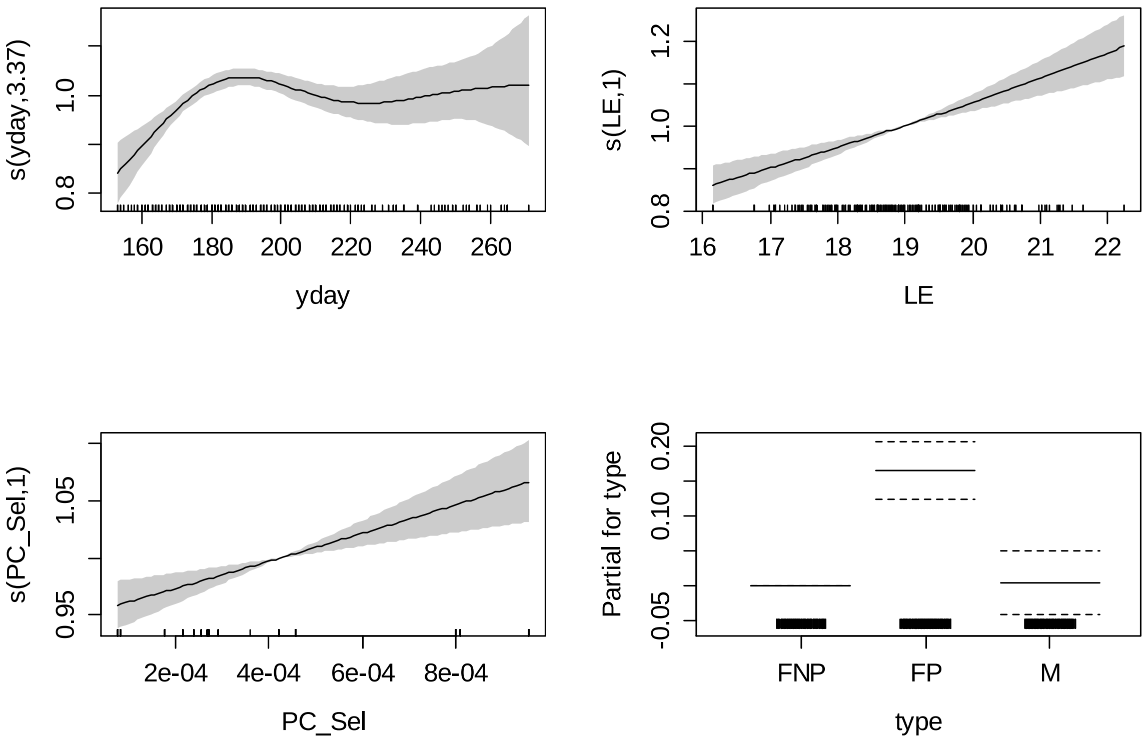

Figure S6. The selected model showing variability in ventral blubber thickness (v4) as termwise/partial functions (i.e. conditional on all other terms being in the model) of day of year (yday), length (LE, in meters), per capita prey availability (PC_Sel) and reproductive type [male (M), pregnant (FP), and non-pregnant females (FNP)]. Vertical ticks on the x-axis (rugplot) shows distribution of raw data. Grey shading indicate 95% confidence intervals on the termwise relationship.

The residual diagnostics look good (Figure S7).

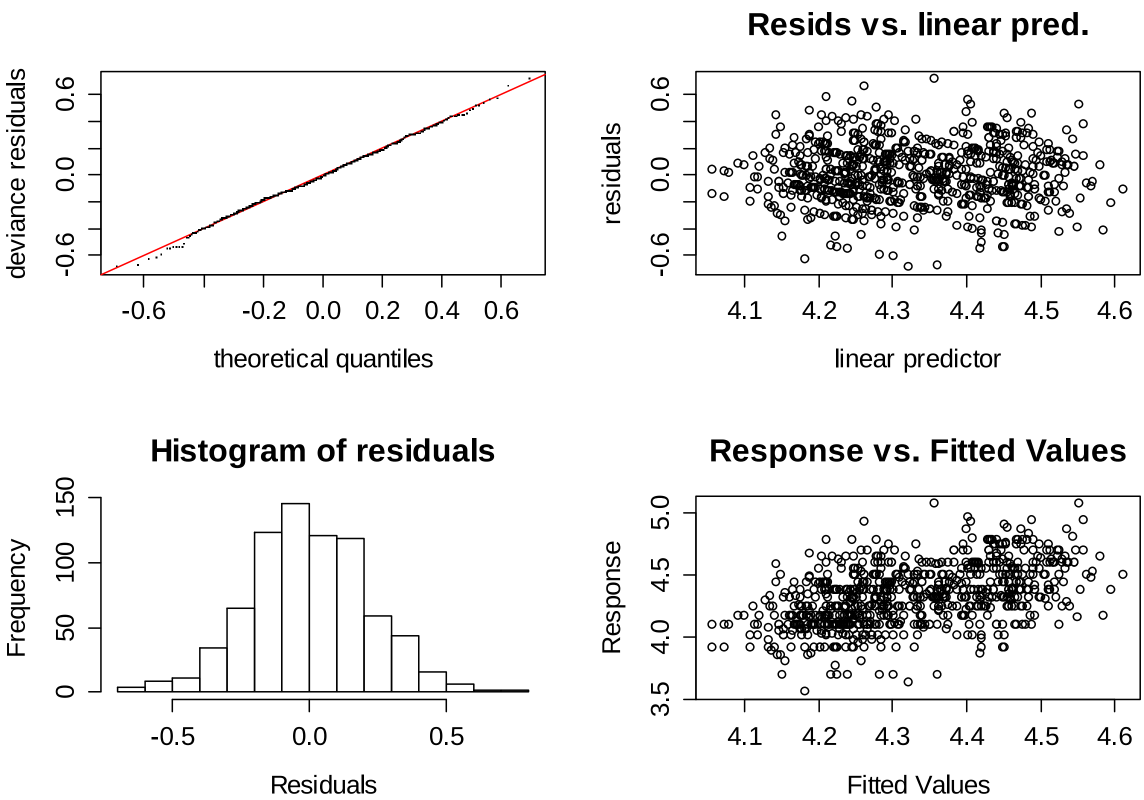

Figure S7. Diagnostic plots from the selected model (8a) showing variability in ventral blubber thickness (v4).

The selected model describing variability in ventral blubber thickness (v4) was of the form:

Family: gaussian
Link function: identity

Formula: lm4 ~ s(yday, k = 5) + type + s(LE) + s(PC_Sel, k = 7)

Parametric coefficients:

	Estimate	Error	*t*-value	Pr(>|*t*|)
(Intercept)	4.257051	0.016914	251.691	< 2e–16 ***
typeFP	0.164412	0.020722	7.934	7.76e–15 ***
typeM	0.005098	0.022993	0.222	0.825

Significance codes: 0, ***; 0.001, **; 0.01, *.

Approximate significance of smooth terms:

	e d.f.	Reference d.f.	*f*	*p*-value
s(yday)	3.373	3.778	6.017	0.000152 ***
s(LE)	1.000	1.000	32.590	1.62e–08 ***
s(PC_Sel)	1.000	1.000	14.994	0.000117 ***

Significance codes: 0, ***; 0.001, **; 0.01, *.

r^2 (adj) = 0.224, deviance explained = 23.1%, GCV score = 0.047249, scale est. = 0.046307, n = 755

The relationship between pregnancy rate and ventral blubber thickness (Figure S8) was similar to the one between pregnancy rate and medial blubber thickness.

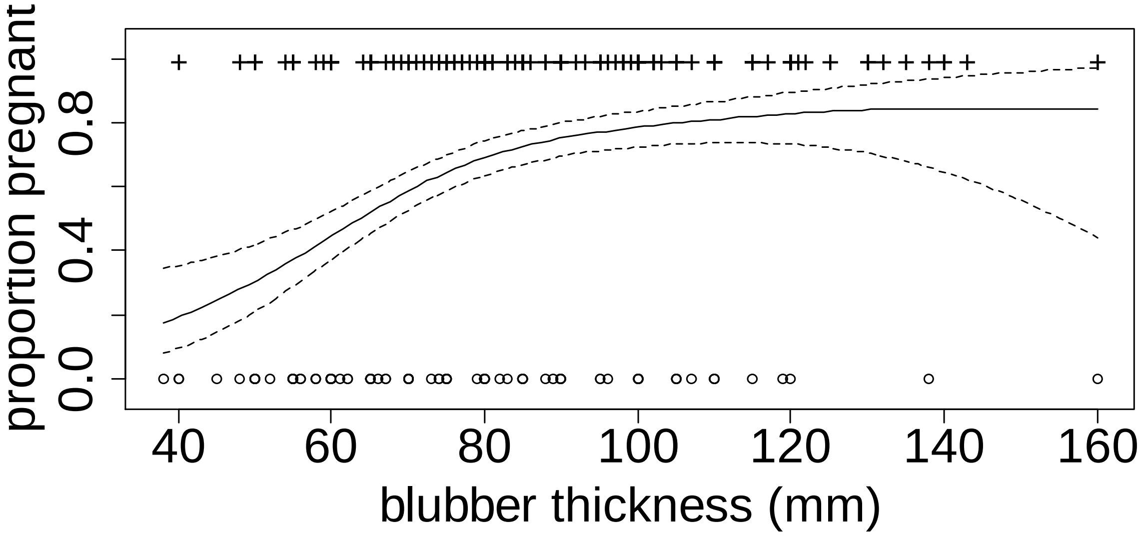

Figure S8. The selected model showing variability in pregnancy (proportion pregnant) as a smooth function of ventral blubber thickness (v4, in mm). Crosses and circles show data values for pregnant and non-pregnant females, respectively. Grey shading indicate 95% confidence intervals.

The selected model describing pregnancy rate as a function of ventral blubber thickness was of the form:

Family: binomial
Link function: logit

Formula: pregnant ~ s(m4)

Parametric coefficients:

	Estimate	**s.e.**	***z*-value**	**Pr(>|z|)**
(Intercept)	0.6769	0.1018	6.65	2.93e–11 ***

Significance codes: 0, ***; 0.001, **; 0.01, *.

Approximate significance of smooth terms:

	e d.f.	**Reference d.f.**	χ^2	***p*-value**
s(m4)	2.262	2.901	48.33	2.42e–10 ***

Significance codes: 0, ***; 0.001, **; 0.01, *.

r^2 (adj) = 0.11, deviance explained = 8.66%, UBRE score = 0.20042, scale est. = 1, n = 487