APPENDICES AND SUPPORTING INFORMATION

Appendix 1: Raw completeness, time series, and proxy data (.xlsx file). Contents of Appendix 1(Excel file):

Sheet 1: Information on every ichthyosaur specimen included in the study
Sheet 2: Guide to metric values for SCM and BSCM
Sheet 3: Abbreviations of museums and institutions holding specimens
Sheet 4: Completeness values for all specimens
Sheet 5: SCM and BSCM 1 (best preserved specimen) for all species
Sheet 6: SCM and BSCM 2 (composite completeness) for all species
Sheet 7: Raw data for correlations and proxy time series
Sheet 8: Raw data for SCM-BSCM 1 time series
Sheet 9: Raw data for SCM-BSCM 2 time series
Sheet 10: Raw data for body size comparisons
Sheet 11: Raw data for hemisphere comparison
Sheet 12: Raw data for grain size comparison
Sheet 13: Raw data for rock composition comparison
Sheet 14: Raw data for overall lithology (grain size + composition) comparison
Sheet 15: References for all specimens taken from the literature

Sheet 1 of Excel file

This sheet contains background information for every specimen coded for completeness in the study. There are a great number of columns divided into broad sections, which I will elaborate on here.

Taxonomy

It is obviously important to know which taxonomic groups each specimen belongs to, in order to create some separation between them. Each specimen is identified by a unique museum code (or assigned one if lacking). Noted is the instance of holotype specimens, since these tend to be the best-preserved fossils. Key papers are listed, from which completeness information is derived. If a specimen is museum-based, it is marked with the designation "MUS" followed by the name of the institution.

Geography

Obtaining a geographical location helps to distinguish fossil localities from each other, enabling comparisons to be made.

Temporal distribution

Knowing which formation an ichthyosaur is from helps to narrow down its location in time. It also allows us to potentially identify the geology and facies of the places specimens are found. The time period a locality was created is noted down to substage and ammonite zone if possible. This helps when studying completeness through the Mesozoic, as scores are divisible into small time bins.

Ecology

Additional data on size and lifestyle were obtained to provide a bigger picture about each ichthyosaur. Collecting the lengths of various body parts enables a rough estimate of body size to be obtained, even if body length is not known. Ecotype and diet are also noted, though not used in this study.

Geology

Knowing about the geological setting of each specimen is important, as one of the main aims of this study is to examine the effects of facies types on completeness. All facies are marine, so this column is not of use in this particular investigation. Each locality was separated into fine, coarse, and non-specific (not noted in the literature) siliciclastic and carbonate categories (or a combination of the two). Sedimentary facies are indicative of depositional environment, and studying differences in fossil completeness can tell us which of these preserve ichthyosaurs with the highest fidelity. Whether a locality is a lagerstätte or not is recorded. Further details about lithology ("comments on geology) are gathered, e.g. the presence of black shales, as these can indicate particular conditions at the time of specimen burial that may help or hinder preservation.

Appendix 2: Extra figures referenced in text

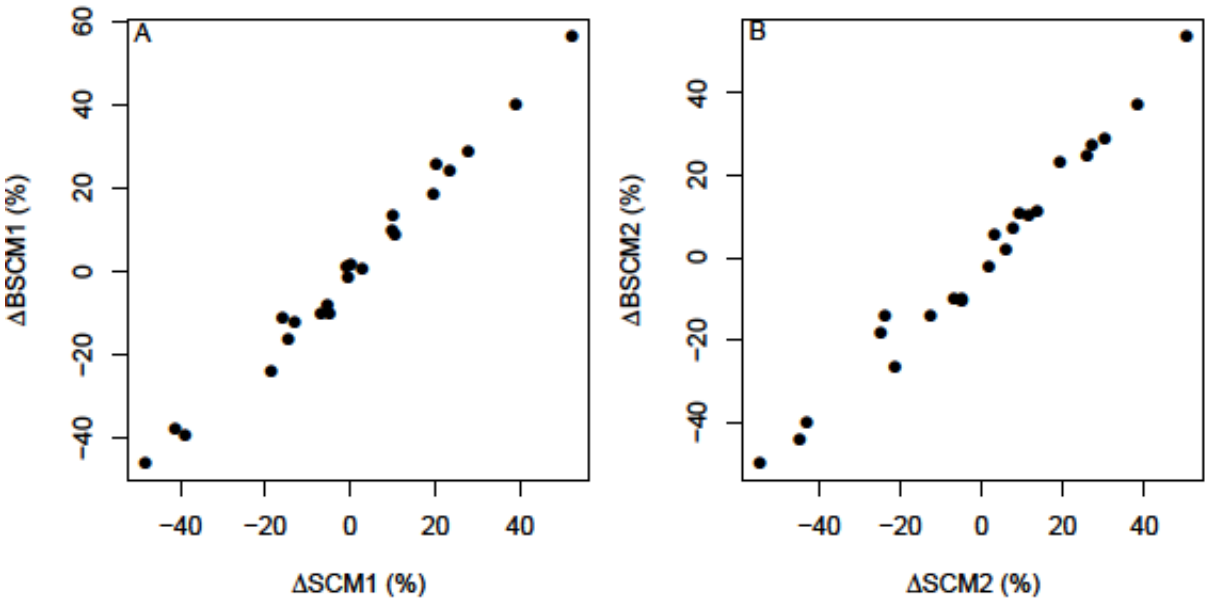

Figure S1: Correlation plots showing the relationships between completeness metrics. Δ indicates that data has undergone generalised-differencing prior to application of Spearman Rank correlation tests. **(A)** SCM1 and BSCM1, and **(B)** SCM2 and BSCM2.

Appendix 3: Extra tables referenced in text

Table S1. Spearman's Rank correlation coefficients (r_s) between time series data for all time bins, Triassic-Jurassic time bins, and Cretaceous time bins including time bins with no ichthyosaur occurrences (i.e. Bathonian and Valanginian). * = *p* significant at 0.05 and ** = *p* significant after false discovery rate corrections (Benjamini & Hochberg 1995).

	All time bins	Triassic-Jurassic	Cretaceous
SCM1 and BSCM1	0.98**	0.99**	0.6
SCM2 and BSCM2	0.99**	0.98**	1*
Diversity and collections	0.48	0.33	0.54
Diversity and FMFs	0.26	-0.13	0.31
Diversity and sea level	-0.03	-0.32	0.03
Diversity and SCM1	0.25	0.35	0.83
Diversity and SCM2	0.29	0.41	0.54
Diversity and BSCM1	0.26	0.31	0.54
Diversity and BSCM2	0.3	0.34	0.54
Collections and FMFs	0.31	0.17	0.89*
Collections and sea level	0.03	0	0.03
Collections and SCM1	0.49*	0.36	0.26
Collections and SCM2	0.43*	0.35	0.66
Collections and BSCM1	0.49*	0.42	0.66
Collections and BSCM2	0.45*	0.35	0.66
FMFs and sea level	0.1	0.46	0.31
FMFs and SCM1	-0.26	-0.39	-0.09
FMFs and SCM2	-0.24	-0.38	0.54
FMFs and BSCM1	-0.23	-0.34	0.54
FMFs and BSCM2	-0.24	-0.4	0.54
Sea level and SCM1	-0.38	-0.74*	0.03
Sea level and SCM2	-0.42	-0.78*	0.54
Sea level and BSCM1	-0.38	-0.72*	0.54
Sea level and BSCM2	-0.42	-0.74*	0.54

Appendix 4: Multiple regression method and results

Four multiple regression models were run with each of the completeness metrics as the dependent variables (SCM1, SCM2, BSCM1, BSCM2), and diversity, collections, formations, sea level and time period (Triassic/Jurassic or Cretaceous) as the independent variables. We omitted the Bathonian and Valaginian time bins as they have no ichthyosaur occurrences and therefore no completeness metric information. All data was log transformed and therefore we added a value of 10.232929658 to all sea level data in order to make the lowest sea level (i.e. -9.232929658 for the Rhaetian) a value of 1 to enable log transformation without the removing the pattern in the data. As the data is log transformed, the first two time bins (i.e. Olenekian and Anisian) are not included in the multiple regression analyses as they do not contain any sea level data. Triassic/Jurassic time bins were labelled as 1 and Cretaceous time bins labelled as 2.

Results

SCM1

Full model:
Call: lm(formula = log_scm1 ~ log_div + log_coll + log_form + log_sl + log_per)

Residuals:

Min	1Q	Median	3Q	Max
-0.7537	-0.1209	-0.0395	0.1343	0.5361

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	5.302292	0.581675	9.116	2.9e-07 ***
log_div	-0.001864	0.114461	-0.016	0.9872
log_coll	0.243813	0.119317	2.043	0.0603 .
log_form	-0.346848	0.132848	-2.611	0.0205 *
log_sl	-0.099553	0.072491	-1.373	0.1913
log_per	-0.312603	0.280814	-1.113	0.2844

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3349 on 14 degrees of freedom
Multiple R-squared: 0.5948, Adjusted R-squared: 0.4501
F-statistic: 4.11 on 5 and 14 DF, p-value: 0.01659

Model selection:
Start: AIC=-38.89
log_scm1 ~ log_div + log_coll + log_form + log_sl + log_per

	Df	Sum of Sq	RSS	AIC
- log_div	1	0.00003	1.5703	-40.890
- log_per	1	0.13899	1.7092	-39.194
<none>			1.5702	-38.890
- log_sl	1	0.21153	1.7818	-38.362
- log_coll	1	0.46833	2.0386	-35.670
- log_form	1	0.76455	2.3348	-32.956

Step: AIC=-40.89
log_scm1 ~ log_coll + log_form + log_sl + log_per

	Df	Sum of Sq	RSS	AIC
- log_per	1	0.14018	1.7105	-41.179
<none>			1.5703	-40.890
- log_sl	1	0.21270	1.7830	-40.349
+ log_div	1	0.00003	1.5702	-38.890
- log_coll	1	0.60433	2.1746	-36.378
- log_form	1	0.76719	2.3375	-34.933

Step: AIC=-41.18
log_scm1 ~ log_coll + log_form + log_sl

	Df	Sum of Sq	RSS	AIC
<none>			1.7105	-41.179
+ log_per	1	0.14018	1.5703	-40.890
+ log_div	1	0.00122	1.7092	-39.194
- log_sl	1	0.53045	2.2409	-37.777
- log_coll	1	0.77288	2.4833	-35.723
- log_form	1	0.85654	2.5670	-35.060

Best model:
Call: lm(formula = log_scm1 ~ log_coll + log_form + log_sl)

Residuals:

Min	1Q	Median	3Q	Max
-0.70370	-0.17392	-0.02218	0.15285	0.59228

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	5.41767	0.55844	9.702	4.18e-08 ***
log_coll	0.26817	0.09974	2.689	0.0161 *
log_form	-0.36414	0.12865	-2.831	0.0121 *
log_sl	-0.13824	0.06206	-2.228	0.0406 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.327 on 16 degrees of freedom
Multiple R-squared: 0.5586, Adjusted R-squared: 0.4758
F-statistic: 6.749 on 3 and 16 DF, p-value: 0.003748

SCM2

Full model:
Call: lm(formula = log_scm2 ~ log_div + log_coll + log_form + log_sl + log_per)

Residuals:

Min	1Q	Median	3Q	Max
-0.75293	-0.13950	-0.03062	0.20493	0.59294

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	5.33967	0.69278	7.708	2.11e-06 ***
log_div	0.05730	0.13632	0.420	0.6806
log_coll	0.26814	0.14211	1.887	0.0801 .
log_form	-0.35321	0.15822	-2.232	0.0424 *
log_sl	-0.12810	0.08634	-1.484	0.1601
log_per	-0.18856	0.33445	-0.564	0.5818

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3989 on 14 degrees of freedom
Multiple R-squared: 0.535,   Adjusted R-squared: 0.3689
F-statistic: 3.221 on 5 and 14 DF, p-value: 0.03835

Model selection:
Start: AIC=-31.9
log_scm2 ~ log_div + log_coll + log_form + log_sl + log_per

	Df	Sum of Sq	RSS	AIC
- log_div	1	0.02811	2.2555	-33.647
- log_per	1	0.05057	2.2780	-33.449
<none>			2.2274	-31.898
- log_sl	1	0.35023	2.5776	-30.977
- log_coll	1	0.56643	2.7938	-29.366
- log_form	1	0.79284	3.0202	-27.808

Step: AIC=-33.65
log_scm2 ~ log_coll + log_form + log_sl + log_per

	Df	Sum of Sq	RSS	AIC
- log_per	1	0.04509	2.3006	-35.251
<none>			2.2555	-33.647
- log_sl	1	0.33916	2.5947	-32.846
+ log_div	1	0.02811	2.2274	-31.898
- log_form	1	0.77935	3.0349	-29.711
- log_coll	1	0.90272	3.1582	-28.914

Step: AIC=-35.25
log_scm2 ~ log_coll + log_form + log_sl

	Df	Sum of Sq	RSS	AIC
<none>			2.3006	-35.251
+ log_per	1	0.04509	2.2555	-33.647
+ log_div	1	0.02263	2.2780	-33.449
- log_sl	1	0.60555	2.9061	-32.578
- log_form	1	0.83458	3.1352	-31.061
- log_coll	1	1.04073	3.3413	-29.787

Best model:
Call: lm(formula = log_scm2 ~ log_coll + log_form + log_sl)

Residuals:

Min	1Q	Median	3Q	Max
-0.75405	-0.20187	0.01092	0.23791	0.58692

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	5.39280	0.64764	8.327	3.29e-07 ***
log_coll	0.31119	0.11567	2.690	0.0161 *
log_form	-0.35945	0.14920	-2.409	0.0284 *
log_sl	-0.14771	0.07197	-2.052	0.0569 .

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3792 on 16 degrees of freedom
Multiple R-squared: 0.5197, Adjusted R-squared: 0.4296
F-statistic: 5.77 on 3 and 16 DF, p-value: 0.007153

BSCM1

Full model:
Call: lm(formula = log_bsc1 ~ log_div + log_coll + log_form + log_sl + log_per)

Residuals:

Min	1Q	Median	3Q	Max
-0.70809	-0.21455	-0.07733	0.23175	0.60339

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	5.25531	0.70418	7.463	3.05e-06 ***
log_div	-0.02775	0.13857	-0.200	0.8442
log_coll	0.28270	0.14445	1.957	0.0706 .
log_form	-0.33729	0.16083	-2.097	0.0546 .
log_sl	-0.11575	0.08776	-1.319	0.2084
log_per	-0.52616	0.33996	-1.548	0.1440

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4054 on 14 degrees of freedom
Multiple R-squared: 0.5938, Adjusted R-squared: 0.4488
F-statistic: 4.093 on 5 and 14 DF, p-value: 0.01684

Model selection:
Start: AIC=-31.24
log_bsc1 ~ log_div + log_coll + log_form + log_sl + log_per

	Df	Sum of Sq	RSS	AIC
- log_div	1	0.00659	2.3079	-33.188

<none>			2.3013	-31.245
- log_sl	1	0.28594	2.5873	-30.903
- log_per	1	0.39377	2.6951	-30.086
- log_coll	1	0.62964	2.9310	-28.408
- log_form	1	0.72299	3.0243	-27.781

Step: AIC=-33.19
log_bsc1 ~ log_coll + log_form + log_sl + log_per

	Df	Sum of Sq	RSS	AIC
<none>			2.3079	-33.188
- log_sl	1	0.29262	2.6005	-32.800
- log_per	1	0.40433	2.7123	-31.959
+ log_div	1	0.00659	2.3013	-31.245
- log_form	1	0.73231	3.0402	-29.676
- log_coll	1	0.74020	3.0481	-29.624

Best model:
Call: lm(formula = log_bsc2 ~ log_div + log_coll + log_form + log_sl + log_per)

Residuals:

Min	1Q	Median	3Q	Max
-0.73591	-0.19337	-0.01607	0.23257	0.56221

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	5.17385	0.72895	7.098	5.36e-06 ***
log_div	0.02678	0.14344	0.187	0.8546
log_coll	0.28122	0.14953	1.881	0.0810 .
log_form	-0.31870	0.16648	-1.914	0.0762 .
log_sl	-0.12083	0.09085	-1.330	0.2048
log_per	-0.46412	0.35191	-1.319	0.2084

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4197 on 14 degrees of freedom
Multiple R-squared: 0.5658, Adjusted R-squared: 0.4107
F-statistic: 3.649 on 5 and 14 DF, p-value: 0.02535

BSCM2

Full model:
Call: lm(formula = log_bsc2 ~ log_div + log_coll + log_form + log_sl + log_per)

Residuals:

Min	1Q	Median	3Q	Max
-0.73591	-0.19337	-0.01607	0.23257	0.56221

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	5.17385	0.72895	7.098	5.36e-06 ***

log_div	0.02678	0.14344	0.187	0.8546
log_coll	0.28122	0.14953	1.881	0.0810 .
log_form	-0.31870	0.16648	-1.914	0.0762 .
log_sl	-0.12083	0.09085	-1.330	0.2048
log_per	-0.46412	0.35191	-1.319	0.2084

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4197 on 14 degrees of freedom
Multiple R-squared: 0.5658,   Adjusted R-squared: 0.4107
F-statistic: 3.649 on 5 and 14 DF, p-value: 0.02535

<u>Model selection:</u>
Start: AIC=-29.86
log_bsc2 ~ log_div + log_coll + log_form + log_sl + log_per

	Df	Sum of Sq	RSS	AIC
- log_div	1	0.00614	2.4722	-31.813
<none>			2.4660	-29.862
- log_per	1	0.30639	2.7724	-29.520
- log_sl	1	0.31161	2.7776	-29.483
- log_coll	1	0.62305	3.0891	-27.357
- log_form	1	0.64548	3.1115	-27.212

Step: AIC=-31.81
log_bsc2 ~ log_coll + log_form + log_sl + log_per

	Df	Sum of Sq	RSS	AIC
<none>			2.4722	-31.813
- log_per	1	0.30146	2.7736	-31.511
- log_sl	1	0.30733	2.7795	-31.469
+ log_div	1	0.00614	2.4660	-29.862
- log_form	1	0.64065	3.1128	-29.204
- log_coll	1	0.88934	3.3615	-27.667

<u>Best model:</u>
Call: lm(formula = log_bsc2 ~ log_coll + log_form + log_sl + log_per)

Residuals:

Min	1Q	Median	3Q	Max
-0.7494	-0.2060	-0.0174	0.2429	0.5431

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	5.1682	0.7045	7.336	2.46e-06 ***
log_coll	0.2946	0.1268	2.323	0.0347 *
log_form	-0.3171	0.1608	-1.972	0.0674 .
log_sl	-0.1198	0.0877	-1.366	0.1922
log_per	-0.4590	0.3393	-1.352	0.1963

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.406 on 15 degrees of freedom
Multiple R-squared: 0.5647,   Adjusted R-squared: 0.4486
F-statistic: 4.865 on 4 and 15 DF, p-value: 0.01023

Appendix 5: Generalized least squares models

Thirty-one models were assessed, comparing each completeness metric (SCM1, SCM2, BSCM1, BSCM2) with various combinations of diversity, collections, formations, sea level and time period (Triassic/Jurassic or Cretaceous) as the independent variables. We omitted the Bathonian and Valaginian time bins as they have no ichthyosaur occurrences and therefore no completeness metric information. In this case, data did not have to be log-transformed, nor the sea levels made all positive. The first two time bins (i.e. Olenekian and Anisian) are not included in analyses as they do not contain any sea level data. Triassic/Jurassic time bins were labelled as 1 and Cretaceous time bins labelled as 2. Advice on suitable measures of goodness of fit from Box *et al*. (1994), Freese and Long (2006), and the Stata web pages on pseudo R-squareds (http://www.ats.ucla.edu/stat/mult_pkg/faq/general/Psuedo_RSquareds.htm).

R code to compare numerous models through generalized least squares (gls) model fitting

```
library(nlme)
library(qpcR)

y1 <- scm1
y2 <- scm2
y3 <- bsc1
y4 <- bsc2
x1 <- diversity
x2 <- collections
x3 <- formations
x4 <- sealevel
x5 <- time

# Comparing null model versus nested models, using AICc

gls.null <- gls(y1~1, correlation=corARMA(p=1), method="ML")
gls.1 <- gls(y1~x1+x2+x3+x4+x5, correlation=corARMA(p=1), method="ML")
gls.2 <- gls(y1~x1+x2+x3+x4, correlation=corARMA(p=1), method="ML")
gls.3 <- gls(y1~x1+x2+x3+x5, correlation=corARMA(p=1), method="ML")
gls.4 <- gls(y1~x1+x2+x4+x5, correlation=corARMA(p=1), method="ML")
gls.5 <- gls(y1~x1+x3+x4+x5, correlation=corARMA(p=1), method="ML")
gls.6 <- gls(y1~x2+x3+x4+x5, correlation=corARMA(p=1), method="ML")
gls.7 <- gls(y1~x1+x2+x3, correlation=corARMA(p=1), method="ML")
gls.8 <- gls(y1~x1+x2+x4, correlation=corARMA(p=1), method="ML")
gls.9 <- gls(y1~x1+x2+x5, correlation=corARMA(p=1), method="ML")
gls.10 <- gls(y1~x1+x3+x4, correlation=corARMA(p=1), method="ML")
gls.11 <- gls(y1~x1+x3+x5, correlation=corARMA(p=1), method="ML")
gls.12 <- gls(y1~x1+x4+x5, correlation=corARMA(p=1), method="ML")
gls.13 <- gls(y1~x2+x3+x4, correlation=corARMA(p=1), method="ML")
gls.14 <- gls(y1~x2+x3+x5, correlation=corARMA(p=1), method="ML")
gls.15 <- gls(y1~x2+x4+x5, correlation=corARMA(p=1), method="ML")
gls.16 <- gls(y1~x3+x4+x5, correlation=corARMA(p=1), method="ML")
gls.17 <- gls(y1~x1+x2, correlation=corARMA(p=1), method="ML")
gls.18 <- gls(y1~x1+x3, correlation=corARMA(p=1), method="ML")
gls.19 <- gls(y1~x1+x4, correlation=corARMA(p=1), method="ML")
gls.20 <- gls(y1~x1+x5, correlation=corARMA(p=1), method="ML")
gls.21 <- gls(y1~x2+x3, correlation=corARMA(p=1), method="ML")
gls.22 <- gls(y1~x2+x4, correlation=corARMA(p=1), method="ML")
gls.23 <- gls(y1~x2+x5, correlation=corARMA(p=1), method="ML")
gls.24 <- gls(y1~x3+x4, correlation=corARMA(p=1), method="ML")
gls.25 <- gls(y1~x3+x5, correlation=corARMA(p=1), method="ML")
gls.26 <- gls(y1~x4+x5, correlation=corARMA(p=1), method="ML")
gls.27 <- gls(y1~x1, correlation=corARMA(p=1), method="ML")
gls.28 <- gls(y1~x2, correlation=corARMA(p=1), method="ML")
```

```
gls.29 <- gls(y1~x3, correlation=corARMA(p=1), method="ML")
gls.30 <- gls(y1~x4, correlation=corARMA(p=1), method="ML")
gls.31 <- gls(y1~x5, correlation=corARMA(p=1), method="ML")

# Computing and comparing the weightings and AICc values for all models:

weighted<-function (aic){
 aic.wt <- exp(-0.5 * (aic - min(aic)))/sum(exp(-0.5 * (aic - min(aic))))
 return(aic.wt)
 }
weighted(c(AICc(gls.null), AICc(gls.1), AICc(gls.2), AICc(gls.3), AICc(gls.4),
 AICc(gls.5), AICc(gls.6), AICc(gls.7), AICc(gls.8), AICc(gls.9), AICc(gls.10),
 AICc(gls.11), AICc(gls.12), AICc(gls.13), AICc(gls.14), AICc(gls.15),
 AICc(gls.16), AICc(gls.17), AICc(gls.18), AICc(gls.19), AICc(gls.20),
 AICc(gls.21), AICc(gls.22), AICc(gls.23), AICc(gls.24), AICc(gls.25),
 AICc(gls.26), AICc(gls.27), AICc(gls.28), AICc(gls.29), AICc(gls.30),
 AICc(gls.31)))

AICcscores <- (c(AICc(gls.null), AICc(gls.1), AICc(gls.2), AICc(gls.3),
 AICc(gls.4), AICc(gls.5), AICc(gls.6), AICc(gls.7), AICc(gls.8), AICc(gls.9),
 AICc(gls.10), AICc(gls.11), AICc(gls.12), AICc(gls.13), AICc(gls.14),
 AICc(gls.15), AICc(gls.16), AICc(gls.17), AICc(gls.18), AICc(gls.19),
 AICc(gls.20), AICc(gls.21), AICc(gls.22), AICc(gls.23), AICc(gls.24),
 AICc(gls.25), AICc(gls.26), AICc(gls.27), AICc(gls.28), AICc(gls.29),
 AICc(gls.30), AICc(gls.31)))
AICcscores

# Comparing numerous gls models, and deriving a global p-value using ANOVA - do
 all the above, and then run this:

anova(gls.null, gls.1, gls.2, gls.3, gls.4, gls.5, gls.6, gls.7, gls.8, gls.9,
 gls.10, gls.11, gls.12, gls.13, gls.14, gls.15, gls.16, gls.17, gls.18, gls.19,
 gls.20, gls.21, gls.22, gls.23, gls.24, gls.25, gls.26, gls.27, gls.28, gls.29,
 gls.30, gls.31)
```

<u>GLS outputs for the four completeness metrics, SCM1, SCM2, BCSM1, BCSM2</u>

C, collections; D, diversity; F, formations; P, period; S, sea level; T, geological time

SCM1

	Model	Parameters	df	weighting	AICc	AIC	BIC	logLik	Test	L.Ratio	*p*
gls.null	1	Null	3	0.004357258	211.2081	211.0082	214.2813	-102.50407			
gls.1	2	DCFST	8	0.014812137	208.7609	203.1609	211.8893	-93.58047	1vs2	17.847208	0.0031
gls.2	3	DCFS	7	0.066702616	205.7513	202.0014	209.6386	-94.00067	2vs3	0.840406	0.3593
gls.3	4	DCFT	7	0.010136944	209.5195	205.7695	213.4068	-95.88473			
gls.4	5	DCST	7	0.007953899	210.0045	206.2545	213.8918	-96.12726			
gls.5	6	DFST	7	0.006128252	210.526	206.776	214.4133	-96.38801			
gls.6	7	CFST	7	0.096742117	205.0077	201.2577	208.895	-93.62887			
gls.7	8	DCF	6	0.001753651	213.0284	210.6755	217.2217	-99.33775	7vs8	11.417756	0.0007
gls.8	9	DCS	6	0.028313555	207.4652	205.1122	211.6585	-96.55611			
gls.9	10	DCT	6	0.009410068	209.6683	207.3153	213.8616	-97.65767			
gls.10	11	DFS	6	0.012125569	209.1612	206.8083	213.3545	-97.40413			
gls.11	12	DFT	6	0.020995867	208.0632	205.7102	212.2565	-96.85512			
gls.12	13	DST	6	0.008787701	209.8051	207.4522	213.9984	-97.72609			
gls.13	14	CFS	6	0.361174758	202.3731	200.0202	206.5664	-94.01009			
gls.14	15	CFT	6	0.035809553	206.9954	204.6425	211.1887	-96.32123			
gls.15	16	CST	6	0.041003561	206.7245	204.3716	210.9178	-96.18579			

gls.16	17	FST	6	0.0026257	212.2211	209.8682	216.4144	-98.9341			
gls.17	18	DC	5	0.001770049	213.0098	211.6765	217.1317	-100.83824	17vs18	3.808289	0.051
gls.18	19	DF	5	0.004510065	211.1392	209.8059	215.2611	-99.90294			
gls.19	20	DS	5	0.016808941	208.508	207.1747	212.6299	-98.58734			
gls.20	21	DT	5	0.024531234	207.7519	206.4186	211.8738	-98.2093			
gls.21	22	CF	5	0.005695216	210.6726	209.3392	214.7945	-99.66962			
gls.22	23	CS	5	0.12545194	204.488	203.1547	208.6099	-96.57733			
gls.23	24	CT	5	0.031556485	207.2483	205.9149	211.3702	-97.95747			
gls.24	25	FS	5	0.004196649	211.2833	209.9499	215.4051	-99.97496			
gls.25	26	FT	5	0.009597938	209.6287	208.2954	213.7506	-99.1477			
gls.26	27	ST	5	0.006212352	210.4988	209.1654	214.6206	-99.58271			
gls.27	28	D	4	0.004730862	211.0436	210.412	214.7762	-101.20602	27vs28	3.246619	0.0716
gls.28	29	C	4	0.005544407	210.7263	210.0947	214.4589	-101.04734			
gls.29	30	F	4	0.002673788	212.1848	211.5533	215.9174	-101.77663			
gls.30	31	S	4	0.009573401	209.6339	209.0023	213.3664	-100.50114			
gls.31	32	T	4	0.018313468	208.3366	207.705	212.0692	-99.85249			

SCM2

	Model	Parameters	df	weighting	AICc	AIC	BIC	logLik	Test	L.Ratio	*p*
gls.null	1	Null	3	0.005317176	215.1094	214.9094	218.1825	-104.4547			
gls.1	2	DCFST	8	0.01517206	213.0124	207.4124	216.1407	-95.70618	1vs2	17.497027	0.0036
gls.2	3	DCFS	7	0.094757391	209.3486	205.5986	213.2359	-95.79932	2vs3	0.186272	0.666
gls.3	4	DCFT	7	0.006856961	214.6008	210.8508	218.488	-98.42538			
gls.4	5	DCST	7	0.010009815	213.8441	210.0942	217.7314	-98.04707			
gls.5	6	DFST	7	0.010553567	213.7384	209.9883	217.6257	-97.99418			
gls.6	7	CFST	7	0.059392352	210.2829	206.5329	214.1703	-96.26647			
gls.7	8	DCF	6	0.002773112	216.4113	214.0584	220.6046	-101.0292	7vs8	9.525449	0.002
gls.8	9	DCS	6	0.047452439	210.7318	208.3789	214.9251	-98.18944			
gls.9	10	DCT	6	0.008023506	214.2865	211.9336	218.4798	-99.96679			
gls.10	11	DFS	6	0.034023503	211.3972	209.0442	215.5905	-98.52212			
gls.11	12	DFT	6	0.023618828	212.1272	209.7742	216.3205	-98.88712			
gls.12	13	DST	6	0.015674082	212.9473	210.5943	217.1406	-99.29716			
gls.13	14	CFS	6	0.316987063	206.9336	204.5806	211.1269	-96.29031			
gls.14	15	CFT	6	0.012283827	213.4347	211.0818	217.628	-99.54089			
gls.15	16	CST	6	0.033248903	211.4432	209.0903	215.6365	-98.54515			
gls.16	17	FST	6	0.001636904	217.4657	215.1127	221.659	-101.55636			
gls.17	18	DC	5	0.002846557	216.3591	215.0257	220.4809	-102.51286	17vs18	1.913001	0.1666
gls.18	19	DF	5	0.009648088	213.9178	212.5844	218.0396	-101.29221			
gls.19	20	DS	5	0.045843582	210.8008	209.4675	214.9227	-99.73374			
gls.20	21	DT	5	0.028766753	211.7328	210.3995	215.8547	-100.19975			
gls.21	22	CF	5	0.005700182	214.9703	213.637	219.0922	-101.81847			
gls.22	23	CS	5	0.140236851	208.5646	207.2313	212.6865	-98.61564			
gls.23	24	CT	5	0.015252744	213.0018	211.6684	217.1236	-100.83421			
gls.24	25	FS	5	0.004226907	215.5683	214.235	219.6902	-102.1175			
gls.25	26	FT	5	0.005306446	215.1134	213.7801	219.2353	-101.89005			
gls.26	27	ST	5	0.004335582	215.5176	214.1842	219.6395	-102.09212			
gls.27	28	D	4	0.009430202	213.9634	213.3319	217.696	-102.66593	27vs28	1.147631	0.284
gls.28	29	C	4	0.006150842	214.8181	214.1865	218.5507	-103.09326			

gls.29	30	F	4	0.003218316	216.1136	215.482	219.8461	-103.74099			
gls.30	31	S	4	0.010346642	213.778	213.1464	217.5105	-102.57319			
gls.31	32	T	4	0.010908817	213.6721	213.0406	217.4047	-102.52028			

BSCM1

	Model	Parameters	df	weighting	AICc	AIC	BIC	logLik	Test	L.Ratio	*p*
gls.null	1	Null	3	0.00205049	213.1671	212.9671	216.2403	-103.48356			
gls.1	2	DCFST	8	0.014874985	209.2039	203.6039	212.3323	-93.80196	1vs2	19.363205	0.0016
gls.2	3	DCFS	7	0.066979619	206.1945	202.4445	210.0818	-94.22225	2vs3	0.840585	0.3592
gls.3	4	DCFT	7	0.003873656	211.8949	208.1449	215.7822	-97.07244			
gls.4	5	DCST	7	0.011687763	209.6862	205.9362	213.5735	-95.9681			
gls.5	6	DFST	7	0.003737393	211.9665	208.2165	215.8538	-97.10825			
gls.6	7	CFST	7	0.101955177	205.3542	201.6042	209.2415	-93.80211			
gls.7	8	DCF	6	0.000470934	216.1094	213.7564	220.3027	-100.87821	7vs8	14.152199	0.0002
gls.8	9	DCS	6	0.040717923	207.1899	204.837	211.3833	-96.4185			
gls.9	10	DCT	6	0.005802001	211.0869	208.7339	215.2802	-98.36697			
gls.10	11	DFS	6	0.006876497	210.7471	208.3941	214.9404	-98.19706			
gls.11	12	DFT	6	0.008919196	210.2269	207.8739	214.4202	-97.93696			
gls.12	13	DST	6	0.007842575	210.4841	208.1312	214.6775	-98.0656			
gls.13	14	CFS	6	0.361386364	202.8234	200.4704	207.0167	-94.23522			
gls.14	15	CFT	6	0.015680872	209.0984	206.7455	213.2917	-97.37273			
gls.15	16	CST	6	0.063610305	206.2977	203.9448	210.4911	-95.9724			
gls.16	17	FST	6	0.00217888	213.0457	210.6927	217.239	-99.34636			
gls.17	18	DC	5	0.000643884	215.4838	214.1504	219.6056	-102.07521	17vs18	5.457705	0.0195
gls.18	19	DF	5	0.001211678	214.2193	212.8859	218.3411	-101.44297			
gls.19	20	DS	5	0.013407719	209.4116	208.0783	213.5335	-99.03914			
gls.20	21	DT	5	0.015311977	209.146	207.8127	213.2679	-98.90634			
gls.21	22	CF	5	0.001777576	213.4528	212.1195	217.5747	-101.05972			
gls.22	23	CS	5	0.183956087	204.1739	202.8406	208.2958	-96.42028			
gls.23	24	CT	5	0.021053037	208.5092	207.1759	212.6311	-98.58793			
gls.24	25	FS	5	0.003285167	212.2244	210.8911	216.3463	-100.44556			
gls.25	26	FT	5	0.006186015	210.9587	209.6254	215.0806	-99.81268			
gls.26	27	ST	5	0.006309147	210.9193	209.5859	215.0412	-99.79297			
gls.27	28	D	4	0.001687316	213.557	212.9254	217.2896	-102.46271	27vs28	5.339477	0.0208
gls.28	29	C	4	0.002226167	213.0027	212.3711	216.7353	-102.18557			
gls.29	30	F	4	0.001096723	214.4186	213.787	218.1512	-102.89353			
gls.30	31	S	4	0.008865627	210.2389	209.6073	213.9715	-100.80367			
gls.31	32	T	4	0.014337254	209.2776	208.646	213.0102	-100.32299			

BSCM2

	Model	Parameters	df	weighting	AICc	AIC	BIC	logLik	Test	L.Ratio	*p*
gls.null	1	Null	3	0.0019698	214.9893	214.7893	218.0624	-104.39463			
gls.1	2	DCFST	8	0.01483369	210.9513	205.3513	214.0797	-94.67566	1vs2	19.437937	0.0016
gls.2	3	DCFS	7	0.072393858	207.7809	204.0309	211.6682	-95.01544	2vs3	0.679559	0.4097
gls.3	4	DCFT	7	0.003922767	213.6115	209.8615	217.4988	-97.93076			
gls.4	5	DCST	7	0.015418379	210.874	207.124	214.7613	-96.562			
gls.5	6	DFST	7	0.007149335	212.4111	208.6611	216.2984	-97.33054			
gls.6	7	CFST	7	0.07252097	207.7774	204.0274	211.6647	-95.01368			

gls.7	8	DCF	6	0.000630591	217.2673	214.9144	221.4606	-101.45719	7vs8	12.887007	0.0003
gls.8	9	DCS	6	0.057433848	208.2439	205.8909	212.4372	-96.94546			
gls.9	10	DCT	6	0.007051888	212.4385	210.0856	216.6318	-99.04279			
gls.10	11	DFS	6	0.015081787	210.9181	208.5652	215.1115	-98.2826			
gls.11	12	DFT	6	0.01330659	211.1686	208.8157	215.3619	-98.40783			
gls.12	13	DST	6	0.01592975	210.8087	208.4558	215.0021	-98.2279			
gls.13	14	CFS	6	0.317549738	204.8239	202.4709	209.0172	-95.23546			
gls.14	15	CFT	6	0.008126653	212.1548	209.8019	216.3481	-98.90094			
gls.15	16	CST	6	0.060671888	208.1342	205.7812	212.3275	-96.89061			
gls.16	17	FST	6	0.001234934	215.9231	213.5701	220.1164	-100.78507			
gls.17	18	DC	5	0.000918463	216.5152	215.1819	220.6371	-102.59095	17vs18	3.61175	0.0574
gls.18	19	DF	5	0.002153198	214.8112	213.4779	218.9331	-101.73894			
gls.19	20	DS	5	0.03057123	209.505	208.1717	213.6269	-99.08583			
gls.20	21	DT	5	0.024355197	209.9596	208.6263	214.0815	-99.31315			
gls.21	22	CF	5	0.001544737	215.4754	214.1421	219.5973	-102.07104			
gls.22	23	CS	5	0.208963978	205.6608	204.3275	209.7827	-97.16373			
gls.23	24	CT	5	0.014192993	211.0396	209.7063	215.1615	-99.85315			
gls.24	25	FS	5	0.002313313	214.6678	213.3344	218.7896	-101.66721			
gls.25	26	FT	5	0.003476676	213.853	212.5196	217.9749	-101.25982			
gls.26	27	ST	5	0.004079134	213.5334	212.2	217.6552	-101.10001			
gls.27	28	D	4	0.002972356	214.1664	213.5348	217.899	-102.76742	27vs28	3.334814	0.0678
gls.28	29	C	4	0.002233246	214.7382	214.1066	218.4708	-103.05331			
gls.29	30	F	4	0.000967184	216.4119	215.7803	220.1444	-103.89014			
gls.30	31	S	4	0.00691629	212.4774	211.8458	216.2099	-101.92289			
gls.31	32	T	4	0.009115537	211.9252	211.2936	215.6577	-101.64679			

Appendix 6: R script

```
# Generalized differencing script from G.Lloyd (http://www.graemetlloyd.com/methgd.html)

gen.diff<-function(x,time)
{
	#if(cor.test(time,x)$p.value > 0.05) print("Warning: variables not significantly
correlated, generalised differencing not recommended")
	dt<-x-((lsfit(time,x)$coefficients[2]*time)+lsfit(time,x)$coefficients[1])
	m<-lsfit(dt[1:(length(dt)-1)],dt[2:length(dt)])$coefficients[2]
	gendiffs<-dt[1:(length(dt)-1)]-(dt[2:length(dt)]*m)
	gendiffs
}

# application of gen.diff example

time<-timeseries1$myr
gd_diversity<-gen.diff(timeseries1$diversity,time)

# Correlation function (A.M.D.)

cor = function(x, y, m = c("pearson", "kendall", "spearman"), N) {
		test = cor.test(x, y, method=m)
		adj = p.adjust(test$p.value, "BH", n=N)
		display = c(test$est, test$p.value, adj)
		names(display)[1] = "stat"
		names(display)[2] = "p"
		names(display)[3] = "adj p"
	return(display)
}

## application of cor example

cor(gd_scm1, gd_bsc1, m="spearman", 24)

## application of lm example
lm_scm1 <- lm(log_scm1 ~ log_div + log_coll + log_form + log_sl + log_per)
summary(lm_scm1)

step_scm1 <- step(lm_scm1, direction = "both")
summary(step_scm1)

## weighted function to weight AICc scores of multiple models (courtesy of Graeme Bell)

weighted<-function (aic){
+   aic.wt <- exp(-0.5 * (aic - min(aic)))/sum(exp(-0.5 * (aic - min(aic))))
+   return(aic.wt)
+   }
```