A Time-Series Analysis of Dinosauria Diversity

Jack K. Horner

MAST

P.O. Box 266

Los Alamos NM 87544 USA

Abstract

It is widely held that the dinosaurs were driven to near extinction because of the Chicxulub asteroid collision with the Earth about 65 million years ago (MA). Without doubt, dinosaur diversity in the fossil record after the collision was at most a percent of what it was prior to the collision. But whether the collision was the principal cause of the extinction is more difficult to assess. Here I compute and compare models of the time series of the number of genera of the Dinosauria and the Mollusca in the period 230 MA - 50 MA. The models are agnostic about whether specific events occurred, and in this sense do not require Chicxulub to explain the Dinosaurian diversity collapse.

Keywords: K-T boundary, time-series, dinosaur extinction, Chicxulub, Cretaceous-Paleogene boundary

1.0 Introduction

Approximately 65.5 million years ago, an asteroid hypothesized to have a speed of ~20 km/s and a diameter of ~10 km struck the ocean near present-day Chicxulub, Yucatan, Mexico ([1]-[3],[5]-[6]). The impact hurled molten rock and rock vapor into the atmosphere, creating, at least briefly, a crater nearly 100 km in diameter and 18 km deep ([7]), spawning fires across the planet. Debris from the impact and fires may have reduced sunlight for years.   It is widely held that the largest land animals, including the dinosaurs of the time, were driven to near extinction by the event ([4],[6]).

Without doubt, dinosaur genera diversity (in this paper, defined as the number of genera in a time bin) in the fossil record after the collision is at most a percent of what it was before the collision. Whether the collision was the principal cause of the near extinction of the superorder Dinosauria ([19]; hereafter, "Dinosauria"), however, is problematic, because there is no definitive evidence that the Chicxulub event caused that demise. Less cataclysmic regimes, such as a small decrease in average annual surface temperature, or a persistent reduction in sunlight reaching the Earth's surface (perhaps caused by volcanic ejecta), could have made the planet untenable for the dinosaurs, who may have been highly sensitive to temperature-induced changes in food distribution ([22]). Unfortunately, even these hypotheses are not uniquely determined by the evidence.

A predictive time series model of the Dinosauria genus-abundance data, because it is inherently agnostic about the occurrence of specific causes, would show that Chicxulub is not *necessary* to explain the observed decline in Dinosauria

diversity. Assuming that unknowns affect the observed values in a time series, we posit that such a model is probabilistic. More specifically, an observation x_t is presumed to be a value of some random variable X_t; the time series $\{x_1, x_2, \ldots x_t, \ldots\}$, a single realization of a stochastic process (*i.e.*, a sequence of random variables) $\{X_1, X_2, \ldots, X_t, \ldots\}$. The fundamental assumption of time series modeling is that the value of the series at time t, X_t, depends only on its previous values (deterministic part) and on a random disturbance (stochastic part). Furthermore, if the dependence of X_t on the previous p values is assumed to be linear, we can write ([21], p. 12)

$$X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \ldots + \varphi_p X_{t-p} + D_t \tag{1}$$

where $\{\varphi_1, \varphi_2, \ldots, \varphi_p\}$ are real constants. D_t is the disturbance at time t, and it is usually modeled as a linear combination of zero-mean, uncorrelated random variables or a zero-mean *white noise model* $\{D_t\}$

$$D_t = Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \ldots + \theta_q Z_{t-q} \tag{2}$$

($\{Z_t\}$ is a white noise model with mean 0 and variance σ^2 if and only if $EZ_t = 0$, $E(Z_t)^2 = \sigma^2$ for all t, and $EZ_s Z_t = 0$ if $s \neq t$, where E denotes the expectation.) Z_t is often referred to as the *random error* or *noise* at time t. The constants $\{\varphi_1, \varphi_2, \ldots, \varphi_p\}$ and $\{\theta_1, \theta_2, \ldots, \theta_q\}$ are called *autoregressive (AR) coefficients* and *moving average (MA) coefficients*, respectively.

Equations 1.1 and 1.2 jointly define a zero-mean *autoregressive moving average (ARMA) model* of orders p and q, or ARMA(p, q). If each of $\theta_1, \theta_2, \ldots, \theta_q$ are 0, Eqs. 1.1 and 1.2 define an *autoregressive* model of order p, or *AR(p)*.

The time-series analysis method used in this study assumes that a series of interest arises from a second-order or weak-stationary process. Roughly put, a process $\{X_t\}$ is stationary if its statistical properties do not change over time. (See [14], Chapter 2 or [21], p. 14, for a detailed account of weak stationarity).

2.0 Method

Two data sets were downloaded from *The Paleobiology Database* ([8]) on 3 January 2011 to a Dell Inspiron 545 with an Intel Core2 Quad CPU Q8200 (clocked @ 2.33 GHz) and 8.00 GB RAM, running under the *Windows Vista Home Premium (SP2)* operating environment and the *Mozilla Firefox* v3.6.13 browser, connected by a 1.5 Mbit/s DSL link to the Internet. The query values used to generate these data sets are shown in Figure 1.

> Dinosauria data set, 230 MA - 50 MA ([10]):
> > Output data = occurrence list
> > Output delimiter = comma-delimited text
> > Taxon or taxa to include = Dinosauria
> > Oldest and youngest intervals = 230 - 50
> > Continents = Africa, Antarctica, Asia, Australia, Europe,
> > > Indian Ocean, Oceania, North America, South America

> Mollusca data set, 230 MA-50 MA ([11]):
> > Output data = occurrence list
> > Output delimiter = comma-delimited text
> > Taxon or taxa to include = Mollusca

Oldest and youngest intervals = 230 - 50

Continents = Africa, Antarctica, Asia, Australia, Europe,

Indian Ocean, Oceania, North America, South America

Figure 1. Query values used to generate the data sets used in this study.

The time interval for the data sets, 230 MA - 50 MA, spans the Dinosauria from their nominal first appearance in the fossil record ([19]) until well after the Chicxulub impact. The Mollusca ([20]) were chosen for comparison with the Dinosauria because there is a large body of Mollusca fossil data available, and the Mollusca are plausible as a "control" for Dinosauria analysis. By default, the query shown in Figure 1 returns data at the genus level.

The genus time-range files from each returned data set were saved. These (comma-separated format) files contain, among other things, the beginning (column *bottom of range),* and end (column *top of range*), times at which [8] reports a genus existed, one row per genus. These range files were imported by Microsoft *Excel* 2007 and the rows in the resulting spreadsheet were sorted in decreasing value of *top of range*. This sorting facilitated rapid visual identification of those genera that arose and perished in the sampled interval.

Excel functions were used to export columns *bottom of range* and *top of range* to a Windows text file. These text files were converted to UNIX textfile format using the Cygwin *dos2unix* utility.   The *count_taxa* software ([12]) running under Cygwin environment (itself under Vista) was then used to count the resulting genera ranges into 1 MA bins, on the hardware described above. A genus was assumed to exist in given bin if the midpoint of that bin lay in the closed interval [*bottom of range, top of range*]. The output of *count_taxa* is a genera-abundance time-series that has a uniform distribution of time-differences (in this case, 1 MA) between adjacent points.

The files output by *count_taxa* were imported to *Mathematica* using the *Time Series* add-on package ([9]). Both data sets were inspected to determine that they did not contradict with the hypothesis that the Chicxulub asteroid could have been the cause of the demise of these taxonomic groups.

Stationary linear time-series models of each of the data sets for the interval 230-50 MA were computed in *Mathematica* as follows (the general theory of the analysis can be found in [14], Chapters 2-3; the entire *Mathematica* script used in this study can be obtained as noted in [18]). Graphs of the time-series derived from *count_taxa* were inspected for trending.   Each of the genus time-range data sets was then zero-meaned (i.e., the mean of the series was subtracted from each of the data values in the series). The Hannan-Rissanen method ([14], Section 5.1.4) was applied to the zero-meaned series to obtain six preliminary models of each of the series. The Akaike Information Criterion values for each of the models were then computed. (The AIC is a penalty function that, when minimized, balances the risks of over-, and under-, fitting ([13]; [14], p. 173).) The two models of each series with the lowest AIC values were selected for subsequent analysis. A conditional maximum likelihood estimate ([14], Section 5.2) of the parameters of each of the selected models was then obtained. The AIC values for the resulting refined models were then determined, and the model with the lowest AIC value was selected for further analysis. The residuals with respect to the selected model were then computed, and the correlation function of the residuals was computed and plotted to assess their convergence. Finally, the portmanteau statistic on the set of residuals for each series, and chi-square statistic for the 95th percentile, were determined;   if the chi-square statistic was greater than the portmanteau statistic the model was accepted ([14], pp. 26, 166-167, 352; [15]).

If *no* model were discovered by this method, it would be strong evidence that no stationary linear model of the time series exists, and, by implication, a strong suggestion that a nonlinear or nonstationary event (e.g., a cataclysmic) event is required to explain the Dinosaurian time series.

The portion of the *Mathematica* script used to produce the time-series analysis for the Dinosauria genera-count in this study ([18]) is shown in Figure 2; the script for the Mollusca series is highly similar.

```
Needs["TimeSeries`TimeSeries`"];
dinodata =
 ReadList[ToFileName[{"C:", "cygwin", "species_abundance"},
   "Dino_230_50_counts.txt"], {Number, Number}];
n = Length[dinodata];
μdino = Mean[dinodata[[All, 2]]];
meanzerodino = (# - μdino) & /@ dinodata[[All, 2]];
dinomodels = HannanRissanenEstimate[meanzerodino, 5, 5, 5, 6];
If[Head[#] === ARMAModel, Head[#][Length[#[[1]]], Length[#[[2]]]],
  Head[#][Length[#[[1]]]]] & /@ dinomodels;
AIC[#, 181] & /@ dinomodels;
{arm2} = ConditionalMLEstimate[meanzerodino, #] & /@
  Take[dinomodels, {2}];
AIC[#, 181] & /@ %;
{arm24} =  ConditionalMLEstimate[meanzerodino, #] & /@
  Take[dinomodels, {6}];
AIC[#, 181] & /@ %;
dinores = Residual[meanzerodino, arm2];
dinocorr = CorrelationFunction[dinores, 12];
plotcorr[corr_, opts___] :=
  ListPlot[corr, DataRange -> {0, Length[corr] - 1}, opts];
plotcorr[# & /@ dinocorr, Joined -> True,
  AxesLabel -> {"k", "ρdino(k)"}];
PortmanteauStatistic[dinores, 12];
Quantile[ChiSquareDistribution[11], 0.95];
```

Figure 2. The portion of the *Mathematica* script ([18], responses not shown) used to compute the model for the Dinosauria genera-count time-series used in this study. The analysis script for the Mollusca series is highly similar.

3.0 Results

The *Paleobiology Database* query described in Section 2.0 yielded 5377 occurrences distributed across 986 genus time-ranges for the Dinosauria data set, and 111412 occurrences distributed across 3658 genus time-ranges for the Mollusca data set. The total time to complete the queries and transmit files to the platform described in Section 2.0 was about one minute. Figure 3 shows the number of Dinosauria and Mollusca genera reported in [8], 230-50 MA, 1 MA binning. The shapes of the time series in Figure 3 are remarkably similar: roughly speaking, at any given time, on average the number of Mollusca genera is ~6 times the number of Dinosauria genera. The mean of the number of Dinosauria genera is 88.5; the mean of the number Mollusca genera, 535.1. The ratio of the mean Mollusca, to the mean Dinosauria, genera-count is 6.0.

The data underlying Figure 3 show a sharp decrease in Dinosauria genera-count at the nominal date (65.5 MA) of the Chicxulub impact, appearing to lend weight to the hypothesis that the collision was a major contributor to the demise of

the Dinosauria. But it is less clear that much of a causal connection can be inferred from this correlation. By way of comparison, at ~112 MA, 127 of the ~175 Dinosauria genera reported in [8], or ~70%, vanished, unassociated with any known asteroid collision with Earth. (New genera were being created at about the same rate as those that vanished, so the net genera count, as shown by Figure 3, decreased only ~10%.) This implies there is at least one apparently non-asteroidal genera-extinction rate in the Dinosauria fossil record whose magnitude is comparable to that which occurred at ~65.5 MA. The Dinosauria genus-diversity collapse at 65.5 MA, therefore, could be explained if the relative extinction rate of ~112 MA were in progress and few new Dinosauria genera were being created past 65.5 MA. Environmental changes less cataclysmic than Chicxulub, such as a small decrease in average annual surface temperature, or a decrease in sunlight reaching the Earth's surface (e.g., due to volcanic ejecta), could have catastrophically compromised the viability of the dinosaurs within a few years. Although the plausibility of these alternative hypothesis shows that the Chicxulub-extinction hypothesis is not uniquely determined by the evidence, the alternative hypotheses are themselves variously problematic..
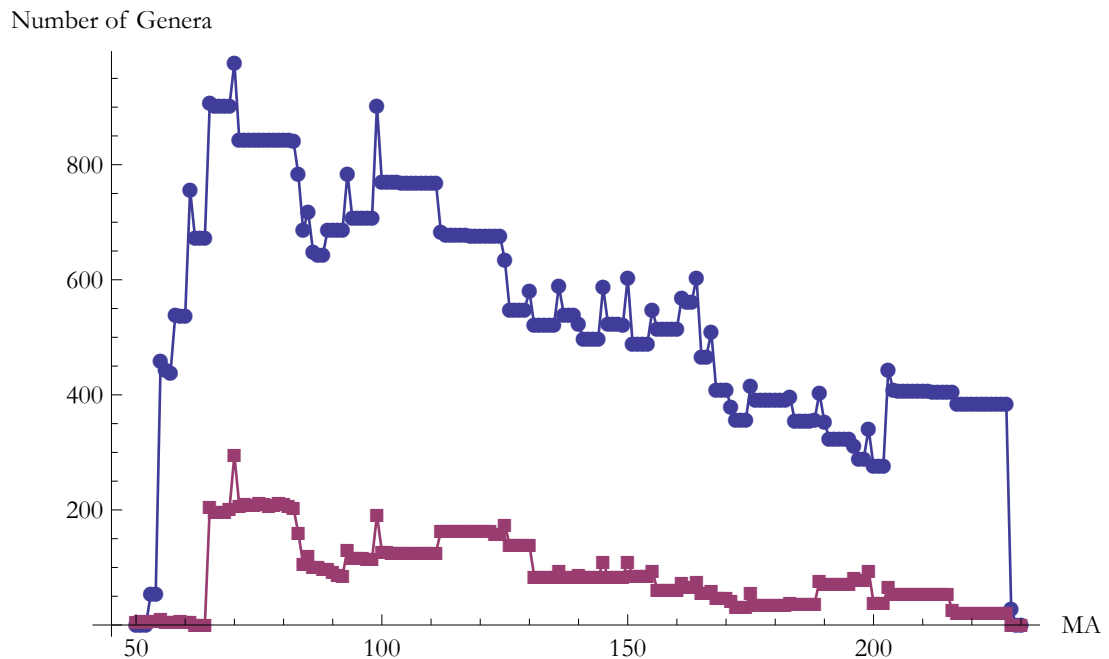

Figure 3. Number of Dinosauria (squares) and Mollusca (circles) genera in [8], 230-50 MA, 1 MA binning. Note the collapse of diversity in both groups beginning at ~65 MA. Note also the shape similarity of these time series. In the figure, time increases from the origin of the Dinosauria (~230 MA) from *right to left*. The figure was generated by the *Mathematica* ([9]) ListLinePlot command.

In any case, a time-series model is agnostic about whether specific events occurred, and such a model can be fitted to the Dinosauria data, showing that the Dinosauria time series does not require Chicxulub per se (nor does it prohibit such an event). In *Mathematica*, the expression

```
ARMAModel[{φ₁, φ₂, …, φₚ}, {θ₁, θ₂, …, θq}, σ²]
```

specifies an ARMA(p, q) model with AR coefficients $\{\varphi_1, \varphi_2, \ldots, \varphi_p\}$, and MA coefficients $\{\theta_1, \theta_2, \ldots, \theta_q\}$, and noise variance σ^2. The *Mathematica* expression

```
ARModel[{φ₁, φ₂, …, φₚ}, σ²]
```

specifies an AR(p) model with AR coefficients $\{\varphi_1, \varphi_2, \ldots, \varphi_p\}$, and noise variance σ^2.

One model for each series passed the portmanteau test of significance described above:

```
Dinosauriamodel(ARM[2]) :
 ARModel[{0.786802, 0.154315}, 8883.25]
```

```
Molluscamodel(ARMA[1, 1]) :
 ARMAModel[{0.963448}, {-0.118255}, 3634.09]
```

4.0 Discussion

Sections 2.0 and 3.0 motivate at least three observations:

1. In this paper, a genus was assumed to exist throughout the interval [*bottom of range, top of range*] reported for that genus in the taxonomic range files described in Section 2.0. More sophisticated existence tests that exploit various abundance weightings (such as *geometric mean abundance* reported in the taxonomic range files) are of course possible. Whether using these weightings would significantly affect the results reported in Section 3.0 will be the subject of future work.

2. The visual similarity of the times series in Figure 3 suggests that the relative creation and extinction rates of Dinosauria, and Mollusca, genera may have a common general dynamic, but the time series analysis models shown in Section 3.0 suggest otherwise.

3. Methods other than the one described in Section 3.0 might produce models that satisfy the portmanteau test of significance described in that section.

5.0 Acknowledgements

----

[1] Baldwin EC, Vocadlo L, and Crawford IA. Influence of target yield stress on crater dimensions: a numerical approach based on Chicxulub. *Lunar and Planetary   Science Conference* XXXVII (2006). Abstract 1887.

[2] Gisler G et al. Two- and three-dimensional simulations of asteroid ocean impacts. *Science of Tsunami Hazards* 21 (2003), 119-134.

[3] Morgan JV et al. Peak ring formation in large impact craters. *Earth and Planetary Science Letters* 183 (2000), 347-354.

[4] Alvarez LW et al. Extraterrestrial cause for the Cretaceous-Tertiary extinction. *Science* 208 (1980), 1095–1108.

[5] Ward W et al. Yucatan subsurface stratigraphy: implications and constraints for the Chicxulub impact. *Geology* 23 (1995), 873-876.

[6] Schulte P et al. The Chicxulub asteroid impact and mass extinction at the Cretaceous-Paleogene boundary. *Science* 327 (5 March 2010), 1214-1218.

[7] Horner JK. Sensitivity of maximum crater depth and diameter to bolide density in a Chicxulub-like asteroid collision. Proceedings of the 2007 International   Conference on Scientific Computing, 200-208. CSREA Press. 2007.

[8] *The Paleobiology Database.* http://paleodb.org/cgi-bin/bridge.pl. 2010.

[9] Wolfram Research. *Mathematica Home Edition* v7.0 (2010) with the *Time Series* v1.4 add-on package (2007).

[10] The major contributors to this data set were Carrano M, Alroy J, Butler R, and Uhen M.

[11] The major contributors to this data set were Kiessling W, Hendy A, Carrano M, Alroy J, Aberhan M, Villier L, Harries P, Ivany L, Fursich F, Miller A,   Patzkowsky M, Bottjer D, Pálfry J, McGowan A, Butler R, Wilf P, and Marshall C.

[12] Horner JK. *count_taxa*, a perl program to count taxa in a taxonomic range file. Available from the author on request.

[15] Ljung GM and Box GEP. On a measure of lack of fit in time series models. *Biometrika* 65 (1978), 297–303.

[18] Horner JK. *Mathematica* notebook *Dino_Mollusca_230_50_time_series.nb*, available from the author on request. 2011.

[19] Weishampel DB, Dodson P, and Osmólska H, eds. *The Dinosauria: Second Edition.* University of California. 2004.

[20] Ponder W and Lindberg DR, eds. *Phylogeny and Evolution of the Mollusca.* University of California. 2008.

[21] Wolfram Research. *Mathematica Time Series* v1.4 manual. 2007. http://media.wolfram.com/documents/TimeSeriesDocumentation.pdf.

[22] Amstrup SC et al. (15 December 2010), 955-958. doi:10.1038/nature09653.