







Role of Taxa Age and Geologic Range: Survival Analysis of Marine Biota over the Last 538 Million Years

Lilian B. PÉREZ-SOSA , Miguel NAKAMURA , Pablo Del MONTE-LUNA , and Alba VICENTE 

Understanding past extinction processes is an important and current matter, and regression methods in the domain of survival analysis can be applied to fossil data. The chief goal of this study is to relate genus-level lifetimes to ancient environmental conditions (e.g., sea level and surface temperatures or carbon dioxide concentration). The Cox proportional hazards model was used to characterize the extinction risks of marine genera, enabling quantitative and graphical comparisons between selected taxonomic groups over geologic time. The environment was confirmed to be a significant factor, and both age-dependent coefficients and time-dependent covariates were required to accommodate the data. The Cox model analyzed in this study provides new insights into the extinction patterns of marine biota over the past 538.8 million years (Ma). Previous works focusing on specific groups at certain geologic times suggest that the extinction risk is related to taxa age, taxonomic group, and, in some particular cases, geologic time. Here, this idea is confirmed for a larger taxonomic group, during a larger time span (ca. 530 Ma), providing a more complete and complex scenario than previous works. After applying survival analyses, conjoint effects were observed between taxa and their age, their time of existence, and the prevailing environmental conditions found at any particular moment.

Key Words: Age-dependent extinction risk; Cox model; Hazard; Sepkoski compendium.

1. INTRODUCTION

The retrospective analysis of extinction, particularly mass extinction, is analogous to how current worldwide environmental changes are reshaping the evolution of life on Earth. To

L. B. Pérez-Sosa · M. Nakamura (✉) Centro de Investigación en Matemáticas A.C. (CIMAT), Callejón Jalisco s/n, Col. Valenciana, 36023 Guanajuato, Gto., Mexico (E-mail: nakamura@cimat.mx). P. Del Monte-Luna · A. Vicente Instituto Politécnico Nacional (IPN), Av. Instituto Politécnico Nacional s/n, Col. Playa Palo de Santa Rita, 23096 La Paz, B.C.S., Mexico. A. Vicente Departament de Dinàmica de la Terra i de l'Oceà, Facultat de Ciències de la Terra, Universitat de Barcelona-UB, 08028 Barcelona, Catalonia, Spain.

© 2023 The Author(s)

Journal of Agricultural, Biological, and Environmental Statistics
<https://doi.org/10.1007/s13253-023-00547-0>

properly understand the potential of the modeling approach to be subsequently developed here, some biological and geologic concepts should first be outlined. In ecology, the term extinction refers to the complete disappearance—the death of the last individual—of a species (IUCN 2019). Although there are some cases where humans may have been involved in ancient extinction (Burney and Flannery 2005; Sandom et al. 2014), the vast majority of them (99%) occurred prior to human appearance, and fossil and geologic records are the only data available to study them. In paleobiology, extinctions are recognized when selected taxa (e.g., species, genera, or class) no longer appear in the fossil record, and mass extinctions are said to occur when over 60% genera are lost in an average lapse of 2 million years (My) (Barnosky et al. 2011).

What exactly triggers extinction is difficult to determine. In modern times (1500 onward, Flemming and MacPhee 1999), overexploitation, environmental stress (e.g., climate change), and habitat modification and loss are the main precursors of extinction. Although similar biotic and abiotic covariates could have triggered past extinctions, determining the relationship between causes and effects using fossil evidence is far more complicated. Climate change has been assumed to be the main cause of large extinctions. However, important environmental factors, such as sea level change, which has an effect on onshore–offshore depth (Alroy 2010b; Alroy et al. 2008; Harrison 2002), temperature, CO₂ concentration, enhanced volcanic activity, extraterrestrial impacts, ocean acidification, and even some astronomical cycles, should also be considered (Rohde and Muller 2005; Bond and Grasby 2017). Recent studies have suggested that biotic interactions also play a key role in the evolution of ecosystems, although they are still difficult to characterize (Fraser et al. 2021; Spiridonov and Lovejoy 2022). Hence, the idea of a single-cause model for explaining past extinctions is dismissed, but a multifactorial event that simultaneously affected different ecosystems in multiple ways is considered the main cause of past biodiversity crises, including major worldwide events (Bond and Grasby 2017).

Because fossil data, the most important source of past biota information, have been widely used and scrutinized, they are generally susceptible to various problems (e.g., Raup 1972; Alroy 2010a; Chao and Jost 2012; Payne and Heim 2020). Fossilization itself is a manifest haphazard phenomenon that depends on the kind of organism (those with hard parts are more likely to fossilize), habitat (which could present differences in preservation potential), and the conjunction of complex sedimentary processes that facilitate fossilization. For that reason, the fossil record is considered incomplete and, in many cases, biased. One of the most remarkable biases observed in the fossil record is the sampling bias on account that not all stratigraphic levels have been homogeneously sampled for several reasons (e.g., sampling availability). Other important biases are the *Signor–Lipps* effect, which highlights the idea that the start and end dates of each genus range (represented by its first and last organism) may not have been preserved (Signor et al. 1982), and the *Pull of the Recent* bias, based on the greater fossilization probability (overestimation) toward the present.

Even though the fossil record is indisputably incomplete (Benton 2009), and despite the biases still linked to it, it has been demonstrated that geologic factors do not affect the evolutionary signal of this record (Benton 1995 and references therein). Moreover, some of these biases have already been proven false, as occurs with the *Pull of the Recent* megabias, which only affects 5% of the Neogene bivalves, and similar percentages have been observed

in other groups, including tetrapods (Benton 2016). This means that even considering these biases, the fossil record still provides valuable data that enable the study of past ecosystems. In fact, many paleodiversity studies employ taxa richness data—with or without corrections to the methodological biases of paleobiological databases (Alroy 2010a)—to detect sudden changes in biodiversity. Similar studies also enable the observation of important evolutionary trends such as the decline in the background extinction pattern from the early Paleozoic to the present (Raup and Sepkoski 1982) or the recognition of the well-known *Big Five* mass extinctions.

Nevertheless, analysis of the fossil record still poses challenges, exemplified by the contrasting results obtained on analyses on the dynamics of past biotic crises on Earth. Van Valen (1973) suggested that extinctions remained constant through time and taxa longevity, whereas (Hagen et al. 2018 and references therein) showed that extinctions are indeed related to taxa age.

In the present study, the marine fossil record was correlated to ancient ecosystems during the *Phanerozoic Eon* (from 538.8 ± 0.2 million years ago (Ma) to the present) by considering abiotic (environmental) and biotic variables, as previously reported (e.g., Fan et al. 2020; Song et al. 2021). Aiming to characterize the relationship between ancient life and past conditions, a sensible statistical model considering the interpretable components was developed. This statistical formulation was based on survival analysis, which enables the study of *lifetimes*—more generally, time until the occurrence of a landmark event—by examining distributional differences as a function of covariates. It is used in many health and industrial reliability contexts; however, it can also be applied to characterize taxa longevity.

Namely, the Cox regression model, which considers the species or genera's lifetimes (in My), multiple *covariates* (e.g., sea level and group membership) that are associated with their *range* (i.e., the interval of geologic time over which a species existed), and *censored observations* (i.e., extant species that appear in the fossil record but have not yet become extinct), was adopted. Exploiting concepts and models originating in survival analysis allowed for an alternative viewpoint, enabling probabilistic statements regarding lifetimes, a visualization of the progression of survival in time as related to covariates in time, group comparisons, and uncertainty quantification. The results obtained by the Cox analysis are helpful to contextualize the current biodiversity crisis by building a detailed past analog as well as testing the feasibility and full potential of *survival analysis* to study fossil data.

The use of Cox regression in the study of extinction is not entirely new. A few studies have used this model in survival analysis to examine past (Doran et al. 2006; Abelson 2019) and extant (Drake 2006) biotas. Previous approaches considered covariates that do not vary in geologic time and considered specific taxon in a limited time span, such as Cenozoic carnivorous (Abelson 2019) or Mesozoic to Cenozoic planktonic foraminifera (Doran et al. 2006). However, we aimed to apply the Cox model to a diverse and representative group of marine biotas during the Phanerozoic considering multiple, time-varying covariates that represent conditions that genera were subjected to during their lifetime. In addition, non-homogeneous relationships with genera's ages were allowed. This innovative formulation resulted in a more elaborate and complex Cox model that enables the examination of lifetimes in two dimensions: geologic time and taxon age.

2. CONCEPTS AND DATA

2.1. FOSSIL DATA

The first accessible paleontological database was created by John Sepkoski and collaborators during the 1980s and 1990s. These authors first presented a compendium of marine families (Sepkoski 1982, 1992) that, after two decades, resulted in a genus-level marine compendium (Sepkoski 2002) built using secondary data (bibliographic data) of marine genera and their relative chronostratigraphic ages. Both compendia have been extensively used in previous studies. Some preliminary works applied time series analysis to search for harmonic components or cycles in biodiversity curves of genus richness (Raup and Sepkoski 1984). Although many researchers have focused on assessing the processes of diversification (origination) and extinctions of marine metazoans (Etter 2007 and references therein), the above compendium also gave rise to studies of the anatomical and ecological constraints of biota (Bambach et al. 2002) as well as evaluations of macroevolutionary processes (Jablonski 2005).

After these first attempts, new statistical methods were developed to reduce the sampling bias of the fossil record and to discuss how to correct biodiversity curves (Alroy et al. 2001; Alroy 2010a; Chao and Jost 2012). These approaches, along with the need to study more detailed directories (i.e., species level), resulted in the proposal of other databases such as the Paleobiology Database (PaleoDB, <https://paleobiodb.org>). This paleontological directory was created in the late 1990s and has been maintained by a number of international paleontologists. Although the Sepkoski compendium and PaleoDB are the preferred databases in many analytical studies, important institutions such as the International Commission on Stratigraphy (ICS, <https://stratigraphy.org>) have proposed their own directories, i.e., the Geodiversity Database (GBDB, <https://geobiodiversity.com>).

Here, the Sepkoski compendium (Sepkoski 2002) was used. It contains over 36,000 entries of genus-level marine fossil data and provides the first and last appearance datum of each genus (FAD and LAD, respectively). A calibration with absolute ages (in Ma) was later included in the compendium and is based mainly on the chronostratigraphic data from Ogg et al. (2004). These data allow for inferring genera *longevity* as well as building richness distribution graphs of the marine Phanerozoic biota (Fig. 1). Although it has been extensively used, this database still contains many flaws: Taxonomic mistakes are common, *singleton* taxa (genera whose FAD and LAD occur in the same temporal interval) are abundant (+15, 797 cases), and chronostratigraphic ages are outdated. These considerations make rigorous statistical modeling challenging since inference on the status of past life must be addressed with probabilistic approaches with a myriad of sources of variation and uncertainty.

The data bits used are the FAD and LAD of each genus, which is represented in the analysis as T_F and T_L , respectively. The interval T_F to T_L is the range, and the lifetime of a particular genus (i.e., duration, longevity, or age at extinction) is calculated by $L = T_F - T_L$. The symbol t is used to indicate the subject's age (in My), while T refers to the date of a rock or the geologic time (in Ma). Singleton taxa provide valuable data in particular cases (Fitzgerald and Carlson 2006), but to avoid misinterpretations, they were

ROLE OF TAXA AGE AND GEOLOGIC RANGE

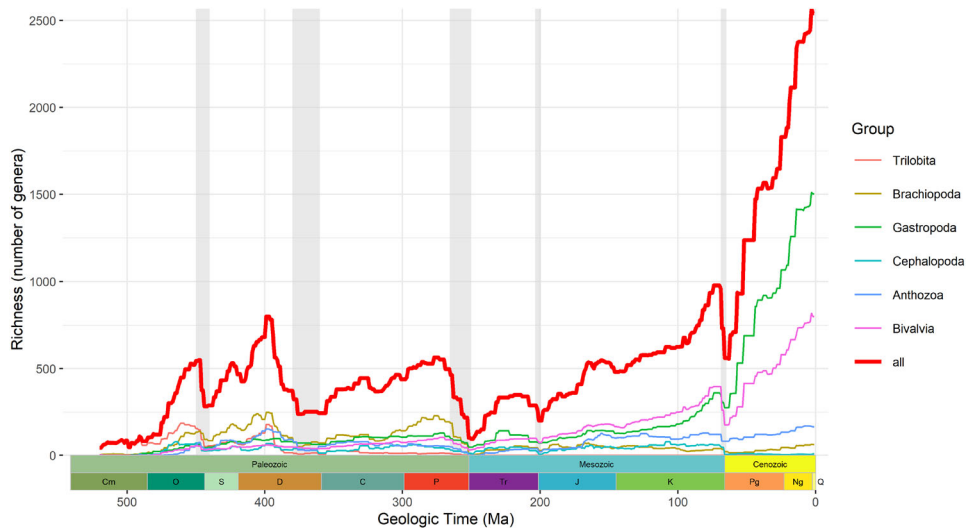


Figure 1. Genera richness (total and by group) during the Phanerozoic Eon. “All” sums the total number of the six groups considered. The shaded time intervals delimit the periods of the five recognized mass extinctions. The bottom strip contains nomenclature for geologic eras and periods. Cm, Cambrian; O, Ordovician; S, Silurian; D, Devonian; C, Carboniferous; P, Permian; Tr, Triassic; J, Jurassic; K, Cretaceous; Pg, Paleogene; Ng, Neogene; Q, Quaternary.

ignored, as is customary in recent analyses (e.g., [Casanovas-Vilar et al. 2010](#)). Extant genera that have unknown (unrealized) extinction data were flagged as censored observations and were retained. Finally, some of the biases described and commented on in this work were disregarded at the outset given that the essential concern was to establish regression methods to incorporate covariates in the analysis of the fossil record.

2.2. COVARIATES RELEVANT TO EXTINCTION

As mentioned above, a number of environmental factors are thought to have influenced past extinctions in the marine realm. Here, only a few abiotic and biotic covariates are applied in the study and are described below (Fig. 2).

Sea levels. The sea levels were determined from a variety of data sources ([Harrison 2002](#)).

The long-range and short-range variations were studied using frequency spectra, and relationships were obtained based on the continental phenomena, temperature changes, and ice age cycles. The sea level curve for the past 590 Ma is known as the Exxon/Haq curve ([Haq et al. 1987](#)), which was illustrated and commented on by [Harrison \(2002\)](#) and is considered here as one of the time-varying covariates that is denoted by $s(T)$.

Sea surface temperature (SST) and CO₂. Temperature and the concentration of carbon dioxide (CO₂) are well-known indicators of past climates on the geologic scale, and their effect on the Earth’s biota has been extensively studied ([Hautmann 2012](#); [Song et al. 2021](#)). These works established that abrupt sea surface temperature changes have strongly affected ancient marine biota. Thus, the covariate SSTdiff (SST difference),

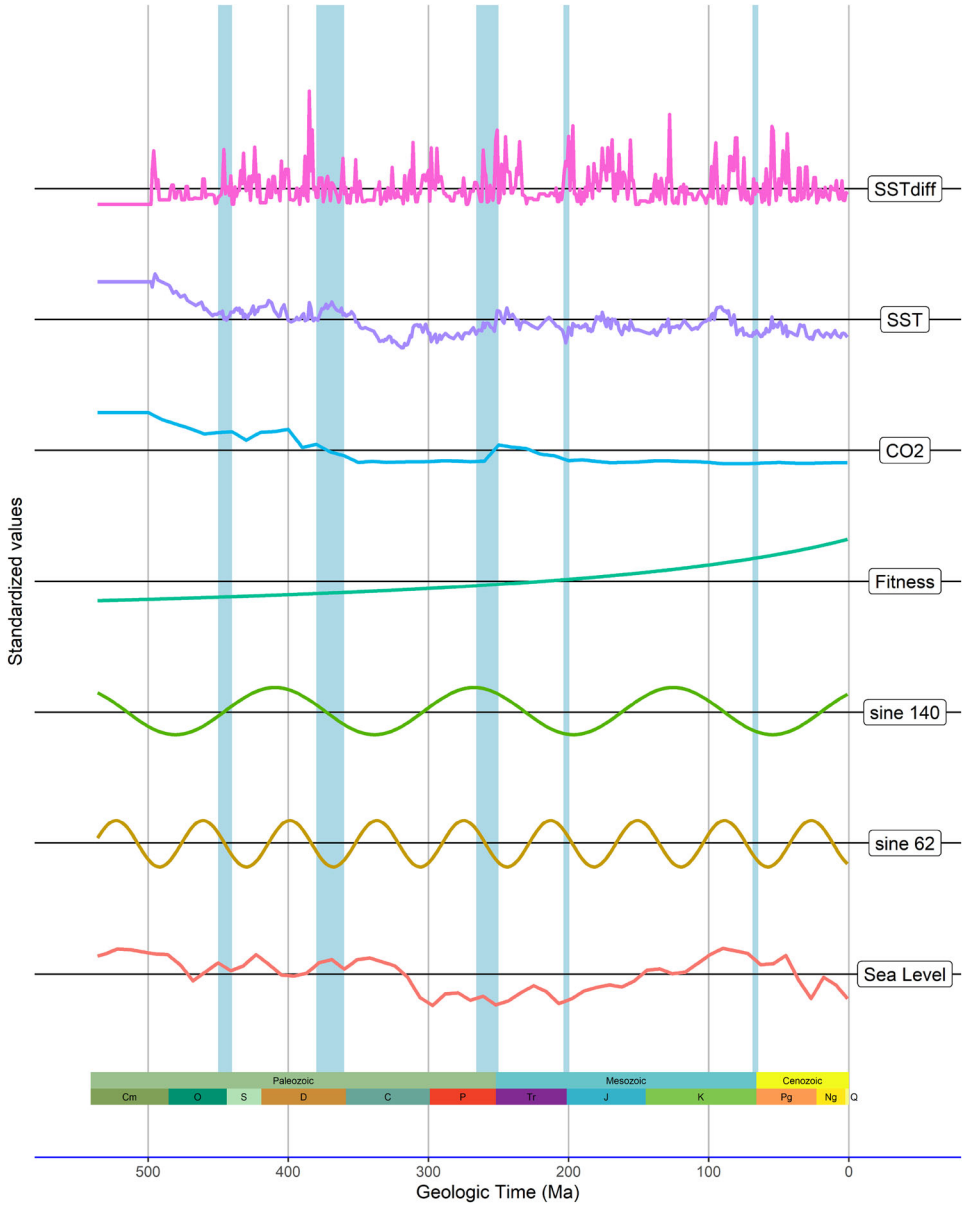


Figure 2. Set of time-varying covariates used in the present study. All variables have been standardized for convenience. Shaded areas over geologic time represent the spans of the five recognized mass extinctions. Abbreviations for geologic eras and periods are given in the caption of Fig. 1.

denoted by $d(T)$, is considered here as the difference in SST between geologic times T and $T - 1$. The covariate CO_2 is intentionally omitted in the models to be described below because of its high correlation with SST.

Sinusoid cycles. Time series analysis of the Sepkoski compendium showed that since the Cambrian Period, two significant cyclic components with periodicities of 62 and 140

My absorbed most of the observed variation in the detrended biodiversity curves of the marine fossil record (Rohde and Muller 2005). Although the origin of these cycles remains speculative, current hypotheses for explaining the 62-My cycle suggest that it correlates particularly well with mass extinctions. Diverse environmental processes may be involved, including changes in the continental margins, sedimentation, and internal rhythmic oscillations of the marine biota. In contrast, the 140-My cycle has been associated with astronomical events occurring on the order of every 10^8 years, particularly the passing of the solar system through the spiral arms of the Milky Way Galaxy, which may have increased the flux of comets and cosmic rays directed at Earth, thereby influencing extinction events (Raup 1986; Raup and Sepkoski 1984; Rohde and Muller 2005). These cycles are denoted in the analysis by $p_{62}(T)$ and $p_{140}(T)$, which have an amplitude of 20 and phases of -20 and 40 , respectively.

Evolutionary fitness. The concept of the “general optimization of fitness” (herein referred to as fitness) is understood as an increasing trend in the performance of the evolutionary process, i.e., evolution becomes progressively more effective. This idea was first introduced by Raup and Sepkoski (1982), but it has rarely been addressed in the literature. The notion of increasing fitness is based on two observations: (1) a negative trend in the total extinction rate of Earth’s biota from the Cambrian Period to the present and (2) a net increase in the total diversity during the same time lapse. The combination of these factors would lead to overall prolonged survival. Since no theoretical or functional mathematical form has been found in the literature to quantitatively describe this concept, the form $e(T) = (T + 250)^{-1}$ was adopted. Although fitness was conceived as an increasing feature, our choice of a specific function solely aimed to illustrate the methodology.

Taxonomic groups. It has been well established that taxa do not behave homogeneously in regard to extinction (e.g., Alroy 2010c; Marshall 2010), suggesting that all analyses should consider different taxonomic groups when possible. Here, six high taxonomic groups are considered: Phylum Brachiopoda (clam-like marine invertebrates) and classes Trilobita (marine arthropods), Gastropoda (snails), Cephalopoda (e.g., nautilus), Bivalvia (true clams), and Anthozoa (corals). Similar groups were chosen in a previous analysis (e.g., Alroy et al. 2001; Alroy 2010c), since they represent the vast majority of the compendium (more than 50% of the genera) but also because they are well-known and well-represented groups in the fossil record. The remaining entries constitute a seventh group, considered the reference group, allowing the inclusion of the whole compendium when estimating model parameters. If records belong to one of G groups and a reference group, the usual representation based on G dummy indicator variables g_1, \dots, g_G is used: All dummies were set to 0 for the reference group, whereas $g_i = 1$ if and only if a record belongs to the i th group of interest. The relationship (Fig. 1) between the number of records—the richness or biodiversity—per group of interest and of all the genera together during the Phanerozoic Eon showed an increasing trend in genera richness of marine biota in all six groups. This finding portrays the notion of an exponentially increasing trend of biodiversity established by many studies (Benton 2016 and references therein). However, a decline in biodiversity was observed during the five major extinction events.

2.3. DATA AVAILABILITY

The compendium of fossil data used in this study was published as Sepkoski (2002), which includes a CD-ROM (4 3/4 in.). Also, see <http://strata.geology.wisc.edu/jack/>. The sea level data were a smoothed version of Haq et al. (1987), Harrison (2002) with a resolution of 1 Ma. The sea surface temperatures and carbon dioxide concentrations were found in the online supplementary materials of Song et al. (2019). The data used, as well as the R script for the fitting of the Cox model, diagnostic residual analyses, and production of plots, can be found at <https://github.com/lpsosa/Fossils>.

3. FORMULATION AND RESULTS

3.1. APPROACH BASED ON SURVIVAL ANALYSIS

It was observed that the Sepkoski compendium provides lifetime data—possibly censored—since each genus is associated with FAD and LAD. Letting t represent the taxon age (My) and T the geologic time (Ma), the genus lifetime is represented by L with a distribution function of $F(t)$ and density of $f(t)$. The *survival* function of a subject in a particular time t is defined as $S(t) = P(L \geq t)$, while the *hazard* function, a convenient characterization for an individual's survival, is defined as the function

$$h(t) = \lim_{\delta t \rightarrow 0} \left\{ \frac{P(t \leq L < t + \delta t \mid L \geq t)}{\delta t} \right\} = -\frac{S'(t)}{S(t)} = \frac{f(t)}{S(t)}.$$

The hazard function is related to the instantaneous death rate or imminent extinction of a subject who has survived up to age t . An approximation for a small δt is given by

$$P(t \leq L < t + \delta t \mid L \geq t) \approx \delta t \times h(t). \quad (1)$$

One minus this last probability (for an implicitly determined value of δt), expressed as a proportion, has been described as *survivorship* in some paleobiological literature (Raup 1986; Hammer and Harper 2006). The terms survival and hazard are recognized in other paleobiological studies (e.g., Drake 2006; Doran et al. 2006) since these are also formulated in the context of survival analysis, albeit with a different approach than the one we adopted.

3.2. EXPLORATORY ANALYSIS

This work aimed to understand how survival functions (or equivalently, hazard functions) vary as geologic time T progresses. One approach to address this issue was to select individuals who represent the situation at a fixed time T and then look at survival times for that subset of individuals. This strategy only applies in this section. Different counting methods and criteria for choosing a subset of individuals at time T have been discussed (Alroy 2010a; Raup 1978). The *range through* criterion was used, giving rise to the subset of cases for which T is covered by the subject's range, a subset termed *risk set* in Doran

et al. (2006). The total number of individuals for each T coincides with the genera richness (Fig. 1).

Once the individuals in the risk set have been determined, the Kaplan–Meier standard, a nonparametric estimator, was applied to assess the survival function of the aggregated species at T . For this preliminary analysis, no grouping was applied, and no covariates were considered. Censoring occurred and was appropriately taken into account by the Kaplan–Meier estimate, especially when T approached the present day. Raup (1978) applied this notion of sub-setting data by taking subsets at time T to observe changes over time but discretizing over stratigraphic periods. In contrast, the Kaplan–Meier approach considers a continuous timescale devoid of stratigraphic naming conventions and explicitly handles censored lifetimes, which is not apparent in the treatment of survivorship in Raup (1986) and Hammer and Harper (2006).

The results obtained by applying the Kaplan–Meier estimator confirm an increasing survival trend from past (538.8 ± 0.2 Ma) to present (0 Ma) biota, suggesting *time-dependent* survival (Fig. 3). During certain geologic times (e.g., $T = 150$ Ma), the survival values vary depending on the taxon’s age t , as young taxa (e.g., 0–50 My old) are more likely to survive than older genera (e.g., 200–300 My old), also suggesting an *age-dependent* survival of the marine metazoans. It is also worth noting that an increase in general survival is apparent immediately after a mass extinction. Some authors considered that the rapid aftermath diversification of the surviving biota represents a well-known macroevolution process (e.g., Jablonski 2005; Hull 2015). Figure 3 is in concordance with the notion of evolutionary fitness described in the bibliography (Raup and Sepkoski 1982). However, that agreement pertains to a rough or low-resolution relationship, and it will be interesting to investigate whether finer details can be perceived, especially by associating them with past environmental conditions.

3.3. COX REGRESSION

The purpose of the Cox model (see Collett 2015) is to evaluate the effect of various risk factors on survival, $S(t)$, via its hazard function, $h(t)$. Allowing the incorporation of covariates, the Cox analysis presents more detailed hazard results than the Kaplan–Meier analysis. First, a straightforward association between a lifetime and covariates could be as simple as the time of its first appearance. Then, if a lifetime L corresponds to a record having value T_F and belongs to a certain group, the covariates in the model would be $s(T_F)$, $d(T_F)$, $p_{62}(T_F)$, $p_{140}(T_F)$, $e(T_F)$, and g_1, \dots, g_6 concatenated into a vector \mathbf{X} of length 11. A function $h_0(t)$, called the baseline hazard function, represents the hazard function when the covariates all take on a value of zero. The Cox model in its most elementary form postulates the hazard function for L to be proportional to $h_0(t)$ and is given by

$$h(t, \mathbf{X}) = h_0(t) \exp \left(\beta_1 s(T_F) + \beta_2 p_{62}(T_F) + \beta_3 p_{140}(T_F) + \beta_4 e(T_F) + \beta_5 d(T_F) + \sum_{i=1}^6 \beta_{5+i} g_i \right).$$

The term within the exponential function that is positively related to hazard is called the *linear predictor*. In this model, no assumptions are made about the specific form of the

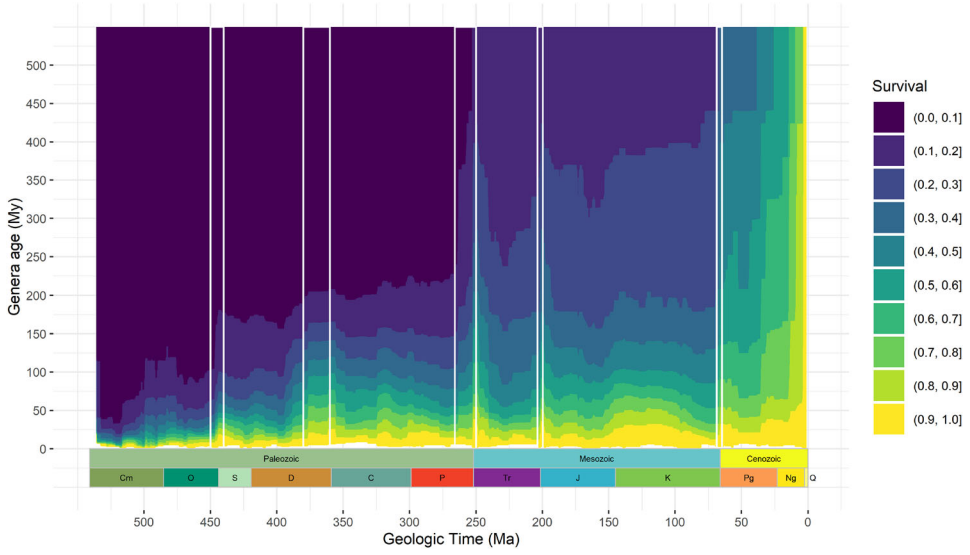


Figure 3. Kaplan–Meier estimates of survival evolving over geologic time. The vertical axis represents the genera age (from younger to older), while the horizontal axis corresponds to the geologic time (from past to present). Abbreviations are given in the caption of Fig. 1. Vertical lines delimit the spans of the five known mass extinctions.

base hazard function $h_0(t)$; hence, the estimation of the regression parameters is enabled using the notion of partial likelihood. The values of parameters β_k regulate the impact of the covariates on the hazard function: A positive value of the coefficient β_k would imply a negative impact on the hazard function since increasing the value of the associated covariate would increase the hazard and, conversely, a negative value would show a decrease in hazard, increasing the probability of survival.

A somewhat more elaborate system of covariates (time-varying) is postulated if instead of considering $s(T_F)$, i.e., a single value of sea level to be associated with a given individual, one considers the function $s(T)$ during the whole of the genus's exposure to sea level during its range. For example, the sea level covariate related to a genus that first appeared at geologic time T_F and has lived for t years is the function $s'(t) = s(T_F - t)$. A generalization of the Cox model may be written as

$$h(t, \mathbf{X}^t) = h_0(t) \exp \left(\beta_1 s'(t) + \beta_2 p'_{62}(t) + \beta_3 p'_{140}(t) + \beta_4 e'(t) + \beta_5 d'(t) + \sum_{i=1}^6 \beta_{5+i} g_i \right).$$

A simple way to encode time-dependent covariates is to consider subintervals (Therneau and Grambsch 2000) with a length of 1 My. The general idea is to extend each original individual record to cover all the 1-My periods during a genus's range. Multiple records are created for each genus, one each to contain a vector holding the current values of the covariates at each 1-My period. For parameter estimation in the model, we rely on the survival (Therneau 2021) package within R (R Core Team 2021). The function `coxph` in this package defines a parameter (`cluster`) that recognizes this extended data format so that records originating from a common genus are treated appropriately.

A more elaborate formulation of the Cox regression, in which its *parameters* are also allowed to be functions of t , can also be considered. The model allowing for time-varying covariates and coefficients is

$$h(t, \mathbf{X}') = h_0(t) \exp \left(\beta_1(t)s'(t) + \beta_2(t)p'_{62}(t) + \beta_3(t)p'_{140}(t) + \beta_4(t)e'(t) + \beta_5(t)d'(t) + \sum_{i=1}^6 \beta_{5+i}(t)g_i \right). \quad (2)$$

The interpretation of the time-varying parameters is that the corresponding covariate has differing effects on the hazard depending on the genus age. Time-dependent parameters are functions, so parametrizations or simplifications incorporating piecewise constant functions of t are the first feasible approach. The varying heights of these functions constitute separate parameter values for each covariate. Graphical exploratory techniques and statistical tests (Therneau and Grambsch 2000) were applied to suggest whether time-dependent parameters are required and over what ranges of t values a parameter $\beta_k(t)$ may be approximated by a piecewise constant function. The final results of this first-stage analysis are implicitly shown in Table 1.

The computer implementation of these piecewise constant parameters involves appending a column to extended data coding specific age intervals that correspond to constant parameter values. The function `survSplit` is designed to split the extended dataset into time-dependent parts in which there is some variation in the coefficients. For parameter estimation, the function `coxph` must be invoked with a special syntax (involving a colon in the model definition) to specify stratifying by age intervals (Therneau and Grambsch 2000).

A second series of techniques based on residual analyses and graphical procedures (Chapter 4 Collett 2015) was also applied to examine the adequacy of Cox proportional hazard models and to evaluate the nested levels of complexity that have been previously described. The conclusion, by examining the results based on deviance and the so-called martingale and the Cox–Snell residuals, the (2) model is adequate in reflecting the intricacies found in the data: Genera lifetimes were related to both time-dependent covariates and parameters.

Parameter estimation is achieved by partial likelihood, and an important point to be made is that in this formulation, no data subsets were selected for estimation purposes at each T . Instead, all data points were involved together in a likelihood function by relationships with covariates, which depend on T . Parameters (Table 1) with large standard errors relative to their estimates were not significantly nonzero (i.e., a confidence interval is prone to include 0). However, the following general interpretations hold true. Only the sinusoid of wavelength 62 and SSTdiff appeared to have age-independent effects on hazard. The remaining parameters, including all parameters identified with group membership, described effects that vary with age to differing degrees.

In some cases, sign reversals occur, meaning that for one age range, hazard increases, but for another, it decreases. The estimates of $\exp(\beta)$ are multiplicative constants that scale hazard per unit covariate. For example, for all groups, increasing the sea level by one unit decreased hazard upon multiplying by 0.998 for very old genera (> 72 My) but increases hazard for younger genera (multiplying by factors > 1). Similarly, the very

Table 1. Estimates, standard errors, and hazard multipliers per unit covariate for parameters in the Cox model (2) with both time-varying parameters and covariates. A covariate may include a set of ranges for age because a stepwise function was used to approximate the time-varying coefficient $\beta(t)$ for that covariate. The wavelength 62 sinusoid and SSTdiff did not require steps, according to the diagnostic procedures applied. The p -value for the test that a parameter is zero is also shown

Age (t) range	Covariates	Estimated β	Estimated $\exp(\beta)$	Standard error	p -value
(0, 9]	Sea level	0.002	1.002	< 0.001	< 0.001
(9, 26]		0.0006	1.000	< 0.001	0.003
(26, 72]		0.0004	1.000	< 0.001	0.075
(72, 530]		-0.001	0.998	< 0.001	< 0.001
(0, 530]		Sinusoid, wavelength = 62	-0.001	0.998	< 0.001
(0, 6]	Sinusoid, wavelength = 140	0.004	1.004	0.001	< 0.001
(6, 17]		-0.002	0.997	0.001	0.008
(17, 42]		-0.005	0.994	0.001	< 0.001
(42, 530]		0.001	1.001	0.001	0.349
(0, 6]		Fitness	-469	< 0.001	24.467
(6, 17]	-380		< 0.001	20.169	< 0.001
(17, 42]	-441		< 0.001	21.282	< 0.001
(42, 72]	-312		< 0.001	33.773	< 0.001
(72, 530]	-350		< 0.001	39.417	< 0.001
(0, 530]	SSTdiff	0.106	1.112	0.005	< 0.001
(0, 5]	Trilobita	0.943	2.568	0.054	< 0.001
(5, 11]		0.648	1.913	0.065	< 0.001
(11, 20]		0.839	2.314	0.073	< 0.001
(20, 30]		0.809	2.246	0.112	< 0.001
(30, 42]		0.738	2.093	0.164	< 0.001
(42, 530]	Brachiopoda	1.335	3.802	0.157	< 0.001
(0, 5]		0.244	1.276	0.056	< 0.001
(5, 11]		0.557	1.746	0.050	< 0.001
(11, 20]		0.682	1.978	0.054	< 0.001
(20, 30]		0.703	2.020	0.074	< 0.001
(30, 42]	Gastropoda	0.859	2.362	0.089	< 0.001
(42, 530]		0.913	2.494	0.083	< 0.001
(0, 5]		-0.571	0.564	0.083	< 0.001
(5, 11]		-0.435	0.646	0.069	< 0.001
(11, 20]		-0.342	0.709	0.067	< 0.001
(20, 30]	Cephalopoda	-0.050	0.950	0.074	0.498
(30, 42]		-0.143	0.866	0.090	0.110
(42, 530]		0.256	1.292	0.056	< 0.001
(0, 5]		1.325	3.763	0.049	< 0.001
(5, 11]		1.040	2.829	0.059	< 0.001
(11, 20]	Anthozoa	0.796	2.217	0.078	< 0.001
(20, 30]		0.666	1.947	0.113	< 0.001
(30, 42]		0.661	1.937	0.145	< 0.001
(42, 530]		0.966	2.628	0.114	< 0.001
(0, 5]		-0.478	0.619	0.098	< 0.001
(5, 11]	Anthozoa	0.184	1.202	0.070	0.008
(11, 20]		0.369	1.447	0.070	< 0.001
(20, 30]		0.350	1.419	0.090	< 0.001
(30, 42]		0.470	1.601	0.106	< 0.001
(42, 530]		0.576	1.780	0.079	< 0.001

Table 1. (Continued)

(0, 5]	Bivalvia	-0.570	0.565	0.096	< 0.001
(5, 11]		-0.344	0.708	0.077	< 0.001
(11, 20]		-0.060	0.940	0.068	0.375
(20, 30]		-0.044	0.956	0.083	0.592
(30, 42]		0.132	1.142	0.091	0.147
(42, 530]		0.207	1.230	0.061	< 0.001

young genera of Anthozoa underwent less hazard than older genera in the same group. Furthermore, although some parameter arrays were monotonic with age (sea level and Brachiopoda/Bivalvia groups), it was more common for this not to be the case. In some instances, the age-tiered parameters represented low-high-low or high-low-high profiles. Allowing for free parameters using the device based on piecewise constant functions enabled us to discover this finding.

3.4. EXPLORATORY PLOTS BASED ON HAZARD

Once the specific model has been discerned and validated and its parameters have been estimated, a few derived plotting devices may provide discovery and understanding in the present context. First (Fig. 4), a group is fixed, and the linear predictor is examined with synchronized covariates \mathbf{X}' so that a genus of this group with an assumed age t (My since its first appearance) is exposed to conditions at geologic time T . This is a theoretical construct approximated by plugging parameter estimates into (2). For example, if a 20-My-old trilobite had been present at geologic time T , the linear predictor addresses the risk it would have been subject to, even if no trilobites were present at time T because they were either extinct or had not yet emerged. Differences between young and old genera hypothetically transplanted to live in distinct geologic times can be visualized in this manner. Similar to Fig. 3, this is a two-dimensional representation—a contour plot—having dimensions of age (t) and geologic time (T) but originating from a parametric model involving explanatory variables and group membership.

According to Fig. 4, all taxa analyzed showed a negative trend in $\log(\text{predictor})$ values, i.e., hazard, suggesting that younger biotas are more likely to survive extinction than past biotas. However, this trend is not as clear-cut as was previously observed using the Kaplan-Meier estimate for survival (Fig. 3) since these results are more complex and complete, probably due to the number of environmental covariates introduced. Considering this, the results allow greater detail to be observed in the risk pattern of the selected biota, confirming that the hazard is not solely time- and age-dependent but also depends on the taxa analyzed (i.e., *taxa-dependent*). An examination of Fig. 4 reveals that, in particular, the Cephalopoda and Trilobita groups were subject to greater hazard than the other taxonomic groups considered. Moreover, each group presented an apparent age pattern in their hazard distribution.

These patterns were not regular among taxa or within one taxon, since the hazard varied throughout its range. To further investigate this, a second plotting device, which evaluates

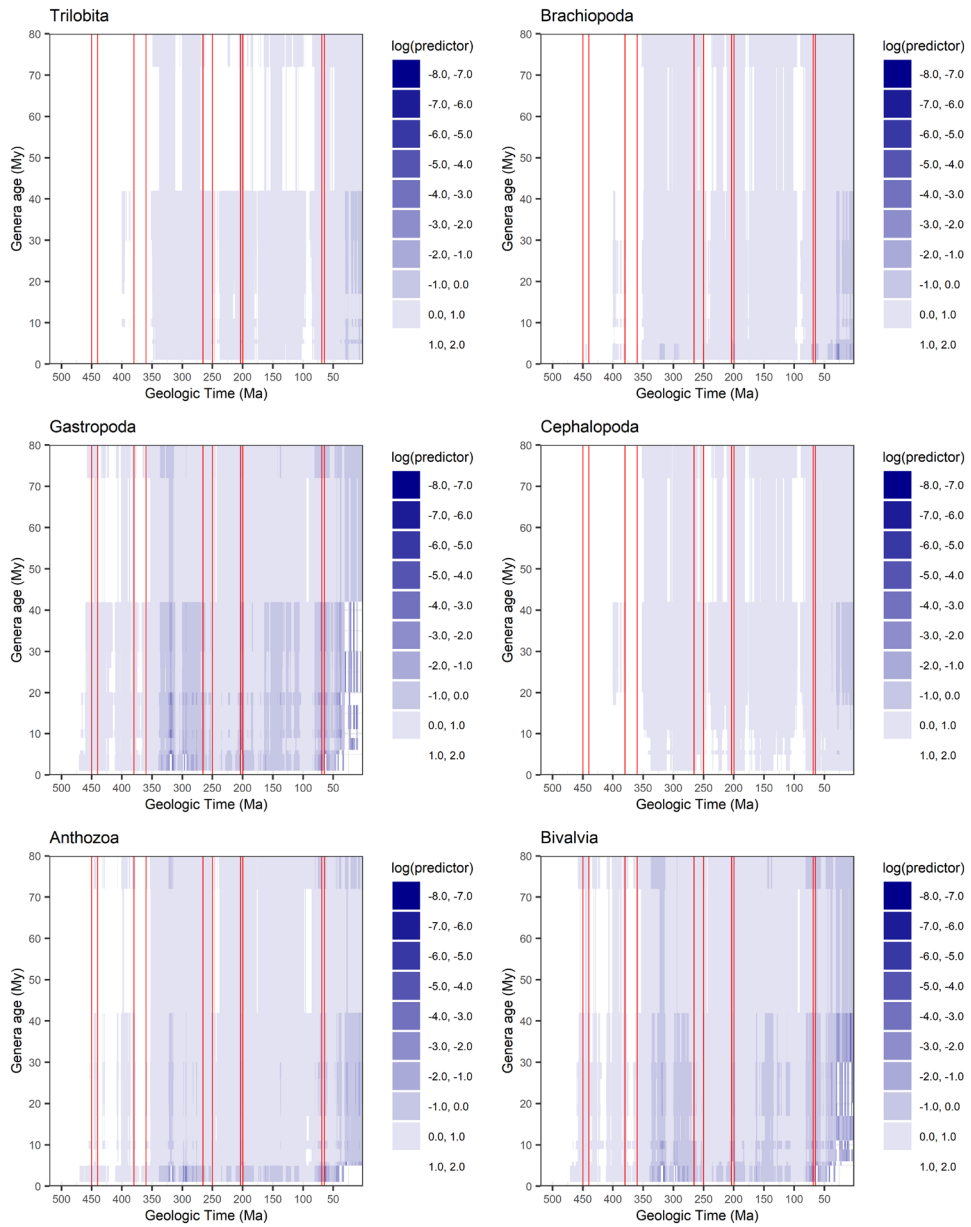


Figure 4. Groupwise contour plots on the (T, t) plane displaying the behavior of the linear predictor for hazard with estimated parameters in the Cox model. A lighter shade depicts a higher hazard, while darker shading represents a lower hazard. Superimposed vertical lines delimit the five recognized mass extinctions. The range for t ends at 80 because the largest observed lifetime that is found to be relevant from diagnostic procedures is 72; then, coefficients with constant behavior are observed.

the hazard values of a taxon at selected representative ages, was applied. The resulting plot (Fig. 5) enables the visualization of the variations in the hazard (i.e., extinction risks) of each selected taxon during the Phanerozoic Eon while considering ages that represent immature, young, mature, and old genera. The results confirm that the hazard is not monotonous for taxa, as was preliminarily observed in Fig. 4, but varies depending on the time of its occurrence (T) and its age (t). The Trilobita and Cephalopoda genera showed more resilience (lower hazard), mature, and older genera, while younger taxa were more resilient in the Brachiopoda, Gastropoda, Bivalvia, and Anthozoa genera. This trend was quite steady during most of their ranges but could be disrupted in some particular events (e.g., Trilobita at $T = 200$ Ma).

A description of cohort behavior in relative terms is afforded by a third graphical device. Let $N_i(T)$ be the number of genera present at geologic time T for the i th group, \mathbf{X}'_j be the set of covariates associated with the j th genus, and $N(T) = \sum_{i=1}^6 N_i(T)$. Figure 6 shows a set of groupwise plots of

$$\frac{N_i(T)^{-1} \sum_{j=1}^{N_i(T)} h(t_j, \mathbf{X}'_j)}{N(T)^{-1} \sum_{k=1}^{N(T)} h(t_k, \mathbf{X}'_k)}, \quad (3)$$

where the sum in the numerator is limited to genera in the i th group, whereas the denominator includes the union of all genera in the six groups of interest. The plotted quantity amounts to the ratio of average hazards in each group in reference to the average hazard in all groups. This ratio is relevant because the baseline hazard $h_0(t)$ drops out, leaving only linear predictors and estimated parameter values. This third device allows a comparison of the mean hazard among groups over geologic time. Note that the ratios of the average hazard of cephalopods, gastropods, and trilobites consistently showed the highest values (Fig. 6). Moreover, the studied taxa showed an increasing trend in their hazard ratios from the Cambrian Period (538.8 ± 0.2 Ma) to the present, with apparent stabilization after the last mass extinction at 66 Ma (K/Pg boundary). This trend suggests that although all groups analyzed show a decreasing hazard trend (Figs. 4 and 5), there is a larger difference between the groups with higher risk and those with lower risk over time.

4. DISCUSSION AND CONCLUSION

The Sepkoski compendium (Sepkoski 2002) was analyzed using the Cox regression model to characterize the extinction patterns of ancient biotas. Previous analyses of the Sepkoski compendium have largely used FAD and LAD data to infer genus longevity and to characterize diversification periods as well as extinction events (Etter 2007). Here, we implemented the Cox regression model in the realm of survival analysis, emphasizing a hazard function for extinction. While previous survival analyses have applied the Cox model to a single taxonomic group in a short time span (Doran et al. 2006; Abelson 2019), here the statistical model involved the complete Sepkoski compendium, composing of $>36,000$ entries of marine metazoan genera found during the whole *Phanerozoic Eon* (from 538.8 ± 0.2 million years ago (Ma) to the present). Moreover, the model included a multifactorial

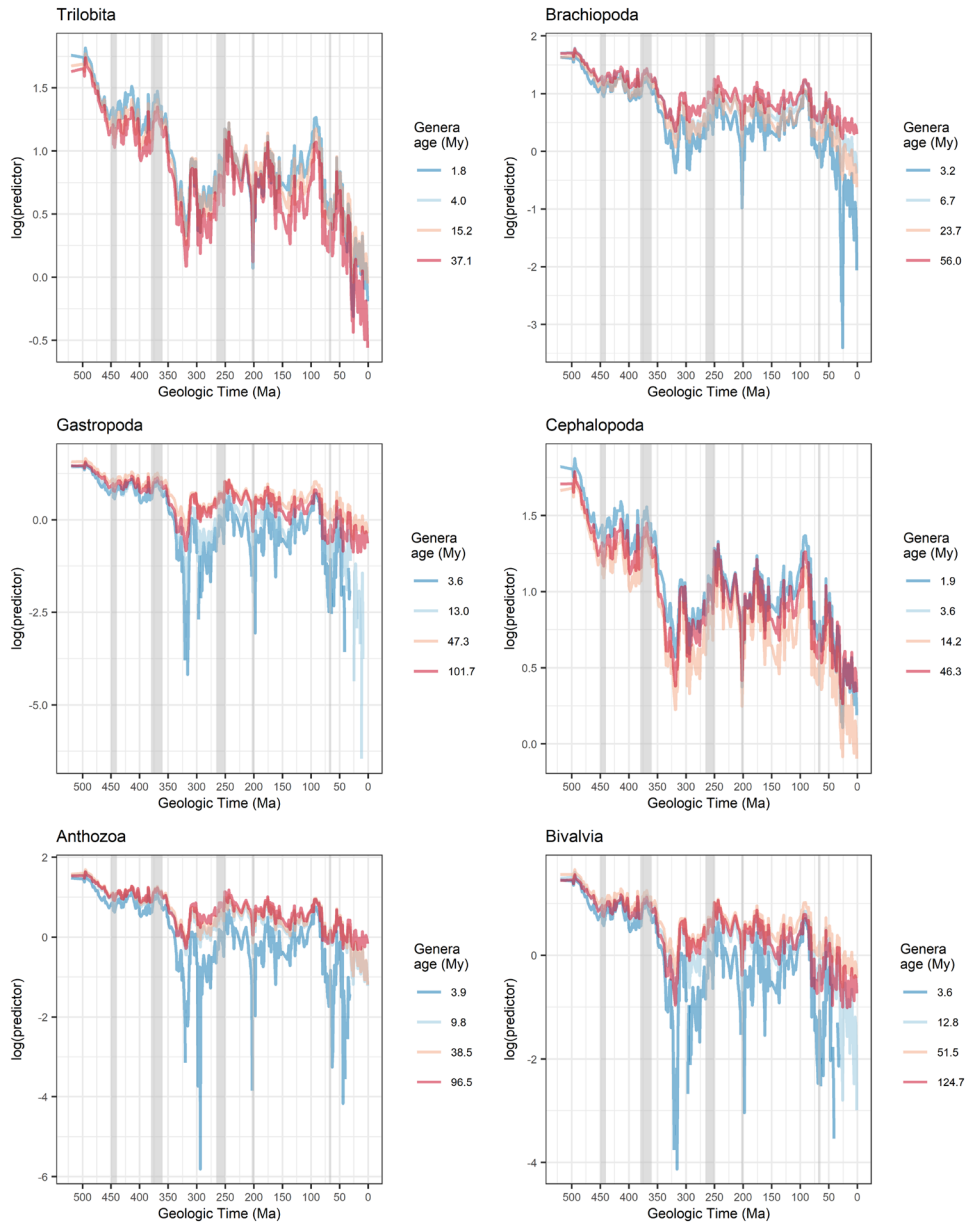


Figure 5. Groupwise evolution of hazards evolving as a function of geologic time (Ma) at selected ages (My). Four ages are plotted for each group, corresponding to sample lifetime quantiles of orders 0.05, 0.25, 0.75, and 0.95. (For example, 25% of trilobites became extinct before age 4.0 My, and 95% of them did so before 37.1 My.) Shaded areas over geologic time delimit the spans of the five recognized mass extinctions.

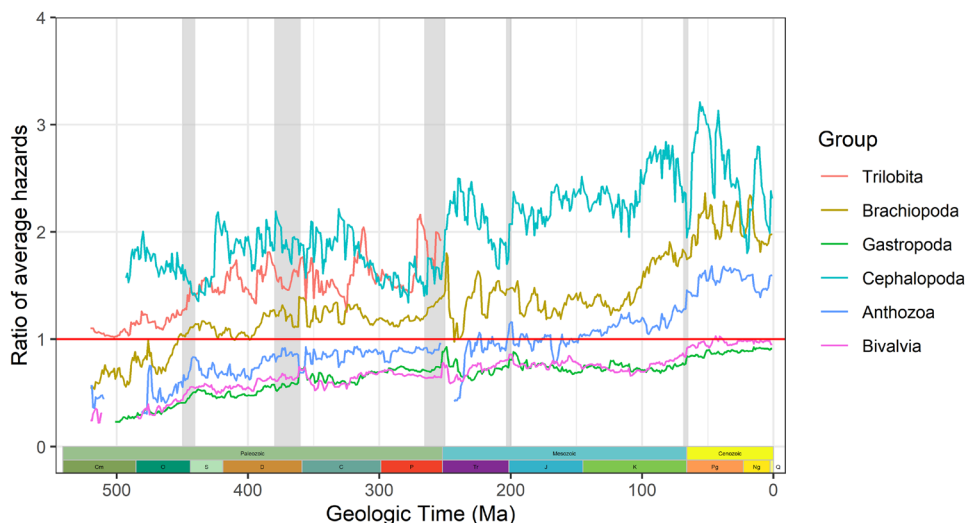


Figure 6. Average hazard per group evolving in geologic time relative to the average hazard over all six groups combined. The horizontal solid line is the reference ratio 1. Shaded areas over geologic time represent the spans of the five recognized mass extinctions. Abbreviations are given in the caption of Fig. 1.

approach, thus relevant environmental covariates that may have affected the marine biota during the past 538.8 Ma (i.e., SST, CO₂ concentration, 62- and 140-My cycles, and sea level) and six relevant taxonomic groups were considered. Employing a bivariate function of taxa age (t) and geologic time (T) and including covariates, this new approach using the Cox model provided a more thorough description of past extinction events with a fresh perspective.

4.1. PALEOBIOLOGICAL SETTING

Regarding how and when past taxa became extinct, [Van Valen \(1973\)](#) proposed the law of constant extinction (LCE), which states that no link exists between the age of a lineage and its risk of extinction (i.e., constant hazard in t). This implies that long-lived genera (older genera t) have the same risk (or hazard) of becoming extinct as short-lived genera (younger genera t). In contrast to the LCE, it has been demonstrated that the risk of extinction of a given taxon changes as a function of its age, whether decreasing, because long-lived taxa are better equipped to endure environmental disturbances (e.g., [Silvestro et al. 2020](#)), or increasing, since newly formed species may be less ecologically specialized (e.g., [Doran et al. 2006](#)). To address this idea, the term *age-dependent extinction* or ADE was proposed in the literature (e.g., [Doran et al. 2006](#); [Alexander et al. 2016](#)).

Further analyses with ADE in mind have determined that extinctions not only depend on the taxa age but also vary according to the taxonomic group and taxonomic level, such as genus or species (e.g., [Hagen et al. 2018](#)), and finally, that selective extinction shifts through time ([Crampton et al. 2016](#)).

The results obtained after the Cox analyses on the six taxonomic groups of Sepkoski's compendium are in line with previous ADE studies, verifying that the extinction risk depends

on taxa age but also confirming that the hazard varies among all the taxa selected in the study (taxa-dependent). As shown in Fig. 4, a time-dependent extinction risk was also observed. Here, all studied taxa show smaller hazard values for recent genera than past genera, which agrees with the declining trend observed that background extinction (Raup and Sepkoski 1982) is related to the ecological differentiation observed over time (Knope et al. 2020). However, the decline in the hazard trend observed from the deep past to the present is not constant within a taxon or between taxa.

A detailed view of the age influence within each group is presented in Fig. 5. Bivalvia, Gastropoda, Brachiopoda, and Anthozoa display young resilient taxa, as was also claimed by Silvestro et al. (2020); however, classes Trilobita and Cephalopoda show mature resilient taxa. Likewise, Fig. 5 indicates that despite a clear trend observed in all groups, mature or young resilient taxa are also linked to geologic time, since in some periods, the trend is inverted (e.g., Trilobita at $T = 200$ Ma). This shift in the extinction mode was previously observed in Graptolites during the Ordovician and Silurian Crampton et al. (2016) but is confirmed here in the six major groups of marine metazoans.

According to Figs. 4 and 5, the hazard trend is affected not only by the taxon age (t) but also by the covariates present during each geologic time (T). This confirms that extinction proneness differs over time in all the marine groups analyzed during the *Phanerozoic Eon*, probably due to the large-scale environmental phenomena that involved some or all of the covariates used in this study. The proposed model, which, contrary to previous works, has been applied in a large and diverse selection of marine metazoans during the whole *Phanerozoic*, indicates that the extinction risk, taxon age, time of existence, and covariates produced a joint effect; as a result, it is impossible to isolate each factor, and thus, the interpretation of the results must also be simultaneous.

In summary, these results suggest that the extinction patterns are more complex than originally conceived by Van Valen; thus, the LCE is still extensively discussed (e.g., Hagen et al. 2018 and references therein) and worthy of further investigation. As such, the relationship between taxon age and extinction risk remains one of the main tenets of biodiversity today (Benton 2016).

4.2. STATISTICAL SETTING

A number of limitations in the modeling approach are evident. Although geologic dating and taxonomy are both conceptually subject to uncertainties (albeit not systematically quantified), they have not been explicitly taken into account. Likewise, the function adopted for describing fitness is not empirically driven, and the covariates included are only those readily available. More sophisticated yet noncompiled covariates are potentially possible, such as the mean body mass (Payne and Heim 2020) or carbon and oxygen excursions (Bond and Grasby 2017).

As in any regression model, the plausibility of *errors in variables*, i.e., that the covariates contain measurement error, is most likely present (Prentice 1982). However, our analysis did not incorporate this. The use of the Bayesian methods to address this complication is a viable possibility, requiring a deep subject matter understanding for the specification of priors (Pan et al. 2022).

Further technical challenges are present in this context. The first challenge is the treatment of the considerable number of singletons so that they are not fully ignored; thus, the analysis that is performed is conditional on positive longevity. The Cox model readily accommodates right censoring for extant taxa. However, double censoring in the sense of a missing FAD and LAD is a matter of greater complexity that entails methods to impute missing values or to enable estimation under this circumstance, especially in the presence of time-dependent covariates and parameters.

As mentioned, the sampling of geologic strata is highly heterogeneous. Quantifying the sampling effort is a particularly complex endeavor in the study of extinction. Even if it were clear-cut, the sampling effort is not a covariate in the same sense when implementing Cox regression. This concept is not an influential factor in the longevity of a species but rather a characteristic of data acquisition, i.e., the process implemented by paleobiologists to search for fossils. The explicit incorporation of the sampling effort would be important to examine its effects on parameter estimation and uncertainty. There is an argument called *weights* in the `coxph` R function that could be fed with a proxy of the sampling effort as a means to reduce the influence of data points stemming from scarce exploration. An additional, potentially confounding factor is preservation bias, i.e., the fact that a dead organism may not always turn into a fossilized specimen. This also poses a modeling challenge, as preservation probabilities may also depend on covariates.

Although residual-based validation was performed on the fitted model, it was not as thorough as simpler versions of the Cox model. A methodological cost related to validation was paid, as insufficiencies in the fit are not readily detected by plotting the residuals versus explanatory variables. However, in the most complex Cox model, for each genus, multiple values for each covariate occur as it varies over time. It is not clear what value of covariate to use for that purpose. The methodology for validating Cox models when they have been complicated by coefficients that are dependent on t as well as covariates that depend on T is recognized as an open problem (Therneau and Grambsch 2000). Estimation of the baseline hazard in this more elaborate situation is also unresolved.

Having adopted the notion of hazard to describe the mechanics of past extinctions (the past 538.8 My), the question of how to compare the estimated hazard obtained by the Cox model with an estimated hazard, presumably attributable to human impacts and obtained from the analysis of modern species (last 500 years), has been posed. Although selected covariates for modern times may be determined, the answer is not evident for several reasons. The main point is that the Cox model has been applied to data at the genus level, whereas the behavior of modern genera has not yet been observed on the same evolutionary timescale. A sensible comparison in terms of hazard would entail transforming from genus to species as well as rescaling time. Determining the uncertainty of statistical inferences would be an important issue to develop, but this comparison with regard to Cox models with differing timescales and sampling units also appears to be an open problem.

ACKNOWLEDGEMENTS

Funding was provided by the Consejo Nacional de Ciencia y Tecnología (Grant No. A1S19598 and scholarship 802877) and Secretaría de Investigación y Posgrado, Instituto Politécnico Nacional (Grant No. 20221243). The authors are also grateful to Editors and two reviewers for enriching comments and suggestions.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Declarations

Conflict of interest The authors declare that there are no conflicts of interest.

[Received September 2022. Revised March 2023. Accepted April 2023.]

REFERENCES

- Abelson ES (2019) Big brains reduce extinction risk in carnivora. *Oecologia* 191(4):721–729. <https://doi.org/10.1007/s00442-019-04527-5>
- Alexander HK, Lambert A, Stadler T (2016) Quantifying age-dependent extinction from species phylogenies. *Syst Biol* 65(1):35–50
- Alroy J (2010a) Fair sampling of taxonomic richness and unbiased estimation of origination and extinction rates. *Paleontol Soc Pap* 16:55–80. <https://doi.org/10.1017/S1089332600001819>
- Alroy J (2010b) Geographical, environmental and intrinsic biotic controls on Phanerozoic marine diversification. *Palaeontology* 53(6):1211–1235. <https://doi.org/10.1111/j.1475-4983.2010.01011.x>
- Alroy J (2010c) The shifting balance of diversity among major marine animal groups. *Science* 329(5996):1191–1194. <https://doi.org/10.1126/science.1189910>
- Alroy J, Marshall CR, Bambach RK et al (2001) Effects of sampling standardization on estimates of Phanerozoic marine diversification. *Proc Natl Acad Sci* 98(11):6261–6266. <https://doi.org/10.1073/pnas.111144698>
- Alroy J, Aberhan M, Bottjer DJ et al (2008) Phanerozoic trends in the global diversity of marine invertebrates. *Science* 321(5885):97–100. <https://doi.org/10.1126/science.1156963>
- Bambach RK, Knoll AH, Sepkoski JJ Jr (2002) Anatomical and ecological constraints on Phanerozoic animal diversity in the marine realm. *Proc Natl Acad Sci* 99(10):6854–6859
- Barnosky AD, Matzke N, Tomiya S et al (2011) Has the Earth's sixth mass extinction already arrived? *Nature* 471(7336):51–57. <https://doi.org/10.1038/nature09678>
- Benton MJ (1995) Diversification and extinction in the history of life. *Science* 268(5207):52–58
- Benton M (2009) The completeness of the fossil record. *Significance* 6(3):117–121. <https://doi.org/10.1111/j.1740-9713.2009.00374.x>
- Benton MJ (2016) Origins of biodiversity. *PLoS Biol* 14(11):e2000724
- Bond DPG, Grasby SE (2017) On the causes of mass extinctions. *Palaeogeogr Palaeoclimatol Palaeoecol* 478:3–29
- Burney DA, Flannery TF (2005) Fifty millennia of catastrophic extinctions after human contact. *Trends Ecol Evol* 20(7):395–401
- Casanovas-Vilar I, García-Paredes I, Alba DM et al (2010) The European Far West: Miocene mammal isolation, diversity and turnover in the Iberian peninsula. *J Biogeogr* 37(6):1079–1093

ROLE OF TAXA AGE AND GEOLOGIC RANGE

- Chao A, Jost L (2012) Coverage-based rarefaction and extrapolation: standardizing samples by completeness rather than size. *Ecology* 93(12):2533–2547. <https://doi.org/10.1890/11-1952.1>
- Collett D (2015) Modelling survival data in medical research, 3rd edn. Chapman & Hall/CRC Texts in Statistical Science Series. CRC Press, Taylor & Francis Group, Boca Raton. ISBN 9781439856789
- Crampton JS, Cooper RA, Sadler PM et al (2016) Greenhouse-icehouse transition in the Late Ordovician marks a step change in extinction regime in the marine plankton. *Proc Natl Acad Sci USA* 113(6):1498–1503. <https://doi.org/10.1073/pnas.1519092113>
- Doran NA, Arnold AJ, Parker WC et al (2006) Is extinction age dependent? *Palaios* 21(6):571–579
- Drake JM (2006) Extinction times in experimental populations. *Ecology* 87(9):2215–2220
- Etter W (2007) Patterns of diversification and extinction. In: Henke W, Tattersall I (eds) *Handbook of paleoanthropology*, vol 1. Springer, Berlin, pp 441–501
- Fan JX, Shen SZ, Erwin DH et al (2020) A high-resolution summary of Cambrian to early Triassic marine invertebrate biodiversity. *Science* 367(6475):272–277
- Fitzgerald PC, Carlson SJ (2006) Examining the latitudinal diversity gradient in Paleozoic terebratulide brachiopods: should singleton data be removed? *Paleobiology* 32(3):367–386
- Flemming C, MacPhee RDE (1999) Redetermination of holotype of *Isolobodon portoricensis* (Rodentia, Capromyidae), with notes on recent mammalian extinctions in Puerto Rico. *Am. Mus. Novit*, p 3278
- Fraser D, Soul LC, Tóth AB et al (2021) Investigating biotic interactions in deep time. *Trends Ecol Evol* 36(1):61–75
- Hagen O, Andermann T, Quental TB et al (2018) Estimating age-dependent extinction: contrasting evidence from fossils and phylogenies. *Syst Biol* 67(3):458–474. <https://doi.org/10.1093/sysbio/syx082>
- Hammer Ø, Harper DAT (2006) *Paleontological data analysis*. Blackwell, Malden. 9781405115445
- Haq BU, Hardenbol J, Vail PR (1987) Chronology of fluctuating sea levels since the Triassic. *Science* 235(4793):1156–1167. <https://doi.org/10.1126/science.235.4793.1156>
- Harrison CGA (2002) Power spectrum of sea level change over fifteen decades of frequency. *Geochem Geophys Geosyst* 3(8):1–17. <https://doi.org/10.1029/2002GC000300>
- Hautmann M (2012) Extinction: end-Triassic mass extinction. In: *Encyclopedia of life sciences*, pp 1–10. Wiley, Chichester. <https://doi.org/10.1002/9780470015902.a0001655.pub3>
- Hull P (2015) Life in the aftermath of mass extinctions. *Curr Biol* 25(19):R941–R952
- IUCN Standards and Petitions Committee (2019) Guidelines for using the IUCN red list categories and criteria. Version 14. Prepared by the Standards and Petitions Committee
- Jablonski D (2005) Mass extinctions and macroevolution. *Paleobiology* 31(S2):192–210
- Knope ML, Bush AM, Frishkoff LO et al (2020) Ecologically diverse clades dominate the oceans via extinction resistance. *Science* 367(6481):1035–1038
- Marshall CR (2010) Marine biodiversity dynamics over deep time. *Science* 329(5996):1156–1157. <https://doi.org/10.1126/science.1194924>
- Ogg JG, Gradstein FM, Smith AG (2004) *A geologic time scale 2004*, vol 86. Cambridge University Press, Cambridge
- Pan A, Song X, Huang H (2022) Bayesian analysis for partly linear Cox model with measurement error and time-varying covariate effect. *Stat Med* 41(23):4666–4681. <https://doi.org/10.1002/sim.9531>
- Payne JL, Heim NA (2020) Body size, sampling completeness, and extinction risk in the marine fossil record. *Paleobiology* 46(1):23–40. <https://doi.org/10.1017/pab.2019.43>
- Prentice RL (1982) Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika* 69(2):331–342. <https://doi.org/10.1093/biomet/69.2.331>
- R Core Team (2021) *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna
- Raup DM (1972) Taxonomic diversity during the Phanerozoic. *Science* 177(4054):1065–1071. <https://doi.org/10.1126/science.177.4054.1065>
- Raup DM (1978) Cohort analysis of generic survivorship. *Paleobiology* 4(1):1–15

- Raup D (1986) Biological extinction in earth history. *Science* 231(4745):1528–1533. <https://doi.org/10.1126/science.11542058>
- Raup DM, Sepkoski JJ Jr (1982) Mass extinctions in the marine fossil record. *Science* 215(4539):1501–1503
- Raup DM, Sepkoski JJ (1984) Periodicity of extinctions in the geologic past. *Proc Natl Acad Sci* 81(3):801–805. <https://doi.org/10.1073/pnas.81.3.801>
- Rohde RA, Muller RA (2005) Cycles in fossil diversity. *Nature* 434(7030):208–210. <https://doi.org/10.1038/nature03339>
- Sandom C, Faurby S, Sandel B et al (2014) Global late Quaternary megafauna extinctions linked to humans, not climate change. *Proc R Soc B Biol Sci* 281(1787):20133254
- Sepkoski JJ Jr (1982) Compendium of fossil marine families. *Milwaukee Public Mus Contrib Biol Geol* 51:1–125
- Sepkoski JJ Jr (1992) A compendium of fossil marine animal families. *Contrib Biol Geol* 83:1–156
- Sepkoski JJ (2002) A compendium of marine fossil genera. Paleontological Research Institution, Ithaca
- Signor PW, Lipps JH, Silver L, et al (1982) Sampling bias, gradual extinction patterns, and catastrophes in the fossil record. In *Geological implications of impacts of large asteroids and comets on the Earth*, vol 190, pp 291–296. Geological Society of America Special Publication
- Silvestro D, Castiglione S, Mondanaro A et al (2020) A 450 million years long latitudinal gradient in age-dependent extinction. *Ecol Lett* 23(3):439–446
- Song H, Wignall PB, Song H et al (2019) Seawater temperature and dissolved oxygen over the past 500 million years. *J Earth Sci* 30(2):236–243. <https://doi.org/10.1007/s12583-018-1002-2>
- Song H, Kemp DB, Tian L et al (2021) Thresholds of temperature change for mass extinctions. *Nat Commun* 12(1):1–8
- Spiridonov A, Lovejoy S (2022) Life rather than climate influences diversity at scales greater than 40 million years. *Nature* 607(7918):307–312. <https://doi.org/10.1038/s41586-022-04867-y>
- Therneau TM (2021) A package for survival analysis in R. R package version 3.2-13
- Therneau TM, Grambsch PM (2000) *Modeling survival data: extending the Cox model*. Statistics for Biology and Health, Springer, New York. 9781441931610 9781475732948. <https://doi.org/10.1007/978-1-4757-3294-8>
- Van Valen L (1973) A new evolutionary law. *Evol Theory* 1:1–30

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.