[American Journal of Science, Vol. 315, April, 2015, P. 275–316, DOI 10.2475/04.2015.01]

American Journal of Science

APRIL 2015

A STATISTICAL ANALYSIS OF THE CARBON ISOTOPE RECORD FROM THE ARCHEAN TO PHANEROZOIC AND IMPLICATIONS FOR THE RISE OF OXYGEN

JOSHUA KRISSANSEN-TOTTON*,†, ROGER BUICK*, and DAVID C. CATLING*
* Department of Earth and Space Sciences/Astrobiology Program, University of Washington, Seattle, Washington 98195, USA

ABSTRACT. Organic and inorganic carbon isotope records reflect the burial of organic carbon over geological timescales. Permanent burial of organic carbon in the crust or mantle oxidizes the surface environment (atmosphere, ocean and biosphere) by removing reduced carbon. It has been claimed that both organic and inorganic carbon isotope ratios have remained approximately constant throughout Earth's history, thereby implying that the flux of organic carbon burial relative to the total carbon input has remained fixed and cannot be invoked to explain the rise of atmospheric oxygen (Schidlowski, 1988; Catling and others, 2001; Holland, 2002; Holland, 2009; Kump and others, 2009; Rothman, 2015). However, the opposite conclusion has been drawn from the same carbon isotope record (Des Marais and others, 1992; Bjerrum and Canfield, 2004). To test these opposing claims, we compiled an updated carbon isotope database and applied both parametric and non-parametric statistical models to the data to quantify trends and mean-level changes in fractional organic carbon burial with associated uncertainties and confidence levels.

We first consider a conventional mass-balance model where carbon input to surficial reservoirs is balanced by burial of sedimentary carbonates and organic carbon. For this model, statistical analysis implies fractional organic burial has increased over Earth history by a factor of 1.5 relative to organic burial at 3.6 Ga, with the 95 percent confidence interval ranging from factors of 1.2 to 2.0. An increase in organic burial by a factor of 1.2 cannot explain the rise of oxygen, whereas an increase by a factor of 2 could conceivably explain the rise of oxygen. There is, however, a highly significant and well constrained increase in organic burial from the Proterozoic to the Phanerozoic.

We also analyze changes in the difference between carbonate and organic carbon isotopic ratios over Earth history. There is a statistically significant increase in this difference from the early to late Archean, possibly caused by increased biological fractionation due to methanotrophic recycling. This transition is consistent with the evolution of oxygenic photosynthesis at 2.8 Ga or earlier.

Finally, we explore how these conclusions change if we modify the traditional mass balance model to include other carbon cycle fluxes, specifically ocean crust carbonatization and authigenic carbonates. Because the size of these fluxes has a large, poorly constrained range, our statistical analysis with this uncertainty implies that the carbon isotope record does not constrain the history of organic burial at all. However, it remains possible that the magnitude of these additional processes has been inconsequential throughout geologic time, in which case conclusions from the conventional model would be valid.

Keywords: oxygen, carbon isotopes, Great Oxidation Event, methanotrophy, carbon cycle, photosynthesis, carbon burial, authigenic carbon

† Corresponding author: joshkt@uw.edu

1) INTRODUCTION

Explaining the oxygenation of Earth's atmosphere and oceans is a great unsolved problem in Earth and planetary science. Its solution would enrich our understanding of the co-evolution of life and the environment because abundant atmospheric oxygen sustains virtually all macroscopic multicellular life on Earth (Catling and others, 2005). Moreover, understanding the 2.4 Ga rise of oxygen and subsequent increases on Earth (for example Lyons and others, 2014) may inform the search for life on exoplanets because abundant atmospheric O_2 or O_3 is a potential biosignature that may be detectable with next-generation large-aperture telescopes (Des Marais and others, 2002).

Multiple lines of geological evidence suggest that atmospheric oxygen levels have increased dramatically over Earth's history (Farquhar and others, 2014). In particular, mass independent fractionation of sulfur isotopes indicates that atmospheric oxygen was $\leq 10^{-5}$ of present atmospheric levels during the Archean. Then around 2.4 Ga, a transition led to levels of oxygen exceeding 10^{-3} to 10^{-2} of present atmospheric levels (Holland, 1994; Pavlov and Kasting, 2002; Farquhar and others, 2007). Atmospheric oxygen is believed to have remained relatively low for the next billion years until the Neoproterozoic when oxygen levels increased again (Shields-Zhou and Och, 2011). Geochemical indicators suggest that this, in turn, was possibly followed by a further increase in the Devonian associated with the colonization of the land by vascular plants (Dahl and others, 2010).

Although it is widely accepted that atmospheric oxygen has increased over Earth history, the precise causes of this increase are debated (for example Kasting, 2013; Catling, 2014; Lyons and others, 2014 for recent reviews). Evidently oxygen source fluxes exceeded oxygen sink fluxes at times during Earth's history; the 2.4 Ga rise, in particular, established a new balance with a permanently oxygenated atmosphere.

There are two broad categories of explanations for the Paleoproterozoic oxic transition: 1) the oxygen source flux permanently increased due to enhanced organic carbon burial (Campbell and Allen, 2008) including the possibility that organics were preferentially subducted into the mantle (Hayes and Waldbauer, 2006), and 2) the oxygen sink flux permanently decreased. Hypotheses for the latter include changes in the redox state of volcanic gases due to mantle redox evolution (Kasting and others, 1993), a transition from subaerial to submarine volcanism (Kump and Barley, 2007; Gaillard and others, 2011), or a change in carbon or sulfur recycling (Holland, 2009). Alternatively, a change in the redox state of low temperature crustal volatiles has been proposed due to hydrogen escape oxidizing the crust (Catling and others, 2001; Claire and others, 2006; Zahnle and others, 2013).

The geological carbon cycle is linked to the oxygen source flux. Earth's atmospheric oxygen is produced almost exclusively by photosynthesis, which can be stoichiometrically approximated by the following:

$$CO_2 + H_2O \rightarrow CH_2O + O_2 \qquad (1.1)$$

However the oxygen produced during photosynthesis doesn't constitute a net source of atmospheric oxygen unless it is coupled with the burial of organic matter. In the absence of organic burial, biogenic oxygen merely re-oxidizes organic matter during respiration, reversing equation (1.1), and there is no net change in atmospheric oxygen. Thus the burial rate of organic carbon equals the rate of oxygen production, and hypotheses that invoke an increased source flux to explain the rise of oxygen necessarily invoke enhanced organic burial as the mechanism. The long-term behavior of the geological carbon cycle is thereby directly related to atmospheric oxygen. It is the majority opinion that the evolution of oxygenic photosynthesis preceded the rise of atmospheric oxygen by at least several hundred million years (Catling, 2014). Note

that if anoxygenic photosynthesis was the source of Archean organic matter then the burial of organic carbon does not add molecular oxygen to that atmosphere. However, anoxygenic photosynthesis still produces an oxidized product (such as ferric iron), and so the burial of organic carbon still implies an oxidation flux for the surface.

The factors that control organic carbon burial are complex and not fully understood. All else being equal, enhanced biological productivity will result in greater deposition of organic matter in sediments. However, only a small fraction of the organic matter deposited in sediments becomes permanently buried (Hedges and Keil, 1995). Instead, most sedimentary organic matter is remineralized to carbon dioxide by microbial metabolisms. Biological productivity and the remineralization fraction are potentially influenced by a variety of geological and biological factors such as tectonic context (Des Marais, 1994), sedimentation rates (Betts and Holland, 1991), oxygen exposure time in sediments (Hartnett and others, 1998), and evolutionary innovations (Logan and others, 1995).

Fortunately, the carbon isotope record provides a window into the geological carbon cycle through deep time (Broecker, 1970). A conventional approach, pioneered by Schidlowski and others (1979; Schidlowski, 1988), assumes that the geological carbon cycle is in steady state over long timescales, or more specifically that (i) the carbon input into surficial reservoirs (atmosphere, oceans and biosphere) is balanced by the burial of carbonates plus organic carbon, and (ii), that the isotopic ratio of carbon inputs equals the average isotopic composition of burial outputs. Taken together these assumptions yield the conservation equation:

$$\delta^{13}C_{in} = f_{org}\delta^{13}C_{org} + (1 - f_{org})\delta^{13}C_{carb} \tag{1.2}$$

Here, $\delta^{13}C_{in}$ is the isotopic abundance of the outgassed carbon inputs, f_{org} is the fraction of the total carbon buried that is buried as organic carbon, and $\delta^{13}C_{org}$ and $\delta^{13}C_{carb}$ are the isotopic abundances of the buried organic and carbonate carbon respectively. Standard isotopic notation is $\delta^{13}C = 1000 \times (R_{sample} - R_{std})/R_{std}$ where R_{sample} is the $^{13}C/^{12}C$ ratio of the sample, and R_{std} is the $^{13}C/^{12}C$ ratio of the Pee Dee Belemnite standard (Faure and Mensing, 2005, p. 620, 705). Oxygenic photosynthesis preferentially fixes isotopically light carbon, and so the two carbon sinks are isotopically distinct with $\delta^{13}C_{org}$ approximately 20 to 30 permil lower than $\delta^{13}C_{carb}$. Rearranging equation (1.2) yields an expression for f_{org} as a function of the isotopic abundances of the inputs and outputs:

$$f_{org} = \frac{(\delta^{13}C_{in} - \delta^{13}C_{carb})}{(\delta^{13}C_{org} - \delta^{13}C_{carb})} \tag{1.3}$$

Peridotititc xenoliths, mantle-derived basalts, and carbonatites suggest that $\delta^{13}C_{in}$ has remained unchanged throughout Earth's history (Mattey, 1987; Holser and others, 1988; Pearson and others, 2003). Shirey and others (2013) found that 72 percent of mantle diamonds have $\delta^{13}C$ values that fall within -5 ± 1 permil (measurements outside this range can be explained by recycled crustal material), and so a constant value of $\delta^{13}C_{in} = -5‰$ is assumed in this study. Consequently, by measuring $\delta^{13}C_{org}$ and $\delta^{13}C_{org}$ over time, the history of f_{org} can be reconstructed. If the absolute rate of total carbon burial can be independently constrained, then f_{org} can be used to infer an absolute oxygen source function through Earth's history. Note that the assumption of constant $\delta^{13}C_{in}$ may not be valid on timescales less than a few hundred million years because preferential weathering of carbonates or organics could modulate the isotopic ratio of riverine carbon input, even if the carbon outgassed remains at -5 permil. However over timescales greater than the rock cycle we can expect the average isotopic value of riverine inputs to equal that of outgassed carbon.

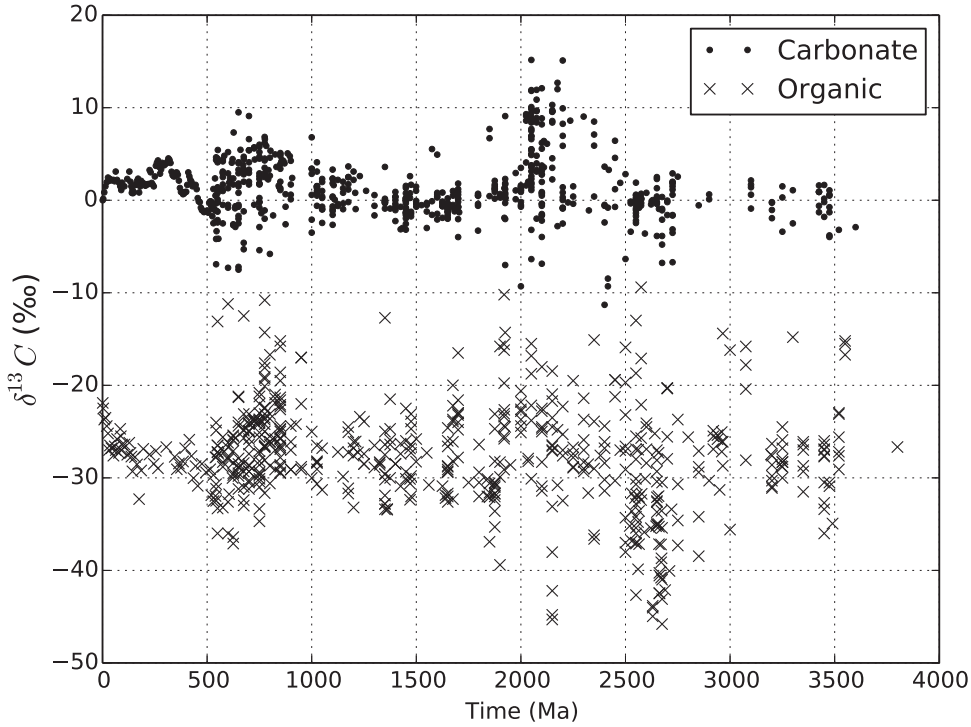

Fig. 1. Carbonate (dots) and organic (crosses) carbon isotope data used in this study. Each data point represents the average per formation per author per lithology. Data exclude lacustrine environments, heavily metamorphosed samples and banded iron formations (see section 2). 13,381 individual carbonate isotope measurements and 4,236 organic isotope measurements were averaged to produce the Precambrian data in this figure (see main text for Phanerozoic data). The method of averaging is described in section 2.

The validity of the steady state assumption is generally accepted. Sundquist (1991) explored a time-dependent model of the carbonate-silicate cycle that yielded response times to perturbations of less than 1 million years. This is consistent with an ocean residence time of around 200 ky; the short response time of the ocean reservoir suggests the steady state approximation is valid. Rothman and others (2003) explicitly modeled the dynamics of the geological carbon cycle and discovered that under certain conditions response times of 10 to 100 million years are possible. However, on longer timescales the steady state approximation can be expected to hold. Any long-duration imbalance in the carbon cycle would deplete or swamp the surface reservoir of carbon or induce large secular changes in isotopic composition.

The carbonate and organic carbon isotope records over Earth history are shown in figure 1. An apparent steady average of both $\delta^{13}C_{org}$ and $\delta^{13}C_{carb}$ has led many authors to conclude that f_{org} has remained approximately constant throughout Earth's history (Schidlowski, 1988; Catling and others, 2001; Holland, 2002; Holland, 2009; Kump and others, 2009; Rothman, 2015). A value of f_{org} around 0.2 is derived if $\delta^{13}C_{org} \approx -25‰$, $\delta^{13}C_{org} \approx 0‰$ and $\delta^{13}C_{in} \approx -5‰$ are assumed in equation (1.3). If f_{org} had been constant, this would imply that secular changes in fractional organic burial - and by extension changes in the oxygen source flux - cannot explain the rise of oxygen. Instead, decreases in one or more oxygen sinks must be responsible for the oxidation of Earth's atmosphere and oceans.

However, this interpretation of the constancy of the carbon isotope record is not universal. For instance, Des Marais and others (1992) incorporated the effects of post-depositional isotopic changes and argued that the carbon isotope record from 2.6 Ga indicates an increase in f_{org} of sufficient magnitude to explain the Paleoproterozoic rise of oxygen. Similarly, Bjerrum and Canfield (2004) suggested that Archean ocean crust carbonatization sequestered light carbon, such that f_{org} could have increased substantially over Earth history as carbonatization declined. Schrag and others (2013) argued that interpreting fluctuations in the carbon isotope record as changes in f_{org} may be incorrect if a sizeable portion of marine carbonate is authigenic in origin. Authigenic carbonates are isotopically distinct from allogenic (sedimentary) carbonates because the carbon that constitutes authigenic carbonates is primarily remineralized organic matter (Schrag and others, 2013).

Even without considering post-depositional alteration of isotopic ratios or additional carbonate sinks there is considerable scatter in the carbon isotope record. In fact it is not obvious whether or not sizeable secular trends or changes in f_{org} could be hidden beneath the noise. Here, we present a rigorous statistical analysis to examine whether secular trends or changes exist in the history of f_{org}.

The carbon isotope time series data are not evenly spaced, do not necessarily have Gaussian noise, and may possess autocorrelation that is a memory of previous states that effectively diminishes the number of independent data points. Thus it is unclear which statistical model best captures the secular changes, noise structure, and temporal autocorrelation of the carbon cycle data. Consequently, our approach is to employ and compare a variety of statistical techniques, both parametric and non-parametric, to ensure robust conclusions. Loosely speaking, parametric techniques assume the data are described by simple functions and distributions, whereas non-parametric techniques make minimal assumptions about the functional form of the data. Section 2 outlines the selection and filtering of carbon isotope data used in this study. Section 3.1 is a non-technical summary of our methods for readers primarily interested in understanding the results and their implications. Sections 3.2 and 3.3 discuss the statistical methods in more detail. Section 3.2 describes the parametric approach, using mean-differencing and linear regressions to evaluate changes in f_{org}. We employ a variety of noise models to characterize the temporal autocorrelation and uncertainty in the data. In section 3.3 we describe an ensemble of non-parametric smoothing techniques used to derive a best estimate of f_{org} over time. Section 4 presents and compares the results from these parametric and non-parametric methodologies. In addition to analyzing f_{org} we also analyze $\varepsilon = \delta^{13}C_{carb} - \delta^{13}C_{org}$ through time. This parameter primarily depends on biological fractionation of carbon with respect to inorganic carbon in seawater and so gives insight into the origin of oxygenic photosynthesis. In section 5, we introduce a modified mass balance equation that includes isotopically distinct sinks such as ocean crust carbonatization and authigenic carbonates. We redo parts of our statistical analysis using this new model. Finally, section 6 discusses the implications of our results for the rise of oxygen, and key conclusions are summarized in section 7.

2) ISOTOPE DATA

We compiled carbon isotope data for both sedimentary carbonates and organic carbon over Earth history back to the early Archean. Carbonate data were taken primarily taken from Shields and Veizer (2002), but were updated with recent data in the literature. Organic carbon data were taken primarily from Schopf and Klein (1992) and references therein, supplemented using recent results in the literature. Rather than report individual isotope measurements, our database uses averaged carbon isotope values; each entry in the database is the mean isotopic value per formation per

author or study per lithology. This averaging is appropriate for the statistical tests we employ in this study because all these tests involve additional averaging or binning of adjacent data points; whether or not data is pre-averaged doesn't change the results appreciably. Where known to be inaccurate, radiometric dates were updated from the literature, and if radiometric dates were unavailable for a given study, then approximate dates were obtained from linear interpolation between dated formations. Spreadsheets containing our dataset are available in the online supplementary materials (http://earth.geology.yale.edu/~ajs/SupplementaryData/2015/KrissansenTotton). The density of Precambrian isotope measurements does not change drastically with time, except for the Archean where there is a scarcity of both organic and carbonate measurements (see supplementary spreadsheet; http://earth.geology.yale.edu/~ajs/SupplementaryData/2015/KrissansenTotton).

The carbon isotope data were also filtered to exclude sediments from non-marine settings, authigenic settings like banded iron formations, or heavily metamorphosed settings. Non-marine sediments were excluded because the steady state mass-balance model relies on carbon-bearing sediments being sourced from an isotopically homogeneous reservoir that reflects the well-mixed global average. This is true for carbon sediments deposited in the Earth's oceans since the typical mixing time in the oceans (~500-1000 years) is small compared to the timescale for sedimentary rock formation. However, the well-mixed assumption doesn't hold for isolated lacustrine environments where isotopic values are potentially variable and may not reflect the global average. Banded iron formations and highly metamorphosed sediments were excluded due to the potential for post-depositional changes in isotopic composition. This filtering is likely imperfect, but it is better than using all carbon isotope data indiscriminately. For completeness we repeated our data analysis on the original unfiltered data and these results are reported only in the Appendices, except where they differed notably from the filtered data analysis.

Since multiple authors often report isotopic ratios for the same formation, and since many geological formations are of similar ages, there are often multiple isotopic values per radiometric date in our data. Standard regression techniques and smoothing algorithms cannot be applied to data with multiple dependent variable values per independent variable value. Consequently, data were binned into 10 million year averages prior to all statistical analyses. Analyses were repeated using 20 million year bins and unique-date bins; results were generally robust to these changes unless stated otherwise. More fundamentally, repeat sampling of the same formation does not constitute independent measurements of globally averaged isotopic ratios in a given geological timescale. Instead it merely refines our knowledge of the globally averaged isotopic ratio at a specific date, rather than improving our knowledge of the mean-level ratio over a larger span of time. Thus binning same-date isotope measurements before applying statistical tests is justified on both pragmatic and theoretical grounds.

Because of the relatively abundant Phanerozoic data, rather than swamp our data set with recent data we used pre-binned data from Hayes and others (1999) to characterize Phanerozoic carbon isotope values. Where possible, data analyses were repeated excluding the Phanerozoic data to test whether results were robust to the exclusion of this qualitatively different data set.

Ideally we would filter organic carbon data using H/C ratios to remove or correct samples with appreciable diagenetic modification (for example Schopf and Klein, 1992; Des Marais, 1997). However H/C ratios are not available for the majority of organic carbon isotope data and so this filtering is not possible. The statistical effects of diagenetic modification are considered briefly in Appendix B.4, and are shown not to be important influences on secular trends.

3) statistical methodologies

3.1) Non-technical Summary of Methods

The carbon isotope data have various features that make them difficult to analyze: they are unevenly spaced in time, serially correlated in time, and there is uncertainty in the nominal radiometric dates. In light of these difficulties we have applied a broad ensemble of statistical techniques to test specific hypotheses about fractional organic burial, and only report conclusions that are robust irrespective of the methodology.

Our methods address three questions about the isotope data:

(i) What is the history of fractional organic burial and biological fractionation over Earth history? In other words, what are $f_{org}(t)$ and $\varepsilon(t)$ with uncertainties? We apply three different smoothing algorithms: locally weighted scatterplot smoothing (LOWESS), kernel regression and Kalman smoothing. These algorithms are each used to estimate $\delta^{13}C_{org}$ and $\delta^{13}C_{carb}$ as functions of time, with uncertainty. The isotope functions are then combined to obtain $f_{org}(t)$ and $\varepsilon(t)$ with uncertainties.

(ii) What is the overall change in fractional organic burial over Earth history with uncertainty? We calculate the linear trend in $\delta^{13}C_{carb}$ and $\delta^{13}C_{org}$ over Earth history, and then combine these to get the linear trend in f_{org} over Earth history. This trend is compared with the null hypothesis that f_{org} has remained constant over Earth history (trend of zero). This analysis is repeated with various assumptions about the correlation structure of data (simple ordinary leased squares (OLS) regression, classical generalized least squares (GLS) regression and bootstrap GLS). Essentially we ask the question under what conditions (if any) can the null hypothesis of no change in f_{org} be rejected, and whether this conclusion is robust to methodology. The start and end points of the trend line in f_{org} can also be used to calculate the absolute and relative increase in f_{org} over Earth history. The relative increase in f_{org} over Earth history is an important quantity for the history of atmospheric oxygen, as explained in the discussion section. We also fit trends to the smoothed curves and check the non-parametric results (i) against parametric results (ii).

(iii) What is the change in f_{org} and ε between key geological time periods? For instance how much does f_{org} change from the early Archean to late Archean, the Archean to the Proterozoic, and the Proterozoic to Phanerozoic? Here, we are testing the null hypothesis that there has been no change in fractional organic burial between these periods, and we quantify any changes that do occur with uncertainties. This analysis is repeated using different methods that make different assumptions about the temporal correlation of data, and only results robust to choice of methodology are reported.

3.2) Parametric Methods

Time series data are often not directly amenable to classical statistical techniques because the data may be correlated in time, as previously mentioned, and thus cannot be treated as independent. To properly characterize uncertainties and perform meaningful statistical tests, temporal autocorrelation must first be accounted for. In our parametric analysis, autocorrelation in the data is explicitly modeled as a first order autoregressive [AR(1)] stationary noise process (see below). Linear regressions and mean differences between two different time periods are then calculated within this autocorrelation framework, and the results are compared to classical statistical techniques where data are assumed to be independent in time.

3.2.1) Autocorrelation model.—Characterizing the autocorrelation in unevenly spaced data is challenging since time series theory is predominately designed for evenly

spaced data. Traditionally, unevenly spaced data are converted to evenly spaced data using linear interpolation. However, if this approach were applied to the carbon isotope data then it would introduce significant bias because the spacing is extremely varied. Instead, we adopt an AR(1) stationary noise model for the unevenly spaced data as described by Mudelsee (2010):

$$X_{noise}(1) = \varepsilon_{N(0,1)}(1) \tag{3.1}$$

$$X_{noise}(i) = X_{noise}(i-1) \times \exp\{-(t(i) - t(i-1))/\tau\} + \varepsilon_{N(0,1-\exp\{-2[t(i) - t(i-1)]/\tau\})}(i) \tag{3.2}$$

Here $i = 1, \ldots, n$, n is the number of data points, and $\varepsilon_{N(0,1)}$ is a random variable from the standard normal distribution with a mean of 0 and standard deviation of 1. Intuitively, at any given time step the noise term $X_{noise}(i)$ is equal to Gaussian noise ε_N plus the previous noise term $X_{noise}(i-1)$, weighted by some autocorrelation parameter, $\exp\{-(t(i) - t(i-1))/\tau\}$, with a magnitude that depends on the time difference between $X_{noise}(i)$ and $X_{noise}(i-1)$. Evidently for closely spaced data the autocorrelation will dominate and consecutive noise terms will be similar, whereas distantly spaced data will approach random noise. AR(1) denotes the noise model is an autoregressive model of order 1. Each noise term is dependent on the previous noise term, but not the 2nd or 3rd previous noise terms, as would be the case for AR(2) and AR(3) models. The persistence time τ is the e-folding time that describes how rapidly autocorrelation falls off as spacing is increased. For time series data $\{t(i), x(i)\}_{i=1}^{n}$ the persistence time τ can be found by minimizing the sum of the squares:

$$S(\tau) = \sum_{i=2}^{n} [x_{noise}(i) - x_{noise}(i-1) \times \exp\{-[t(i) - t(i-1)]/\tau\}]^2 \tag{3.3}$$

Note that we have switched to lower case notation to refer to the actual data, as opposed to upper case notation for the general model. Given this noise model, bootstrapping algorithms can be applied to generate confidence intervals for variables of interest such as the mean change in fractional organic burial over Earth history.

The application of this model to regression analysis assumes that the carbon isotope record can be described as a secular background trend plus autocorrelated noise. The noise in the data consists of measurement noise along with process noise such as local environmental conditions, diagenetic changes and rock cycle feedbacks. Specifically, we would expect the carbon cycle to have memory on timescales less than the ~200 million year rock cycle (Holland, 1978) since this is the time it takes to weather away carbonates and organic carbon on the continents. The presence of some degree of autocorrelation in these various noise processes is confirmed by non-zero τ values in our analysis, and thus the use of an autocorrelation model rather than treating all data as independent is appropriate. The specific choice of an AR(1) model as opposed to higher order autoregressive moving average models is pragmatic; higher order models do not exist for unevenly spaced data (Mudelsee, 2010). Furthermore, the AR(1) model is preferable because it is a correct embedding of a continuous process in time, unlike higher order models which have no continuous analog (Mudelsee, 2010).

3.2.2) Regression analysis.—The regression analysis method can be summarized as follows. A linear regression is performed for both organic carbon isotope data and carbonate isotope data. These regression results, with corresponding uncertainties, are then combined to calculate f_{org}. Following Mudelsee (2010) we incorporate non-stationarity (that is a change in mean with time) into the model as follows:

$$X(i) = X_{trend}(i) + X_{noise}(i) \times V \quad (3.4)$$

Here $X_{trend}(i)$ is (in this case) a linear trend, $X_{noise}(i)$ is the noise term from the AR(1) stationary noise model above (eqs 3.1 and 3.2), and V is the magnitude of the noise, which is taken to be an unknown constant for simplicity.

Three different approaches were used to calculate the linear regression for the carbon isotope data:

I) Ordinary Least Squares (OLS) fit, with no temporal autocorrelation model. This method treats data as independent in time. This is the ubiquitous technique of "simple linear regression."

II) Classical Generalized Least Squares (GLS) fit. The algorithm for this method is provided in Appendix A.1. The basic idea is to provide an initial guess for τ and V, compute the GLS fit along with the uncertainties in fit parameters, and then use these outputs to re-calculate τ and V with a new GLS fit, and repeat until the fitted parameters converge. The GLS fit is an extension of an OLS regression that accounts for AR(1) autocorrelation by preferentially weighting more spaced (less autocorrelated) data points and assuming Gaussian distributions for all uncertainties.

III) Bootstrap GLS. The basic idea is to first perform the classical GLS regression (method II) to estimate the trend, and then subtract the trend from the data to obtain *correlated* residuals. The best-fit τ is then used to perform a weighted differencing to obtain *uncorrelated* residuals. The uncorrelated residuals are resampled with replacement, and the previous operation is inverted to retrieve resampled *correlated* residuals. These correlated residuals are added back to the GLS trend line, and the GLS regression is performed again. Best fit parameters (intercept and gradient) are stored, and the resampling is repeated to build up a probability distribution for each parameter and hence uncertainties. The algorithm for this method is described in Appendix A.2.

Each method is performed independently for carbonate and organic carbon data, thereby bootstrapping distributions for the gradient and intercept for each data set. These distributions are in turn sampled to obtain a realization of $f_{org}(t)$. Finally, a linear regression through this $f_{org}(t)$ realization is performed, and this process is repeated to build up a distribution for the trend in f_{org}. The trend in f_{org} with uncertainty can then be compared to the null hypothesis of zero trend to determine if there is a statistically significant change in fractional organic burial over Earth history.

3.2.3) Mean difference analysis.—It is known independently of carbon isotope data that there have been transitional periods in Earth's redox-history whereby atmospheric oxygen has increased dramatically, the Great Oxidation Event (GOE) at ~2.4 Ga (Catling, 2014; Farquhar and others, 2014) and the second rise of oxygen in the Neoproterozoic (Och and Shields-Zhou, 2012). Changes in Earth's biogeochemical cycles accompanied these redox transitions, and consequently there are sizeable carbon isotope excursions during these transitional periods, the precise causes of which are debated (Melezhik and others, 2013; Canfield, 2014). Given that the modern residence time of O_2 against geologic sinks is relatively short (~2 m.y.), for isolated changes in organic carbon burial to permanently increase steady state atmospheric oxygen, the mean f_{org} of non-transitional periods must change, that is the average f_{org} prior to a rise of oxygen must be less than the mean f_{org} after the rise of oxygen. Hence calculating the mean difference between two non-transitional regions in the carbon isotope record provides a method to test whether f_{org} has permanently increased, and whether this can quantitatively explain oxygen increases. The mean difference approach is complimentary to regression analysis.

The mean difference method can be summarized as follows. The "mean level" was calculated for different subsections of organic carbon and carbonate isotope data.

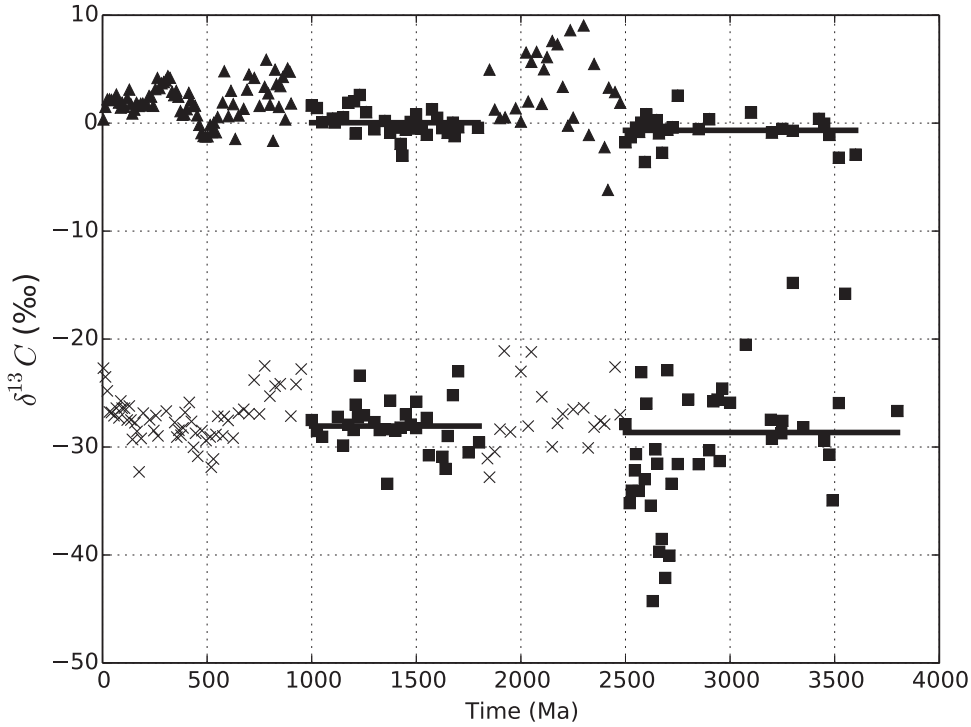

Fig. 2. Binned carbonate (triangles) and organic carbon isotopes (crosses) illustrating the mean-difference methodology. Two regions of interest are defined (1.8-1.0 Ga and 3.8-2.5 Ga) and data in these regions, highlighted by squares, are fitted with a zero-gradient regression to determine mean values. A black horizontal line denotes the mean value in each region. This process is repeated for resampled data to build up distribution for mean values (and f_{org}) within each interval of interest.

These mean levels were determined by a variety of methods such as a simple arithmetic mean or a best fit GLS regression where the regression line was flat because gradient of zero was imposed. The mean levels of different subsections of the data were then compared as described below. The probability distributions for organic and carbonate subsections were also sampled to determine if f_{org} had changed between different time intervals. Figure 2 illustrates the general methodology.

Two different approaches were used to calculate and compare mean levels for the carbon isotope data:

I) Student's t-test. Firstly, two non-transitional intervals of interest were defined, for example 1.8 to 1.0 Ga and 3.8 to 2.5 Ga (fig. 2). The means and standard errors of the carbon isotope data in each interval were calculated, assuming Gaussian distributions, and then the means of each region were quantitatively compared using the Student's t-test. This approach will overstate the significance of differences in mean-levels since it does not account for temporal autocorrelation; however it provides a useful comparison with other methods.

II) a) Bootstrap "flat" GLS regression. Firstly we adopted the autocorrelation model described in section 3.2.1. This model was used to describe the noise, and the bootstrap GLS regression was used to find the mean level for each subsection of the data (gradient specified as zero). The resulting mean-level probability distributions for each subsection of the data were then sampled to make statistical comparisons between different subsections of the data. The approach is essentially identical to the algorithm

used to perform the bootstrap GLS regression above, except that in equation (3.4), $X_{trend}(i)$ is replaced by the constant mean-level X_{mean}, and the regression algebra is modified accordingly since we are only solving for one parameter (mean-level) as opposed to two (gradient and intercept). For completeness the full algorithm is described in Appendix A.3.

b) Identical to IIa) except that persistence times, τ, were fitted to each interval individually rather than fitting the persistence time to the entire data set.

The mean difference methods were used to calculate probability distributions for the mean-level of organic carbon (or carbonate) isotopes for two different intervals of the data. Given these probability distributions, we determined whether the organic carbon (or carbonate) mean-level had changed by sampling from the resulting distributions and building a probability distribution for the difference between the two intervals. To determine whether f_{org} from two different intervals was different, the organic and carbonate isotope distributions in each of the intervals were sampled, realizations of f_{org} were calculated, and a f_{org} distribution was created for each of the two subsections. The two f_{org} distributions corresponding to different intervals of time were then sampled and differenced to determine if there was a statistically significant change in f_{org} between the two intervals. In other words, here we were testing the null hypothesis that the two f_{org} distributions have the same mean. The difference between carbonate and organic carbon isotopic ratios, ε, was also calculated using the same approach.

Four pairs of intervals were chosen for mean level comparison:

a) The late Archean (2.8–2.5 Ga) and the early Archean (3.8–2.8 Ga). This division was chosen to investigate possible changes in $\delta^{13}C_{org}$ and ε associated with proposed methanotrophic recycling of carbon during the late Archean (Hayes and Waldbauer, 2006; Thomazo and others, 2009).

b) The "boring billion" Proterozoic (1.8–1.0 Ga) and the Archean (3.8–2.5 Ga). This division was chosen to investigate changes in steady-state organic burial following the Great Oxidation Event, excluding the 2.2 to 2.06 Ga Lomagundi excursion of very positive $\delta^{13}C_{carb}$ (Schidlowski and others, 1976; Bekker and others, 2008).

c) The "boring billion" Proterozoic (1.8–1.0 Ga) and the early Archean (3.8–2.8 Ga). This division was also chosen to compare pre-GOE and post-GOE steady state organic burial, but excludes the late Archean excursion because of ostensible methanotrophic recycling of carbon.

d) The Phanerozoic (0.54–0 Ga) and the "boring billion" Proterozoic (1.8–1.0 Ga). This division was chosen to examine increased steady-state organic burial following the Neoproterozoic Oxygenation Event (Och and Shields-Zhou, 2012).

3.3) Non-parametric Methods

To complement the parametric models, carbon isotope data were also analyzed using a variety of non-parametric smoothing algorithms to ensure robust conclusions. In particular, the change in f_{org} over Earth history from the parametric analyses above was compared to the same change in f_{org} from non-parametric smoothing. Smoothing algorithms also reveal $f_{org}(t)$ with uncertainties; this is potentially useful for quantifying carbon cycle excursions and validating or parameterizing geological carbon cycle models.

We applied three smoothing methods to the carbon isotope data: locally weighted scatterplot smoothing (LOWESS), kernel regression, and Kalman smoothing. LOWESS is a standard non-parametric smoothing algorithm often applied to unevenly spaced time series data. It utilizes a moving window approach where the bandwidth of the window is a fixed fraction of the total number of data points. For each data point $\{t(i),x(t(i))\}$ a quadratic is fitted to all the data pairs within the surrounding window (each point is weighted by its distance from the center of the window). The smoothed

estimate, $\bar{x}(t)$, is the value of the fitted polynomial evaluated at t. Each data point is then re-weighted by its residual and the procedure is repeated to ensure outliers are discounted. The only free parameter in the LOWESS algorithm is the smoothing bandwidth, that is the fraction of data points used for each polynomial fit. Various heuristic methods exist for selecting an optimal bandwidth; we adopted "leave-out-one cross-validation". The LOWESS algorithm and the cross-validation procedure are described in full in Cleveland (1979). Uncertainties in the smoothed curve were generated by bootstrapping: the residuals from each fit were repeatedly resampled and added back to the smoothed curve to generate an ensemble of smoothed curves.

The second smoothing method adopted is Kernel regression using the Nadaraya-Watson estimator (Fan and Yao, 2003). This is a moving-average scheme that estimates the conditional expectation value of the dependent variable with a Gaussian weighting function. The bandwidth of the weighting function is a free parameter, and this was once again optimized using leave-out-one cross validation. Uncertainties were generated by bootstrap resampling and refitting. Both LOWESS and Kernel regression algorithms are standard functions in the statsmodels module for python, the programming language that was used for these analyses. We provide the python code used to perform this smoothing analysis as supplementary material (http://earth.geology.yale.edu/~ajs/SupplementaryData/2015/KrissansenTotton), and they are mirrored on the website of the senior author (DCC).

The third smoothing method adopted is the Kalman smoother (Shumway and Stoffer, 2011). We implemented a local level state-space model whereby the state of the system at time t, x_t, represents the globally averaged isotopic ratio and evolves stochastically according to:

$$x_{t+1} = x_t + w_t \tag{3.5}$$

Here, w_t is the process noise that is interpreted as representing real changes in the carbon cycle. It is a normally distributed random variable with a mean of zero and unknown variance. The Kalman smoother represents imperfect observations of the isotope record with:

$$y_t = x_t + z_t \tag{3.6}$$

Here, y_t is the observed isotope value and z_t is a normally distributed random variable with a mean of zero and unknown variance. z_t is interpreted as the measurement noise, which is any source of noise that causes a difference between observed $\delta^{13}C$ values and the true globally averaged $\delta^{13}C$ values at the time of deposition. Possible sources of measurement noise include preservation or sampling biases and diagenetic modifications. A Gaussian distribution is justified on the grounds that there are many sources of error and that therefore the Central Limit Theorem applies. The Kalman smoother fits this model to the data using a Bayesian framework, optimizing the variance of both the process noise and the measurement noise. Equation (3.6) can be modified to account for missing observations in otherwise evenly spaced data. This allows unevenly spaced data to be analyzed using the Kalman smoother; carbon isotope data were averaged into 10 my bins and the empty bins were taken to be missing observations (the analyses were repeated using 50 my bins with negligible difference in results). Equation (3.5) was also modified to include a variable trend term to improve the fit provided by a simple random walk.

Strictly speaking the Kalman smoother is a parametric model. However it is useful for comparison with non-parametric smoothers because of its theoretical optimality. It can be shown that the Kalman smoother provides estimates of the state vector with expectation values that match the true expectation values and minimize variance (Simon, 2001). The Kalman smoother was implemented using the KFAS package in

the statistical freeware 'r' (where KFAS documentation provides information on algorithms). Uncertainties in smoothed curves were generated using KFAS simulations of the best fit state-space parameters; an ensemble of smoothed curves was generated to obtain confidence intervals.

4) RESULTS

All the results reported in this section assume that the carbon cycle is in steady state. This means that the isotopic composition of outgassed carbon into the surficial reservoir is constant ($-5‰$), and that the simple mass balance representation of the carbon cycle in equation (1.2) is valid.

4.1) History of Organic Burial and Biological Fractionation Over Time

The carbonate and organic $\delta^{13}C$ time series were smoothed independently using all three smoothing techniques and the resulting ensemble of curves were sampled to produce a distribution of smoothed f_{org} curves. Figures 3A and 3B show smoothed carbonate and organic carbon curves, respectively, for all three smoothers with 95 percent confidence intervals. The three different smoothing techniques produce consistent results. The smoothed carbonate curve(s) provide confirmation of positive $\delta^{13}C_{carb}$ excursions in the Paleoproterozoic (the Lomagundi excursion) and Neoproterozoic. The smoothed organic curves are relatively constant within uncertainties except for the negative excursion around 2.7 Ga.

Figure 3C shows the smoothed f_{org} curves from the three smoothing algorithms with 95 percent confidence intervals. Once again the three algorithms produce congruent results. The Kalman smoother produces slightly larger uncertainties - especially in the Archean - since it incorporates a scheme for missing data and has a more robust method of error calculations than mere resampling of residuals. Neoproterozoic and Paleoproterozoic excursions in the carbonate data appear as pulses of organic burial in the smoothed f_{org} reconstruction. Outside of these excursions the smoothed curves suggest that there is little change in steady-state f_{org} from the Archean to the "boring billion", but that there is an increase in f_{org} from the Proterozoic to the Phanerozoic. These changes in steady-state f_{org} are quantified in the section 4.3 below.

The sensitivity of the smoothed f_{org} curves to different assumptions was briefly investigated. It makes no difference whether the LOWESS and kernel regression uncertainties are generated by simple resampling and refitting (effectively a white noise model), or by applying the autoregressive noise model described in section 3.2; smoothed f_{org} curves and confidence intervals are virtually identical in both cases. This suggests that there is no autocorrelation in the post-smoothing noise which justifies the application of the white-noise approach. Rather than assuming a constant variance noise model we also fitted a LOWESS curve to the squared residuals and scaled the noise terms accordingly to better capture the changes in variance over time. This is a somewhat *ad hoc* approach, but arguably better captures the fact that carbon isotope measurements are well constrained in the Phanerozoic compared to the Proterozoic and Archean. Modifying the uncertainties in this way increases the mean-value and confidence intervals for the total change in f_{org} over Earth history, but once again the differences are slight.

Introducing uncertainty in the independent variable also does not change the smoothed results appreciably. Gaussian time-scale noise with a standard deviation of 100 m.y. was added to the data and new uncertainties were generated by bootstrapping as described in section 4.2.2. This damped the f_{org} fluctuations slightly but otherwise had not effect. Finally, excluding Phanerozoic data from the analysis did not change the results appreciably.

Our reconstruction of f_{org} over Earth history, with uncertainties, could be used as input for (or validation of) dynamical models of the rise of oxygen. We have

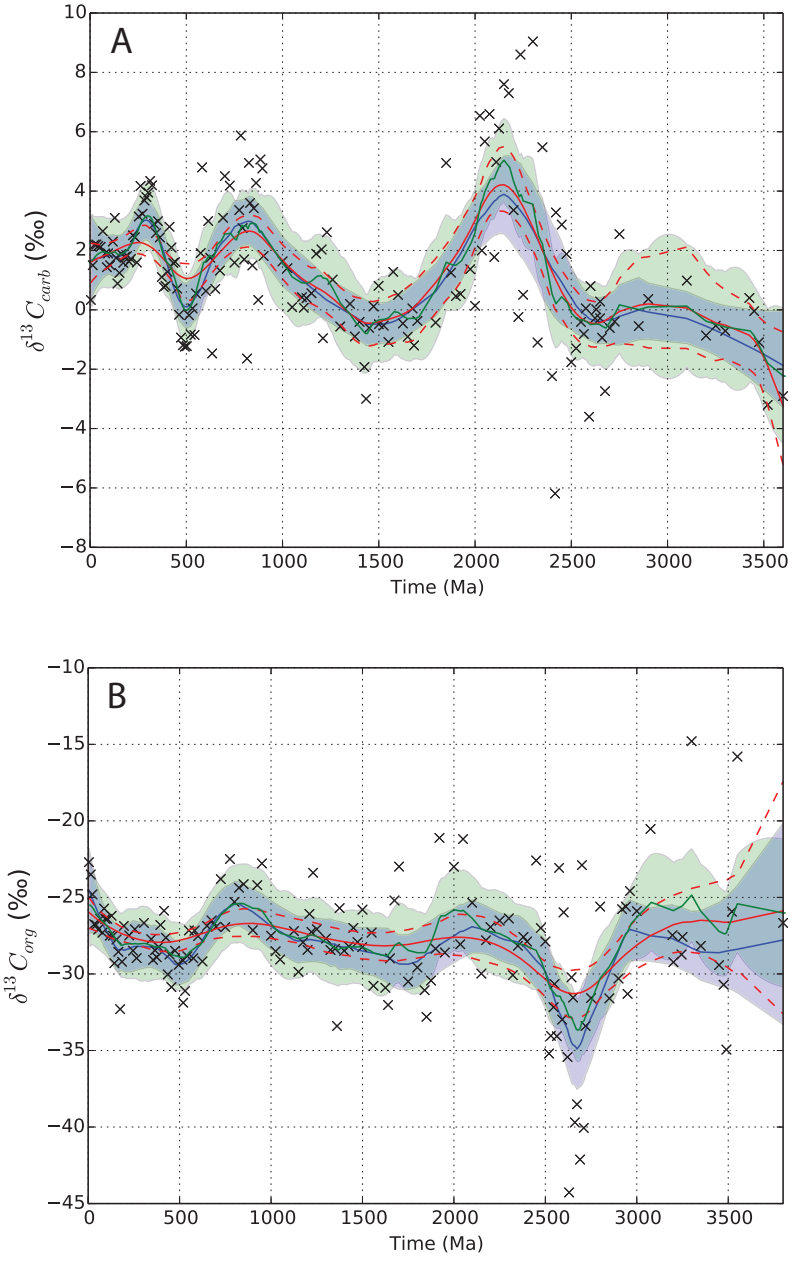

Fig. 3. LOWESS and Kernel regression smoothed curves are denoted by blue and red solid lines respectively, with 95% confidence denoted by blue shading and red dashed lines, respectively. The Kalman smoother curve is the green solid line, with 95% confidence shown with green shading. (A) Shows carbonate results with binned (10 my) carbonate data denoted by black crosses, (B) shows organic results with binned (10 my) organic data denoted by black crosses, (C) shows the resulting f_{org} curves, and D) shows the resulting ε curves. The black dashed line in (C) shows the previously accepted canonical value for fractional organic burial of 0.2.

included the python scripts used to generate the smoothed curves as supplementary material (http://earth.geology.yale.edu/~ajs/SupplementaryData/2015/KrissansenTotton).

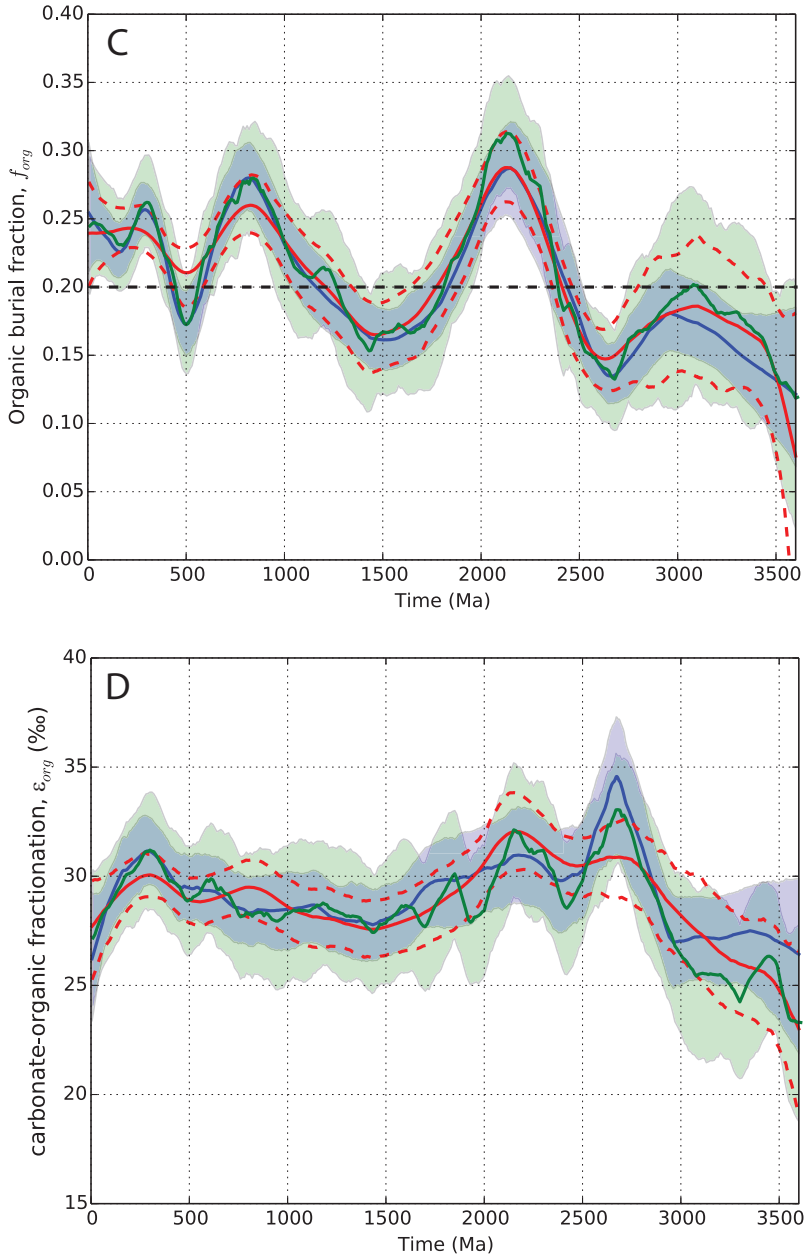

Fig. 3. (continued)

Figure 3D shows the smoothed ε curve over Earth history (where $\varepsilon = \delta^{13}C_{carb} - \delta^{13}C_{org}$). The difference, ε, is largely determined by biological fractionation during carbon fixation and subsequent pathways of remineralization of organic carbon, although changes in climate or carbon dioxide abundance could result in small variations in ε (Hayes, 1994). Compared to the fractional organic burial record, there has been comparatively less variation in the fractionation between organic and

TABLE 1

Total linear change in $\delta^{13}C_{org}$ and $\delta^{13}C_{carb}$ (‰) over Earth history (3.6-0 Ga) using 10 my binned isotope data. Uncertainties in GLS methods generated from bootstrapping (4000 iterations)

Method	Organic		Carbonate	
	Change in $\delta^{13}C_{org}$ (w. 95%)	P value	Change in $\delta^{13}C_{carb}$ (w. 95%)	P value
I) OLS	2.17 [-0.02,4.38]	0.05	2.32 [0.99,3.64]	0.0005
II) Classical GLS	0.83 [-2.11,3.76]	0.58	2.39 [0.42,4.36]	0.02
III) Bootstrap GLS	0.81 [-2.19,3.86]	0.58	2.38 [0.08,4.59]	0.04

inorganic carbon. Visually, it appears as though there is an increase in fractionation from the early Archean to late Archean, followed by a decline in fractionation from the late Archean to the Paleoproterozoic and no further significant changes through to the Phanerozoic. These results are quantified in section 4.3.

4.2) Total Change in Fractional Organic Burial over Earth History

Here, rather than estimating f_{org} as a function of time we calculate the overall change in f_{org} over Earth history. This is achieved by fitting a linear regression to both the carbonates and organics, and multiplying the consequent slope in f_{org} by the duration of the rock record. This absolute change in f_{org} can be compared to the fitted f_{org} value at 3.6 Ga to get the relative increase over Earth history, which equals the fitted modern value of f_{org} divided by the fitted 3.6 Ga value of f_{org}. This quantity is useful for determining whether increases in organic burial can explain the transition from an anoxic to oxic atmosphere, as explored in section 6.

Regression analyses suggest that fractional organic burial increased over Earth history, but the magnitude of this increase has a large spread centered around 1.5 (increase relative to 3.6 Ga). Tables 1 and 2 show the results for the three parametric regression methods outlined in section 3.2.2. The bootstrap GLS regression leads us to conclude with 95 percent confidence that the increase in f_{org} over Earth history is between 0.008 (factor of 1.04 increase relative to 3.6 Ga) and 0.139 (factor of 2.10 increase relative to 3.6 Ga). This is a very large confidence interval: the lower bound would constitute a negligible increase in the oxygen source flux, whereas the upper

TABLE 2

Total linear change in f_{org} over Earth history (3.6-0 Ga) from a variety of statistical techniques

Method	Change in f_{org} (w. 95%)	Change in f_{org} relative to the 3.6 Ga (w. 95% confidence)	P value
I) OLS	0.078 [0.039,0.119]	1.50 [1.19,1.99]	<0.0001
II) Classical GLS	0.073 [0.012,0.139]	1.50 [1.06,2.43]	0.01
III) Bootstrap GLS	0.072 [0.008,0.139]	1.45 [1.04,2.08]	0.03
LOWESS	0.096 [0.056,0.132]	1.64 [1.32,2.04]	<0.0001
Kernel regression	0.089 [0.048,0.132]	1.57 [1.26,2.03]	<0.0001
Kalman smoother	0.081 [0.038,0.126]	1.49 [1.20,1.93]	<0.0001

Rows 1-3 below the header row give results for the parametric regression methods described in section 3.2.2, and rows 4-6 gives results for smoothing methods described in section 3.3. LOWESS and Kernel regression uncertainties were generated using bootstrap resampling (10,000 iterations) with no time-scale uncertainty. Data were filtered and averaged into 10 my bins prior to all analyses.

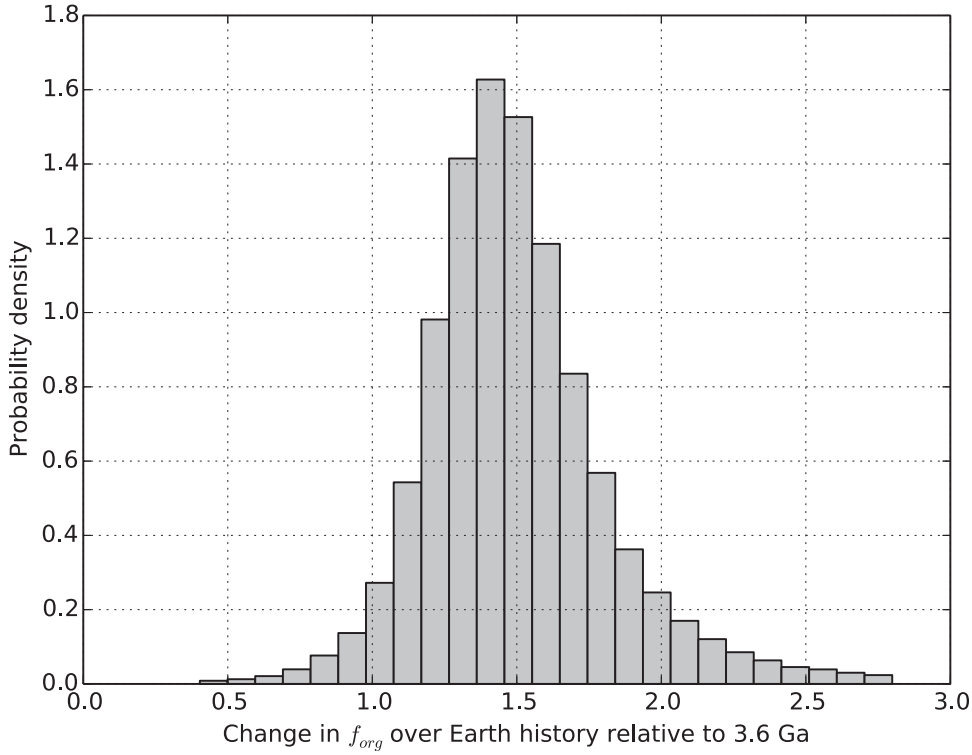

Fig. 4. Probability distribution for the change in the fraction of carbon buried as organics, f_{org}, over Earth history (relative to 3.6 Ga) from the bootstrap GLS regression.

bound would approximately double the source flux. The probability distribution for the change in f_{org} over Earth history from the GLS bootstrap is shown in figure 4. Even if we ignore autocorrelation and assume the carbon isotope data are temporally independent (OLS) then the 95 percent confidence interval still ranges from a factor of 1.19 to 1.99 increase in f_{org} relative to 3.6 Ga. The low p-values for the change in f_{org} merely indicate that f_{org} probably hasn't decreased given this carbon cycle model. This result, that a wide spread of possible f_{org} increases are possible, is robust to data selection and filtering. If the analysis is repeated but Phanerozoic data is excluded, then the uncertainty in the change in f_{org} increases (GLS bootstrap 95% confidence interval ranges from a factor of 0.96 to 2.5 change relative to 3.6 Ga). If the analysis is repeated for unfiltered carbon isotope data then the GLS bootstrap 95 percent confidence interval for the increase in f_{org} extends from 1.02 to 2.43 relative to 3.6 Ga. Tables showing the analysis results for different data selection criteria and filtering are available in Appendix B.

Introducing autocorrelation does not dramatically change the results compared to a simple OLS regression, thereby indicating that there is minimal autocorrelation in the carbon isotope data. The confidence intervals are slightly wider for the GLS results, but OLS regression analysis is sufficient to provide a similar answer. However it was necessary to run GLS fits to prove this.

4.2.1) Comparison with smoothing algorithms.—In order to directly compare the results from the regression analyses in the section 3.2 with the smoothing algorithms in section 3.3, the change in f_{org} over Earth history (with confidence intervals) was

calculated for each of the smoothing techniques. This was achieved by taking the ensemble of f_{org} curves from each of the smoothing algorithms and producing a corresponding ensemble of linear regressions. Thus a distribution for the linear change in f_{org} over Earth history was obtained and compared with equivalent parametric results.

These results are summarized in table 2 (rows 4-6). Both the smoothing algorithms and the parametric regression techniques produce a similar f_{org} change over Earth history (factor of 1.5 increase relative to 3.6 Ga) with similar 95 percent confidence intervals (approximately 1.2 to 2.0 relative to 3.6 Ga). The consistently low p-values suggest that some positive increase in f_{org} has probably occurred, but it is clear from the confidence intervals that the magnitude of this increase is poorly constrained. The parametric regression techniques generally produce larger uncertainties than the smoothing techniques; this is unsurprising since the smoothing of data reduces extremes compared to unsmoothed linear regressions.

4.2.2) Timescale uncertainty.—In the analysis above we assumed that there were no uncertainties in the carbon isotope sample ages. This assumption is unrealistic. Although the radiometric ages of dated strata are tightly constrained, the stratigraphic distance between datable layers and carbon-bearing layers is often large, and thus interpolation is necessary, thereby introducing considerable uncertainty in ages. Errors in the database are also likely to exist and introduce further uncertainty in nominal ages. Rather than attempt the task of estimating the uncertainty in the radiometric age for every point in the carbon isotope database, time-axis uncertainty was introduced by adding Gaussian noise to each radiometric date, and then bootstrapping over many different realizations of the time-uncertain data to generate probability distributions for parameters of interest. Adding noise in this way is undoubtedly an oversimplification because some samples have well constrained ages, whereas other samples have uncertainties of tens to hundreds of millions of years. However this approach provides some indication of the sensitivity of the results to uncertainties in the formation ages. The alterations to the algorithms made to account for timescale uncertainty are described in Appendix A.4.

Because the choice for the width of the Gaussian used to introduce noise to the time values is unclear, the analysis was repeated for a range of Gaussian widths. Figure 5 shows the change in f_{org} over Earth history (relative to 3.6 Ga) as a function of t_{noise}, where t_{noise} is the standard deviation of the Gaussian (in millions of years) used to add uncertainty to the nominal ages of carbon isotope samples. Evidently, adding a small amount of noise to the nominal carbon isotope ages causes a sharp increase in the uncertainty in the f_{org} change with GLS bootstrapping. This is because for $t_{noise} = 0$ there are many closely spaced observations due to preservation biases, which imply low τ values, a larger effective sample size, and therefore smaller uncertainties in the final parameters. However when t_{noise} is increased there are fewer closely spaced observations, the τ estimate is larger, effective sample sizes are smaller, and final uncertainties are larger. The initial sharp increase in bootstrap GLS confidence intervals as t_{noise} increases perhaps suggests that the true autocorrelation is low, and that by adding time-axis noise we are artificially introducing autocorrelation by increasing the spacing between data points.

As t_{noise} is increased further, the autocorrelation is destroyed entirely and the uncertainties converge toward the OLS confidence interval. A similar decay in uncertainties is visible in the classical GLS confidence intervals (though the initial sharp increase is absent for reasons that are unclear). For reasonable t_{noise} magnitudes the mean value for the change in f_{org} is constant. This indicates that it is only the autocorrelation structure that is being altered by adding time-scale uncertainty.

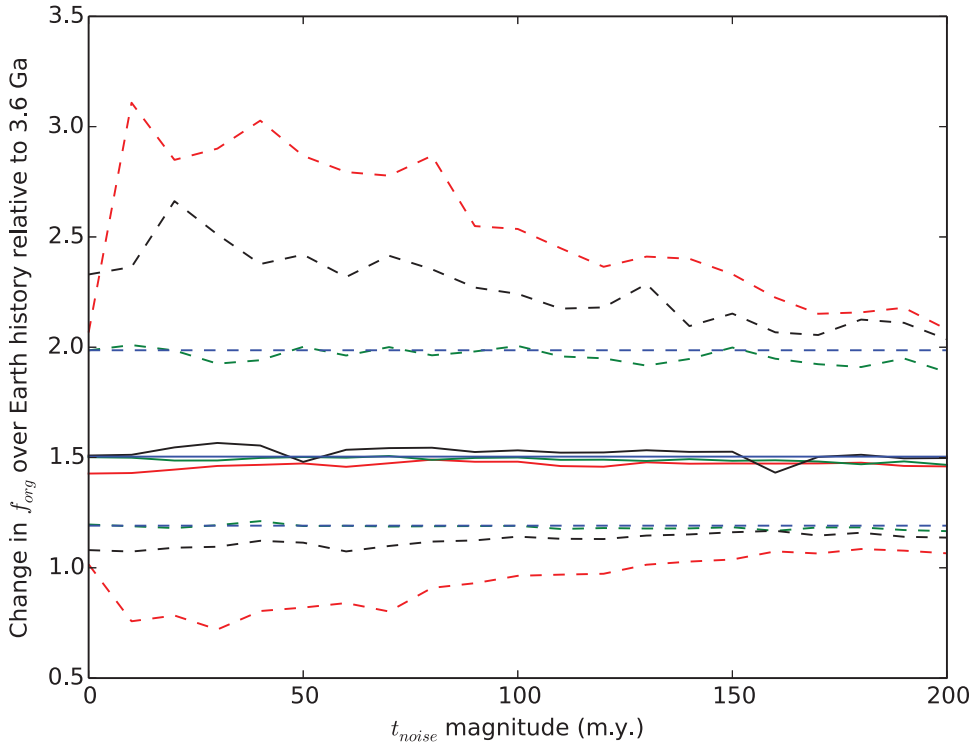

Fig. 5. Change in f_{org} over Earth history relative to 3.6 Ga as a function of t_{noise}, where t_{noise} is the standard deviation of the Gaussian (in millions of years) used to add uncertainty to the nominal ages of carbon isotope samples. Solid lines are median changes and dashed lines denote 95% confidence intervals for different methods. Median rather than mean changes are used to reduce fluctuations due to extreme increases or decreases. The blue line is a reference line that shows the OLS regression result, which does not account for timescale uncertainty in any way. The green line is the OLS regression modified to account for timescale uncertainty (method I, section 3.2.2), the black line is the classical GLS regression modified to account for timescale uncertainty (method II, section 3.2.2), and the red line is the GLS bootstrap regression modified to account for timescale uncertainty (method III, section 3.2.2).

However, for very large values of t_{noise} the mean change in f_{org} does approach zero as we would expect (not shown).

Unsurprisingly, the confidence interval for the OLS, zero time-uncertainty reference case (blue line) is essentially identical to the OLS regression confidence interval that incorporates time-uncertainty (green line). Again, this indicates that adding t_{noise} does not change the confidence intervals due to changes in the data points themselves, but rather it changes the autocorrelation structure of the data.

4.3) Change in f_{org} and ε Between Key Geological Time Periods

Table 3 shows the results of mean difference comparisons between the four intervals defined in section 3.2.3. For each sub-table, the first six rows give the mean level with standard error for two time intervals and three different methods. The seventh to ninth rows give the mean difference between the two intervals for the three methods, and the tenth to twelfth rows give the corresponding p-values for the mean level differences (the null hypothesis being that there is no difference between mean levels). The analysis is reported separately for organic carbon isotopes, carbonate isotopes, f_{org} and ε (four columns).

TABLE 3

Mean difference comparisons between four non-transitional pairs of intervals in Earth's history

		$\delta^{13}C_{org}$ (‰)	$\delta^{13}C_{carb}$ (‰)	f_{org}	$\varepsilon = \delta^{13}C_{carb}$ - $\delta^{13}C_{org}$ (‰)
A. Late Archean (2.8-2.5) vs Early Archean (3.8-2.8 Ga)					
Mean level 2.8-2.5 Ga	t-test	-32.9±1.3	-0.59±0.36	0.137±0.011	32.3±1.4
	GLSa	-32.0±2.4	-0.69±0.34	0.139±0.014	31.3±2.5
	GLSb	-32.7±1.3	-0.70±0.35	0.135±0.011	32.0±1.3
Mean level 3.8-2.8 Ga	t-test	-26.7±1.1	-0.74±0.41	0.164±0.015	26.0±1.2
	GLSa	-26.2±1.2	-0.72±0.37	0.139±0.015	25.4±1.3
	GLSb	-26.1±1.2	-0.93±1.7	0.158±0.059	25.1±2.1
Mean level difference [95% confidence]	t-test	-6.23 [-9.55,-2.91]	0.15 [-0.92,1.23]	**-0.028 [-0.06,0.009]**	**6.38 [2.87,9.88]**
	GLSa	-5.90 [-11.1,-0.64]	0.021 [-0.94,1.06]	**-0.030 [-0.069,0.011]**	**5.90 [0.44,11.47]**
	GLSb	-6.59 [-10.1,-3.20]	0.24 [-2.79,3.97]	**-0.023[-0.12,0.12]**	**6.85 [2.20,11.20]**
P value	t-test	0.0006	0.78	**0.14**	**0.00028**
	GLSa	0.026	0.99	**0.15**	**0.03**
	GLSb	0.0002	0.95	**0.56**	**0.0020**
B. Middle Proterozoic (1.8-1.0 Ga) vs Archean (3.8-2.5 Ga)					
Mean level 1.8-1.0 Ga	t-test	-28.1±0.4	0.04±0.21	**0.179±0.007**	**28.1±0.5**
	GLSa	-28.1±0.6	0.08±0.20	**0.181±0.007**	**28.3±0.7**
	GLSb	-28.0±0.46	0.19±0.32	**0.184±0.010**	**28.2±0.6**
Mean level 3.8-2.5 Ga	t-test	-29.9±1.0	-0.65±0.26	**0.149±0.009**	**29.3±1.0**
	GLSa	-28.6±1.4	-0.71±0.25	**0.154±0.011**	**27.8±1.4**
	GLSb	-29.4±1.1	-0.71±0.25	**0.150±0.009**	**28.6±1.1**
Mean level difference [95% confidence]	t-test	1.86 [-0.24,3.94]	0.69 [0.03,1.35]	**0.031 [0.008,0.053]**	**-1.16 [-3.36,1.05]**
	GLSa	0.52 [-2.49,3.35]	0.79 [0.18,1.43]	**0.026 [0.0007,0.051]**	**0.45 [-2.65,3.64]**
	GLSb	1.33 [-0.99,3.64]	0.90 [0.10,1.70]	**0.034 [0.007,0.060]**	**-0.39 [-2.88,2.09]**
P value	t-test	0.08	0.04	**0.0074**	**0.30**
	GLSa	0.72	0.011	**0.045**	**0.79**
	GLSb	0.26	0.03	**0.014**	**0.75**

TABLE 3
(continued)

		$\delta^{13}C_{org}$ (‰)	$\delta^{13}C_{carb}$ (‰)	f_{org}	$\varepsilon = \delta^{13}C_{carb} - \delta^{13}C_{org}$ (‰)
C. Middle Proterozoic (1.8-1.0) vs Early Archean (3.8-2.8 Ga)					
Mean level 1.8-1.0 Ga	t-test	-28.1±0.4	0.04±0.21	0.179±0.007	28.1±0.5
	GLSa	-28.1±0.6	0.09±0.21	0.181±0.007	28.2±0.7
	GLSb	-28.0±0.46	0.18±0.32	0.184±0.010	28.2±0.6
Mean level 3.8-2.8 Ga	t-test	-26.7±1.1	-0.74±0.41	0.164±0.015	26.0±1.2
	GLSa	-26.2±1.2	-0.72±0.39	0.169±0.015	25.4±1.3
	GLSb	-26.1±1.3	-0.91±1.6	0.159±0.060	25.2±2.1
Mean level difference [95% confidence]	t-test	-1.36 [-3.63,0.89]	0.78 [-0.12,1.68]	**0.015 [-0.017,0.047]**	**2.14 [-0.29,4.61]**
	GLSa	-1.91 [-4.71,0.59]	0.80 [-0.009,1.720]	**0.012 [-0.020,0.046]**	**2.81 [0.04,5.73]**
	GLSb	-1.94 [-4.64,0.57]	1.08 [-1.83,4.75]	**0.024 [-0.069,0.162]**	**3.03 [-1.02,7.59]**
P value	t-test	0.24	0.10	**0.35**	**0.08**
	GLSa	0.15	0.053	**0.47**	**0.05**
	GLSb	0.14	0.48	**0.71**	**0.16**
D. Phanerozoic (0.54-0 Ga) vs Middle Proterozoic (1.8-1.0 Ga)					
Mean level 0.54-0 Ga	t-test	-27.8±0.3	1.75±0.19	0.228±0.006	29.6±0.4
	GLSa	-27.6±0.4	1.72±0.19	0.230±0.006	29.2±0.5
	GLSb	-27.5±0.5	1.35±0.70	0.220±0.019	28.8±0.9
Mean level 1.8-1.0 Ga	t-test	-28.1±0.4	0.04±0.21	0.179±0.007	28.1±0.5
	GLSa	-28.1±0.6	0.09±0.21	0.181±0.007	28.3±0.7
	GLSb	-28.0±0.5	0.18±0.32	0.184±0.010	28.2±0.6
Mean level difference [95% confidence]	t-test	0.26 [-0.75,1.27]	1.71 [1.15,2.27]	**0.049 [0.032,0.066]**	**1.46 [0.32,2.60]**
	GLSa	0.50 [-0.93,1.92]	1.63 [1.06,2.19]	**0.049 [0.030,0.067]**	**0.92 [0.71,2.62]**
	GLSb	0.53 [-0.87,1.92]	1.68 [-0.36,2.69]	**0.036 [-0.008,0.078]**	**0.58 [-1.47,-2.67]**
P value	t-test	0.62	<0.0001	**<0.0001**	**0.01**
	GLSa	0.48	<0.0001	**<0.0001**	**0.28**
	GLSb	0.45	0.13	**0.10**	**0.59**

For each interval pair (A, B, C, and D), the mean level with standard error for both intervals is listed for organic carbon, carbonates and fractional organic burial (rows 1 to 6 below the header row). The difference between the two intervals with 95% confidence is shown in rows 7 to 9, with corresponding p-values in rows 10 to 12. The analysis is repeated for the three different methods described in the section 3.2.3, hence there are 12 rows rather than 4. The mean difference in fractional organic burial between the two intervals, with confidence intervals and p-values, is highlighted in bold.

We also directly compared the mean-difference results from the parametric analyses in section 3.2.3 to the smoothing analysis. We adopted a heuristic method for comparing the mean-difference between two sections of smoothed data since the number of independent measurements is unknown. For a given time interval the distribution for the mean-level was obtained from the ensemble of (means of) smoothed curves within that interval. Complete tables comparing mean-difference results from both parametric analyses and smoothing analyses are available in Appendix B.3. Key results are briefly summarized here.

4.3.1) Early Archean to late Archean.—Both the Student's t-test and the GLS bootstrap regression indicate a significant decrease in $\delta^{13}C_{org}$ from the early Archean to the late Archean (table 3A). This result is robust to the precise date chosen to divide the early and late Archean; the difference is statistically significant for any division date between 2.7 and 3.0 Ga. The mean difference between the early and late Archean is maximized when the division is placed at 2.78 or 2.90 Ga, where the former correlates with a change in sulfur cycling (Stüeken and others, 2012) and the latter roughly coincides with a glaciation episode (Young and others, 1998).

Despite the decline in $\delta^{13}C_{org}$ there is no statistically significant decrease in f_{org} from the early Archean to the late Archean. This is confirmed by both the parametric methods above and smoothing analysis (Appendix table B.4A). There is, however, a statistically significant increase in ε by about 6 permil from the early to late Archean. This change is apparent in figure 3D, and has also been quantified by mean-difference analysis as shown in table 3A. These results suggest that changes in the carbon isotope record from the early to late Archean are best explained by changes in biological fractionation between inorganic and organic carbon, and not by changes in fractional organic burial.

4.3.2) Archean to Proterozoic.—There is a marginally significant increase in f_{org} from the Archean (3.8-2.5 Ga) to the Proterozoic (1.8-1.0 Ga), although the magnitude of the increase is not well constrained by either the Student's t-test or the GLS bootstrap regression (table 3B). The mean (absolute) increase in fractional organic burial from the Archean to the Proterozoic is ~0.03, with the 95 percent confidence interval ranging from approximately 0 and 0.05. Smoothing analysis also confirms this result (table B.4B). One way of testing whether f_{org} increased from the Archean to the Proterozoic is to exclude the late Archean and to difference only the early Archean and the Proterozoic (table 3C). This is because, as established above, there is a marked decrease in $\delta^{13}C_{org}$ in the late Archean plausibly due to methanotrophic-recycling. When only the early Archean is considered, there is clearly no statistically significant difference between pre-GOE f_{org} and post-GOE f_{org}. Smoothing analysis confirms this result (table B.4C). However, the 95 percent confidence interval for the change in f_{org} between these two levels is quite large in absolute terms, and so sizeable changes in f_{org} could exist buried within the noise.

There is no statistical difference in ε between the Archean and the Proterozoic. However, this is only because the low values of ε in the early Archean cancel out the high values in the late Archean. If the late Archean and Proterozoic are compared, we find that there is a statistically significant decrease in ε to levels marginally greater than the early Archean (not shown). This is consistent with figure 3D.

4.3.3) Proterozoic to Phanerozoic.—Both mean difference analysis (table 3D) and smoothing analysis (table B.4D) indicate that there is a statistically significant and well-constrained increase in f_{org} from the Proterozoic to the Phanerozoic. The mean absolute increase in f_{org} from the Proterozoic to the Phanerozoic is ~0.04 with a 95 percent confidence interval ranging from approximately 0.03 to 0.065. In short, the carbon isotope record suggests that there is a real increase in organic burial temporally correlated with the Neoproterozoic rise of oxygen. This challenges the widely held

belief that the carbon isotope record doesn't support secular changes in organic carbon burial. It is unclear as to whether there is a change in ε from the Proterozoic to Phanerozoic; the different methods reported in table 3D produce conflicting results.

All the mean-difference results described above are robust to changes in data binning and filtering. If unfiltered carbon isotope data are used, then all the key results above are unchanged, except that there appears to be a statistically significant (albeit poorly constrained) increase in f_{org} from the early to late Archean. If only non-unique valued data points are binned (as opposed to 10 m.y. bins) then the key results are also unchanged, though there is some variation in confidence intervals and p-values (not shown). Adding noise to the time-axis increases the size of the confidence intervals for the mean-level change in f_{org} by approximately a factor of 2 (not shown).

4.4) Results Summary

Trend analysis implies that f_{org} increased over Earth history and that the null hypothesis of constant f_{org} can be rejected. However, the magnitude of this increase has a large range. The simplest regression, OLS, which does not account for timescale uncertainty or autocorrelation, gives a mean change in f_{org} over Earth history of 1.5 (relative to 3.6 Ga), with a broad 95 percent confidence interval that ranges from 1.19 to 1.99 (relative to 3.6 Ga). Accounting for autocorrelation and/or timescale uncertainty widens this confidence interval. There is, however, a statistically significant increase in fractional organic burial from the Proterozoic to the Phanerozoic, and a significant increase in ε from the early Archean to late Archean which is not accompanied by a change in f_{org}. The analysis above assumes that the carbon cycle is in steady state, that the isotopic composition of outgassed carbon into the surficial reservoir is constant (−5‰), and that the simple mass balance representation of the carbon cycle in equation (1.2) is valid. In the next section we explore how these results change when we consider a more complex carbon cycle model.

5) complex mass balance model

The mass-balance model used to calculate f_{org} in the analysis above is a simplification of the geological carbon cycle with only two isotopically distinct carbon burial sinks: organic and carbonate sediments. However, in practice there are at least two other distinct sinks, which may or may not be important: the burial of authigenic carbonates and carbonate burial via oceanic crust carbonatization (OCC). Authigenic carbonates are isotopically distinct from allogenic (sedimentary) carbonates because their carbon is primarily remineralized organic matter (Schrag and others, 2013). Oceanic crust carbonates are potentially isotopically distinct from sedimentary carbonates because the sedimentary carbonates mostly form near the surface (above the carbonate compensation depth), whereas oceanic crust carbonatization occurs at depth or below the sea floor. If there is an isotopic gradient in the dissolved inorganic carbon from the surface to the sea floor, then the two carbonate sinks will be distinct (Bjerrum and Canfield, 2004). The question arises as to whether a more complex isotopic mass balance could change our results.

We can consider a modified version of the mass-balance model [eq (1.2)], which includes the additional isotopically distinct carbonate sinks described above. The new mass-balance equation is given by:

$$\delta^{13}C_{in} = f_{org}\delta^{13}C_{org} + (1-\alpha-\beta)(1-f_{org})\delta^{13}C_{carb} + \underbrace{\alpha(1-f_{org})\delta^{13}C_{ac}}_{\text{authigenic fraction}} + \underbrace{\beta(1-f_{org})\delta^{13}C_{OCC}}_{\text{ocean crust fraction}} \quad (5.1)$$

Here, $\delta^{13}C_{org}$ is the isotopic ratio of organic matter, $\delta^{13}C_{carb}$ is isotopic ratio of allogenic (non authigenic sedimentary) carbonates, $\delta^{13}C_{OCC} = \delta^{13}C_{carb} + \Delta S$ is the isotopic

ratio of oceanic crust carbonates, and ΔS is the difference in dissolved inorganic carbon isotopic ratio between the surface ocean and the seafloor. Also, f_{org} is the fraction of carbon buried as organic carbon, α is the fraction of inorganic carbon deposited as authigenic carbonates, β is the fraction of inorganic carbon deposited during OCC, and $\delta^{13}C_{oc}$ is the isotopic ratio of authigenic carbonates. We also introduce an apparent carbonate isotopic abundance, $\delta^{13}C^{AP}_{carb}$:

$$\delta^{13}C^{AP}_{carb} = \frac{(1-\alpha-\beta)\delta^{13}C_{carb} + \alpha\delta^{13}C_{ac}\lambda}{(1-\alpha-\beta)+\alpha\lambda} \tag{5.2}$$

This is the isotopic composition we observe in buried carbonates, which is some weighted mixture of true allogenic carbonates and authigenic carbonates. The fraction of authigenic carbonate that are preserved in the observable rock record, λ, accounts for the fact that even though authigenic burial may be large, it might not be apparent from the isotope record if allogenic carbonates are preferentially preserved. The range of λ is from zero to one. The denominator in equation (5.2) is simply the sum of the weightings to ensure normalization.

By solving these equations (5.1) and (5.2) we obtain fractional organic burial,

$$f_{org} = \frac{\delta^{13}C_{in} - (1-\alpha-\beta)\delta^{13}C_{carb} - \beta(\Delta S + \delta^{13}C_{carb}) - \alpha\delta^{13}C_{ac}}{\delta^{13}C_{org} - (1-\alpha-\beta)\delta^{13}C_{carb} - \beta(\Delta S + \delta^{13}C_{carb}) - \alpha\delta^{13}C_{ac}} \tag{5.3}$$

where $\delta^{13}C_{carb} = \lambda\alpha(\delta^{13}C^{AP}_{carb} - \delta^{13}C_{ac})/(1-\alpha-\beta) + \delta^{13}C^{AP}_{carb}$ is the actual isotopic value in sedimentary carbonates only, not what is observed.

We investigated how the results of the smoothing analysis changed using this more complex carbon cycle model. The model contains five free parameters that are currently uncertain: α, λ, $\delta^{13}C_{ac}$, β, and ΔS. However it is possible to perform bootstrapping over geologically plausible parameter ranges to reconstruct probabilistic f_{org} histories. The ensemble of $\delta^{13}C^{AP}_{carb}$ and $\delta^{13}C_{org}$ curves from the LOWESS analysis in section 4.1 were used as inputs. To reconstruct f_{org} over Earth history, probability distributions for the five unknown parameters were specified and sampled along with the ensemble of smoothed curves for $\delta^{13}C^{AP}_{carb}$ and $\delta^{13}C_{org}$ to obtain a smoothed estimate of $f_{org}(t)$ with confidence intervals.

Our choice of parameter ranges is based on literature data. The difference between dissolved inorganic carbon isotopes in surface and seafloor water, ΔS, was taken to be a uniform distribution between -2 permil (modern) and 0 permil. There is some evidence from seafloor basalts for constant $\delta^{13}C$ with depth in the Archean ocean (Nakamura and Kato, 2004), that is $\triangle S \approx 0‰$, but the depth of the ocean floor carbonate used to infer this is not well constrained and so large ΔS values cannot be excluded. However, a conservative parameter range was chosen to demonstrate that a large sea floor to surface isotopic gradient is not required to considerably magnify uncertainties in f_{org}. The fraction of carbonates deposited via OCC, β, was taken to be a linear function from some unknown initial Archean value (0.3-0.9 uniform distribution) to the known present value, 0.0. The initial OCC fraction depends strongly on the composition of the Archean seafloor and our chosen parameter range is derived from theoretical estimates by Sleep and Zahnle (2001). It is not yet known whether authigenic carbonates are quantitatively important in the interpretation of the carbon isotope record. It has been argued that the fraction of authigenic carbonates was greater at some times in Earth's history than today (Schrag and others, 2013), although even the modern fraction of authigenic carbonates is not well constrained (Sun and Turchyn, 2014). If there is a preservation bias against authigenic carbonates then the conventional mass balance model will not correctly predict f_{org} from the isotope record, and hence the need for the complex mass balance model described above. We are

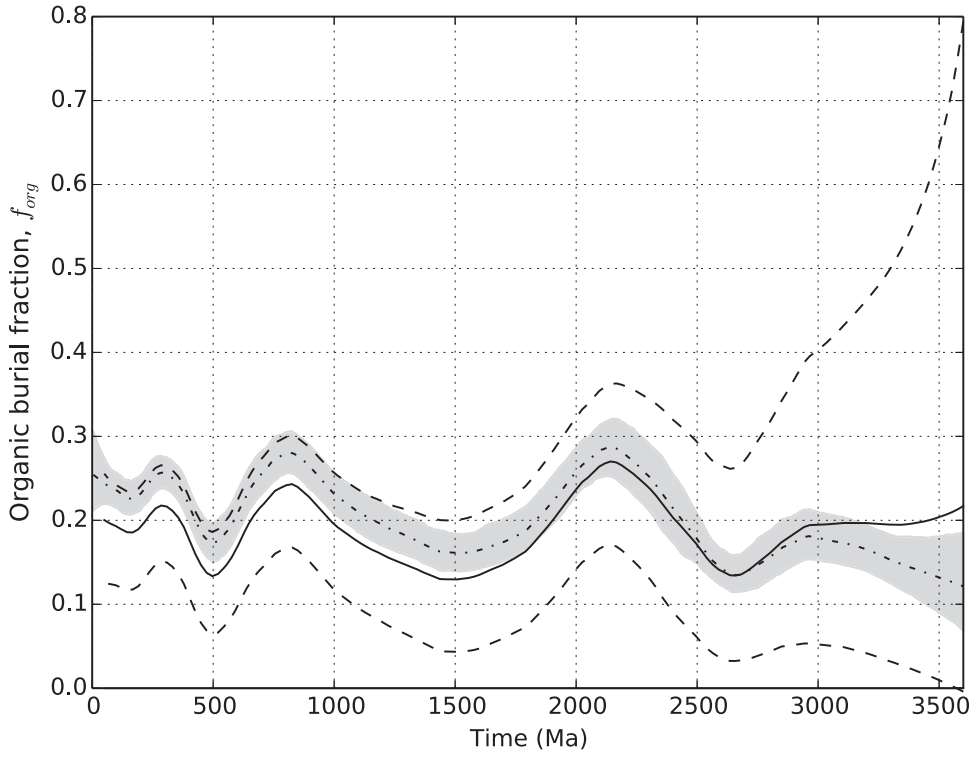

Fig. 6. Smoothed f_{org} (solid line) with 95% confidence intervals (dashed lines) from the updated carbon cycle model and parameter distributions described in the main text. Smoothed f_{org} from the simple carbon cycle model using LOWESS (section 4.1) is denoted by the dot-dash line for comparison, with 95% confidence intervals shaded gray.

(mostly) agnostic as to whether authigenic carbonates are quantitatively important, and simply vary the model parameters describing authigenic carbonates across plausible ranges (loosely speaking from 'authigenic carbonates negligible' to 'authigenic carbonates quantitatively important'). The fraction of authigenic carbonates, α, and the preservation rate of authigenic carbonates, λ, cannot be independently estimated from the isotope record since the carbonate time series is a product of both of these parameters. Heuristic lower bounds can arguably be put on α and λ by inspection of the isotope record; no more than ∼10 percent of Archean data lie below the bulk-Earth composition of -5 permil, although this observation does not constrain each parameter individually. Authigenic fraction α was taken to be a linear function from 0 to 0.3 (uniform distribution) in the Archean to 0.07 to 0.20 (uniform distribution) at the present. The modern value was taken from Sun and Turchyn (2014) and the Archean upper limit was an arbitrary cut-off chosen to ensure the allogenic carbonate fraction was generally non-zero (zero or near-zero allogenic carbonates is unphysical). In the absence of strong constraints the preservation fraction, λ, was sampled from a uniform distribution from 0.1 to 1.0. The authigenic isotopic abundance was set to a constant value of $\delta^{13}C_{ac} = -23‰$ which is the approximate modern value for organic carbon sediments (later it will be shown that results are insensitive to assumptions about $\delta^{13}C_{ac}$).

Figure 6 shows $f_{org}(t)$ with 95 percent confidence intervals given the parameter choices specified above. The smoothed LOWESS curve from figure 3C (simple carbon

cycle model) with 95 percent confidence intervals is also plotted for comparison. The confidence intervals from the new model are much larger; this illustrates the capacity for uncertainties in carbon cycle processes to greatly magnify the uncertainties in organic burial. The mean linear change in f_{org} over Earth history was also evaluated using the model parameters described above and found to be a factor of 1.15 (relative to 3.6 Ga), with 95 percent confidence ranging from 0.29 to 2.44.

Naturally the precise parameter ranges we have chosen are somewhat arbitrary; the result above is more illustrative than definitive. It demonstrates that unless the five carbon cycle parameters can be meaningfully constrained then the carbon isotope record does not constrain either the direction or the magnitude of the change in organic burial over Earth history. The confidence interval for f_{org} in the Archean nominally extends to zero organic burial. Of course, zero organic burial is impossible since we know kerogenous sediments exist from the Archean, but this result indicates that very low fractional organic burial in the Archean cannot be entirely excluded. Alternatively, this result may indicate that the ranges we have chosen for the parameters are too generous. Whatever the explanation, the probability distributions for the change in f_{org} over Earth history have long tails extending to factors of 4 or more (see below). The sensitivity of the uncertainty in f_{org} changes to different parameter assumptions is explored in Appendix C and key findings are summarized in table 4. The remainder of this section can be skipped without loss of continuity by readers not interested in this sensitivity analysis.

In table 4, each row represents one set of parameter choices, and the final two columns give the change in f_{org} over Earth history (both absolute and relative) given these parameter distributions and the ensemble of smoothed LOWESS curves for $\delta^{13}C_{org}$ and $\delta^{13}C^{AP}_{carb}$ from the section 4.1. The first row (italicized) denotes the "best-estimate" parameter choices that generated figure 6. The 13th row (also italicized) denotes the parameter choices for the simple carbon cycle model that generated the smoothed LOWESS curve in figure 3C (and repeated in fig. 6). Rows 2 to 13 explore the sensitivity of the results to different parameter assumptions; where parameters have been changed from their original values in row 1 they are highlighted in bold.

In general we see that the large uncertainty in the change-in-f_{org} over Earth history is not attributable to any one parameter. In all cases except when Archean OCC fraction is low (rows 5 and 8), neither the sign nor magnitude of the change-in-f_{org} over Earth history is determinable from the carbon isotope data. The authigenic preservation parameter λ has a sizeable impact on the change-in-f_{org} confidence intervals (rows 2-4). Thus any geological observations or theoretical arguments that constrain this parameter would greatly refine our knowledge of the geological carbon cycle. Similarly, constraining the inorganic carbonate fractions α (authigenic carbonates) and β (OCC) would also dramatically reduce the uncertainty in organic burial changes (rows 5-8). Determining the magnitude of the Archean ocean isotope gradient would also be desirable, though this is of secondary importance since it seems to only extend the tail of the change-in-f_{org} distribution. Finally, the precise value for the isotopic composition of authigenic carbonates $\delta^{13}C_{ac}$ doesn't influence uncertainties very much so long as it is approximately equal to that of organic carbon.

6) DISCUSSION

The precise extent to which secular changes in f_{org} of differing magnitudes can explain the rise of oxygen is beyond the scope of this study. However by considering absolute fluxes we can draw some conclusions. In steady state the total O_2 oxygen source flux (in Tmol O_2/yr where one Tmol $= 10^{12}$ mol) due to the burial of organic matter is given by:

$$F_{O_2.carbon} = f_{org}(F_{total.carbon.burial}) = f_{org}(F_{volc + meta.CO_2} + F_{weather.CO_2}) \quad (6.1)$$

TABLE 4

Each row represents one choice of parameter ranges. The first row contains the parameter choices used to generate fig. 6 (solid and dashed lines), and the 13th row contains the parameter choices used to generate the original LOWESS smoothing (dot-dash lines and shaded, fig. 6). "rnd(0,1)" in row four denotes that λ was not a constant, but rather a random variable between 0 and 1 for each time-value. Note that β + α < 1 was enforced in cases where the distributions could have yielded total fractions greater than one (distributions were resampled until β + α < 1)

	β		ΔS (‰)	$\delta^{13}C_{ac}$ (‰)	α		λ	Change in f_{org} since 3.6 Ga [95% c.]	Change in f_{org} relative to 3.6 Ga [95% c.]
	Archean	Modern			Archean	Modern			
1	*[0.3,0.9]*	*0.0*	*[-2.0,0.0]*	*-23*	*[0.0,0.3]*	*[0.07-0.20]*	*[0.1,1.0]*	*0.03[-0.37,0.15]*	*1.15[0.29,2.44]*
2	[0.3,0.9]	0.0	[-2.0,0.0]	-23	[0.0,0.3]	[0.07-0.20]	**1.0**	-0.03[-0.47,0.11]	0.88[0.25,1.66]
3	[0.3,0.9]	0.0	[-2.0,0.0]	-23	[0.0,0.3]	[0.07-0.20]	**0.2**	0.06[-0.13,0.19]	1.52[0.56,4.36]
4	[0.3,0.9]	0.0	[-2.0,0.0]	-23	[0.0,0.3]	[0.07-0.20]	**rnd(0,1)**	0.03[-0.27,0.12]	1.20[0.35,2.01]
5	**0.3**	**0.0**	[-2.0,0.0]	-23	[0.0,0.3]	[0.07-0.20]	[0.1,1.0]	0.07[-0.004,0.19]	1.52[1.01,3.49]
6	**0.8**	**0.0**	[-2.0,0.0]	-23	[0.0,0.3]	[0.07-0.20]	[0.1,1.0]	-0.06[-0.45,0.09]	0.74[0.20,1.60]
7	[0.3,0.9]	0.0	[-2.0,0.0]	-23	**0.1**	**0.1**	[0.1,1.0]	0.05[-0.25,0.12]	1.28[0.41,2.33]
8	**0.3**	**0.0**	[-2.0,0.0]	-23	**0.1**	**0.1**	[0.1,1.0]	0.09[0.04,0.13]	1.68[1.29,2.56]
9	[0.3,0.9]	0.0	[-2.0,0.0]	**[-28,-18]**	[0.0,0.3]	[0.07-0.20]	[0.1,1.0]	0.03[-0.37,0.15]	1.16[0.29,2.42]
10	[0.3,0.9]	0.0	[-2.0,0.0]	$\delta^{13}C_{org}$	[0.0,0.3]	[0.07-0.20]	[0.1,1.0]	0.01[-0.43,0.16]	1.02[0.19,2.38]
11	[0.3,0.9]	0.0	**0.0**	-23	[0.0,0.3]	[0.07-0.20]	[0.1,1.0]	0.01[-0.37,0.13]	1.05[0.29,2.08]
12	[0.3,0.9]	0.0	**[-5.0,0.0]**	-23	[0.0,0.3]	[0.07-0.20]	[0.1,1.0]	0.05[-0.36,0.19]	1.30[0.25,3.44]
13	***0.0***	*0.0*	***0.0***	***N/A***	***0.0***	***0.0***	***1.0***	*0.1[0.06,0.13]*	*1.64[1.32,2.04]*

Here $F_{volc+meta.CO_2}$ is the carbon outgassed into the surficial reservoir (Tmol C/yr) from volcanic and metamorphic processes, and $F_{weather.CO_2}$ is the carbon released into the surficial reservoir (Tmol C/yr) from the weathering of carbonates and organic carbon sediments. The combined term in the brackets is the total source flux of carbon into the surface reservoir, which we assume to equal total carbon burial, $F_{total.carbon.burial}$. In the discussion that follows the carbon cycle is always assumed to be in steady state.

A useful metric to conceptualize the rise of oxygen is the dimensionless K_{OXY} parameter, which is defined as the ratio of the oxygen source flux to fast and efficient oxygen sinks. The parameter excludes hydrogen escape, which is a small relative flux at the oxic transition, and oxidative weathering, which increases only afterwards (Claire and others, 2006; Kasting, 2013). Thus, we specify K_{OXY} as:

$$K_{OXY} = (F_{O_2.carbon} + F_{O_2.\mathrm{other}})/F_{reduced} \tag{6.2}$$

Here $F_{O_2.\mathrm{other}}$ is the flux of oxygen from the burial of other reduced species such as sulfide and iron(II) oxide, and $F_{reduced}$ is the flux of outgassed reductants from volcanic and metamorphic processes in O_2 consuming equivalents (Tmol O_2/yr). The utility of K_{OXY} is illustrated by the biogeochemical model of Claire and others (2006) of the Great Oxidation Event. In the Archean, fast and efficient oxygen sinks overwhelm the oxygen source flux from organic burial. This implies that $K_{OXY} < 1$ and that the atmosphere was anoxic. An approximate steady state was maintained by the escape of hydrogen to space. However, if the source flux increases or the sink flux decreases then atmospheric oxygen may accumulate (until oxidative weathering provides a negative feedback). If $K_{OXY} > 1$ then the oxygen source flux dominates fast and efficient oxygen sinks and the resulting atmosphere is oxic. For instance, in Claire and others (2006) f_{org} is held constant over Earth history; the transition to an oxic state is caused by the escape of hydrogen to space, which causes a secular decline in reduced metamorphic and volcanic outgassing fluxes ($F_{reduced}$) due to the gradual oxidation of the crust. Cerium anomalies in zircons suggest that the Earth's crust has in fact become more oxidized over time; Yang and others (2014) suggest that this oxidation proceeded until 3.6 Ga but the exact duration is uncertain due to noise in the data, and the trend seemingly continues into the late Archean. Also, inventories of total O_2 equivalents in the crust that include ocean, atmosphere, sediments, and oxidized hard rocks (igneous and metamorphic) show 1.7 to 2.3 times as much O_2 as moles of organic carbon (Catling and others, 2001; Sleep, 2005; Hayes and Waldbauer, 2006).

The Great Oxidation Event occurs when K_{OXY} transitions from being less than unity to greater than unity. Today, $K_{OXY} \approx 6$, which indicates an oxic atmosphere (Claire and others, 2006; Kasting, 2013). Kasting (2013) adopted K_{OXY} as a simple means of exploring the capacity of different hypotheses to explain the Great Oxidation Event: K_{OXY} was calculated for different assumptions about source/sink fluxes in the Archean (relative to modern fluxes) to see if such changes are capable of implying $K_{OXY} < 1$ in the Archean. If the conjectured fluxes in the Archean do not ensure $K_{OXY} < 1$, then the hypothesis cannot explain the Great Oxidation Event since they would predict that Earth always had an oxic atmosphere.

We adopt the approach of Kasting (2013) to explore whether changes in f_{org} can explain the rise of oxygen. For the modern day fluxes, let us take $F_{O_2.carbon} = 10 \pm 3.3$ Tmol O_2/yr, $F_{O_2.\mathrm{other}} = 5.15 \pm 1.4$ Tmol O_2/yr, and $F_{reduced} = 2.4 \pm 1.8$ Tmol O_2/yr from Holland (2002). Note that our numbers are slightly different to previous studies because there is an error in the stoichiometric coefficients in row 2, Table A2 of Holland (2002), which has propagated through the literature. Using corrected flux values we find K_{OXY}(modern) = (10 + 5.15)/2.4 = 6.3 with a 95 percent confidence interval from 3.3 to 22.3. Thus even if we take the lower end of the 95 percent confidence interval of K_{OXY}(modern) and generously assume that $F_{O_2.\mathrm{other}}$

scales with f_{org} in the same way that $F_{O_2.carbon}$ scales, then a doubling of f_{org} over Earth history – the upper bound of the 95 percent confidence interval from our statistical analysis – implies that $K_{OXY}(Archean) = 0.5 \times 3.3 \geq 1$. Evidently a secular increase in organic burial cannot on its own explain the Great Oxidation Event. This is consistent with the mean-difference analysis that indicated no statistically significant difference between f_{org} in the early Archean and middle Proterozoic. The only caveat on this result is that exceptionally large changes in f_{org} (increases by a factor of a few) are not completely ruled out with the more complicated carbon cycle model in section 5.

However, merely changing f_{org} and holding all other fluxes constant is an unrealistic scenario. The total volcanic outgassing on the early Earth was probably 3 to 5 times larger than the present flux due to greater heat flow from the interior (Sleep and Zahnle, 2001). The overall effect on K_{OXY} of elevated carbon outgassing - and therefore total burial - is partially offset by the corresponding increase in reductant outgassing. However if we assume $F_{volc+meta.CO_2}$ and $F_{reduced}$ are scaled by the same amount for a given increase in total outgassing, then there is still a change in K_{OXY} since presumably $F_{weather.CO_2}$ does not have the same scaling. Additionally, the input of carbon from organic weathering would have been lower in the Archean due to diminished oxidative weathering in anoxic conditions and carbonate weathering may have also been lower if continental size was smaller. Thus total organic and carbonate weathering, $F_{weather.CO_2}$, was likely lower in the Archean than today (Flament and others, 2008; Bekker and Holland, 2012).

To explore whether changes in f_{org} could explain the rise of oxygen given these more realistic flux assumptions we calculated an Archean K_{OXY} using the following expression: f_{org}

$$K_{OXY}(\text{Archean}) = \frac{f_{org} F_{total.carbon.burial} + F_{O_2.\text{other}}}{MF_{reduced}} = \frac{f_{org}(MF_{volc+meta.CO_2} + WF_{weather.CO_2}) + F_{O_2.\text{other}}}{MF_{reduced}} \tag{6.3}$$

Here M (dimensionless) is the total outgassing in the Archean relative to modern levels, and W (dimensionless) is the weathering of carbonate and organic sediments in the Archean as a fraction of the modern flux. Given that the total modern carbon burial rate is ~50 Tmol C/yr (Holland, 2002; Berner, 2004), and that the carbon cycle is in steady state, this implies $F_{volc+meta.CO_2} + F_{weather.CO_2} = 50$ Tmol C/yr. Furthermore, since $F_{volc+meta.CO_2}$ must equal the silicate weathering rate (modern value ~10 Tmol C/yr), we partitioned the modern carbon source fluxes by $F_{volc+meta.CO_2} = 10$ Tmol C/yr and $F_{weather.CO_2} = 40$ Tmol C/yr. All other variables were unchanged from their modern values.

Figure 7 shows Archean K_{OXY} as a function of M and W assuming a doubling of f_{org} over Earth history. We observe that for low Archean weathering fluxes and higher total outgassing rates, there are regions of parameter space with Archean $K_{OXY} < 1$, thus implying that a doubling of f_{org} over Earth history could explain the rise of oxygen. For instance if $M = 2$ then the Archean carbonate and organic weathering flux, W, must be less than 0.31 to ensure an anoxic Archean atmosphere ($K_{OXY} < 1$). Note that it is possible for M to increase and W to decrease because the weathering term only includes carbonate and organic weathering and not silicate weathering; the carbon cycle is in steady state by assumption. In contrast, if the increase in f_{org} over Earth history is only 1.2 relative to 3.6 Ga, the lower limit of our 95 percent confidence interval, then there is no region in this parameter space with $K_{OXY} < 1$. The inability of a factor of 1.2 increase in f_{org} to explain the rise of oxygen is not because the integrated O_2 imbalance is too small (it is not), but rather because such a modest change in the oxygen source

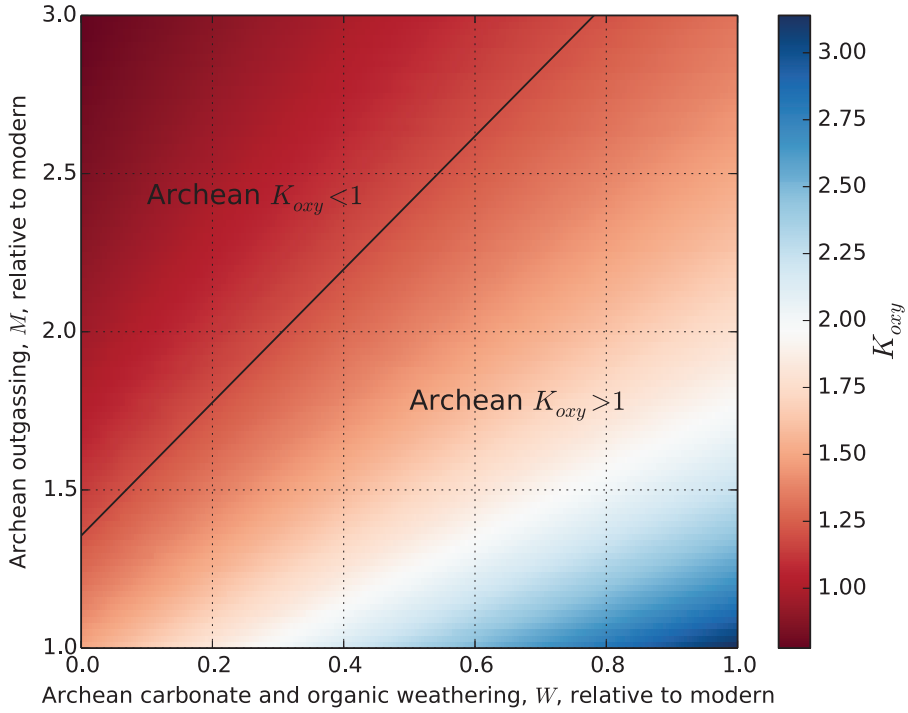

Fig. 7. Archean K_{OXY} as a function of Archean carbonate plus organic weathering plotted as a fraction, W, of the modern flux, and total Archean outgassing plotted as a scalar multiple, M, of the modern flux. This plot assumes a doubling of f_{org} from the Archean to the modern Earth. The black line denotes $K_{OXY} = 1$, the cross-over from anoxic to oxic atmospheres. Regions above the line predict an anoxic Archean atmosphere and could potentially explain the rise of oxygen ($K_{OXY} < 1$), whereas regions below the line predict an oxic Archean atmosphere and cannot explain the rise of oxygen ($K_{OXY} > 1$). If the same figure is plotted for a factor of 1.2 increase in f_{org} over Earth history, then Archean $K_{OXY} > 1$ everywhere.

flux would imply an oxic Archean atmosphere unless Archean O_2 sink fluxes were larger than today. This illustrates that given the uncertainty in f_{org} over Earth history it is difficult to evaluate whether changes in f_{org} constitute an important contribution to the rise of oxygen when other fluxes are changing. This uncertainty arises from noise inherent to the isotope record, and not from uncertainty in our interpretations. Similar conclusions are obtained if we consider changes in the reductant flux in combination with f_{org} increases. This result ignores the sizeable uncertainties in the modern fluxes, and unrealistically assumes that the burial flux from sulfides and iron(II) oxide has remained constant, although similar results are obtained if the burial flux from other reduced species scales with f_{org}, or if $F_{O_2,\mathrm{other}} = 0$ in the Archean.

Evidently to make progress we require better constraints on the absolute O_2 source and sink fluxes through time, which includes the contribution to carbon input from the weathering flux of carbonates and organics on the early Earth (eq. 6.1). The weight percent of organic carbon and carbonates in modern, globally averaged sedimentary rocks is consistent with $f_{org} \approx 0.2$ (Li, 2000). This suggests that organic content could be used to inform total burial rates through time. However the rock record is too incomplete to provide global averages for the weight percent of organic carbon in Precambrian sedimentary rocks. The weight percent of organics in Archean

and Proterozoic shales is within a factor of 2 to 3 of that of Phanerozoic shales (Holland, 1984). From our organic carbon data set, the weight percent organic content of Archean sedimentary rocks (3.59%) is indistinguishable from that of Proterozoic sedimentary rocks (3.56%), and Lyons and others (2014) showed the cumulative distribution of total organic content in organic rich Archean sedimentary rocks is identical to that of Neogene rocks. Thus assuming that sedimentation rates have not changed markedly this would imply the total rate of organic carbon burial also hasn't changed dramatically. However it is difficult to be quantitative about this constraint given the absence of globally averaged sedimentary rock organic content for the Archean; modest changes in the total burial flux of less than an order of magnitude probably cannot be ruled out by sedimentary organic concentrations. Even these modest changes in absolute fluxes could have dramatic implications for the surficial redox balance. Figure 7 and the discussion above demonstrate that relatively small changes in carbonate and organic weathering (or total outgassing) can affect whether changes in fractional organic burial can trigger surface oxidation. Thus, even if the conventional mass balance model of the carbon cycle is accepted, the redox implications of the carbon isotope record will be ambiguous until these absolute fluxes are more tightly constrained.

Our analysis also neglects the oxygen source flux from other biogeochemical cycles. Microbial reduction of sulfate and subsequent burial of sulfide is a sizeable source of oxygen on the modern Earth (Holland, 2002; Berner, 2004). The rise of oxygen could be better understood if fluxes of sulfur were constrained, but this is challenging because unlike the carbon cycle, the sulfur cycle is not in mass balance through the Proterozoic and so sulfur isotopes are of less utility for determining fractional sulfide burial (Canfield, 2004).

Although the carbon isotope record is too noisy to determine whether changes in organic burial can explain the transition from an anoxic to oxic atmosphere, our analysis does indicate a permanent increase in organic burial coincident with the Neoproterozoic rise of oxygen. This is broadly consistent with previous analyses of the carbon isotope record (Des Marais and others, 1992; Jiang and others, 2010), and with mineralogical, tectonic and biological arguments for enhanced Neoproterozoic organic burial (Lenton and others, 2014). For instance, elevated phosphorus delivery (Planavsky and others, 2010), the breakup of the supercontinent Rodinia (Knoll and others, 1986) and biological innovations (Logan and others, 1995) have all been suggested as possible explanations for enhanced organic burial in the Neoproterozoic. Naturally, this increase in fractional organic burial from the Proterozoic to the Phanerozoic is not necessarily the cause of the Neoproterozoic rise of oxygen. We have not quantitatively evaluated whether this enhanced organic burial that could explain the Neoproterozoic rise of oxygen, or even whether the timing is consistent with it being a cause or effect of the rise. Laakso and Schrag (2014)'s model of the Neoproterozoic rise of oxygen implicates an increase in fractional organic burial, and is thus qualitatively consistent with the carbon isotope record. However, their model requires changes in fractional organic burial well in excess of the upper limit of our confidence interval for the change in fractional organic burial from the Proterozoic to the Phanerozoic (assuming the conventional mass balance model is correct).

Our analysis suggests that changes in the carbon isotope record from the early to late Archean are best explained by changes in biological fractionation between inorganic and organic carbon, and not by changes in fractional organic burial, given that the latter is not statistically significant. This is consistent with previous studies that attributed strongly negative late Archean $\delta^{13}C_{org}$ values to the recycling of organic carbon by methanotrophs (Hayes, 1994; Hinrichs, 2002; Hayes and Waldbauer, 2006). In anoxic sediments, isotopically depleted methane produced by fermentation is

converted to biomass by methanotrophs; the resultant organic carbon may be extremely depleted relative to carbonates, with $\delta^{13}C_{org}$ as low as -60 permil (Hayes, 1994). Our analysis also indicates that ϵ declined from the late Archean to the Proterozoic, which suggests methanotrophic recycling was pushed deep into the sediments due to increasingly oxidizing conditions following the great oxidation event (Hayes, 1994; Fallick and others, 2008). Methanotrophy requires either molecular oxygen or oxygen-derived electron acceptors such as sulfates (Hinrichs, 2002). There would have been some abiotically produced sulfate in the Archean atmosphere due to photochemical oxygen production in the upper atmosphere. However, the low partial pressure of oxygen makes this abiotic source an unlikely electron acceptor for the apparent pulse of methanotrophy observed in the globally integrated carbon isotope record. It is more likely that oxygenic photosynthesis produced the abundant electron acceptors that enabled methanotrophic recycling in the late Archean. This ostensible menthanotrophic transition occurs at 2.8 Ga which suggests the evolutionary origin of oxygenic photosynthesis is at least this ancient. This statistical argument supports molybdenum isotope analysis which also indicates an early origin of oxygenic photosynthesis (Planavsky and others, 2014).

7) CONCLUSIONS

- Analyses invoking a conventional carbon isotope mass balance model imply that fractional organic burial, f_{org}, has increased over Earth history by around 0.08 (factor of 1.5 increase relative 3.6 Ga), but with a large 95 percent confidence interval extending from about 0.03 to 0.13 (factor of 1.2-2.0 increase relative to 3.6 Ga).

- The carbon isotope record does not constrain the history of organic burial well enough to evaluate whether changes in fractional organic burial can explain the rise of oxygen. Uncertainties in absolute fluxes such as total outgassing and the weathering of carbonates and organics, coupled with uncertainties in the fractional organic burial change imply that the carbon isotope record neither supports nor negates the hypothesis that secular changes in organic burial were responsible for the oxidation of the surface environment.

- There is a well constrained and permanent increase in organic burial temporally correlated with the Neoproterozoic oxygenation event.

- Analysis of the difference between carbonate and organic carbon isotope ratios, ϵ, implies a statistically significant increase in ε from the early to late Archean. This shift is consistent with enhanced biological fractionation due to methanotrophic recycling, which in turn could indicate that oxygenic photosynthesis evolved 2.8 Ga or earlier.

- All of the conclusions above are derived assuming a conventional mass balance model of the carbon cycle. When authigenic carbonates and ocean crust carbonatization are included in a more complex carbon cycle model, the utility of the carbon isotope record for reconstructing the fractional burial of organic carbon is dramatically limited. If these processes are important, then the various parameters describing the burial of allogenic, authigenic and carbonatization carbonates (see section 5) must be independently determined and specified before the carbon isotope record can be used to constrain the history of organic burial and oxygen production.

ACKNOWLEDGMENTS

We thank Eva Stüeken, Megan Smith, Eric Steig, John Hayes and Donald B. Percival for insightful discussions and assistance in compiling carbon isotope data. We also thank Aviv Bachan, Lee Kump, Timothy Lyons, Stephanie Olson and Daniel Rothman for numerous helpful comments that greatly improved the manuscript. Support for this work included NASA Exobiology Program Grant NNX10AQ90G awarded to DCC, the Virtual Planetary Laboratory award of NASA Astrobiology

Institute, NSF Frontiers in Earth System Dynamics award No. 1338810, and Fulbright New Zealand.

8) Appendix

Appendix A: Algorithms

A.1) Classical Generalized Least Squares (GLS) Fit Algorithm

This is the algorithm for method II, section 3.3.2 in the main text. This algorithm closely follows Mudelsee (2010), Chapter 4, p. 122. We start with either the carbonate or organic carbon isotope time series data, $\{t(i),x(i)\}_{i\,=\,1,\ldots,n}$ where t is time and x is the carbon isotope value for point i.

1) Make an initial guess for the persistence time, τ, and the variability in the noise, V. These variables are defined in main text.

2) Calculate the covariance matrix, $\mathbf{C}$:

$$C(i,j) \;=\; V^2 \times \exp(-|\,t(i) - t(j)\,|/\tau) \tag{A.1}$$

Here $i \;=\; 1, \ldots n\, j \;=\; 1, \ldots n$ index the n data points in the time series, and $t(i)$ is the radiometric age of the i th carbon isotope data point.

3) Perform the GLS regression:

$$\mathbf{B} \;=\; (\mathbf{T}^T\mathbf{C}^{-1}\mathbf{T})^{-1}\,\mathbf{T}^T\,\mathbf{C}^{-1}\mathbf{x} \tag{A.2}$$

$$\text{Where } \mathbf{B} \;=\; \begin{pmatrix} B_0 \\ B_1 \end{pmatrix}, \mathbf{x} = (x(1)\;\ldots\;x(n))^T \text{ and } \mathbf{T} = \begin{pmatrix} 1 & \ldots & 1 \\ t(1) & \ldots & t(n) \end{pmatrix}^T$$

Here $x(i)$ is the i th carbon isotope value, corresponding to time $t(i)$. B_0 and B_1 are the fitted linear regression parameters, intercept and gradient, respectively. Superscript T denotes matrix transpose.

4) Update V using the GLS parameters:

$$V \;=\; \sum_{k=1}^{n} (x(k) - B_1 \times t(k) - B_0)^2/(n-2) \tag{A.3}$$

Use equation (A.1) to update the covariance matrix using the new V. The new autocorrelated residuals $r(i)$ can be found using:

$$r(i) \;=\; (x(i) - B_1 \times t(i) - B_0)/V \tag{A.4}$$

$r(i)$ is the fitted $X_{noise}(i)$ in equation (3.4) in the main text. Substitute $r(i)$ into equation (3.3) in the main text and minimize the sum of the squares to find the best-fit τ.

5) Perform GLS regression on the original data (step 3) but using the new fitted parameters, τ, V and $C(i,j)$. Repeat steps 3-5 until converged, meaning B_0 and B_1 are not changing after each iteration. The B_0 and B_1 obtained at convergence are the best-fit regression parameters. The variance in these parameters (Gaussian uncertainties by assumption) is given by:

$$\mathbf{Var}(\mathbf{B}) \;=\; (\mathbf{T}^T V^{-1}\mathbf{T})^{-1} \tag{A.5}$$

GLS Bootstrap Algorithm

This is the algorithm for method III, section 3.2.2 in the main text. This algorithm closely follows Mudelsee (2010), Chapter 4, p. 123. We start with either the carbonate or organic carbon isotope time series data, $\{t(i),x(i)\}_{i\,=\,1,\ldots,n}$.

1) Perform steps 1-5 from method II.

2) For convenience define the autocorrelation parameter:

$$\hat{a}\,(i) \;=\; \exp(-[t(i) - t(i-1)]/\tau) \tag{A.6}$$

Convert correlated residuals, $r(i)$, (eq (A.4)) to uncorrelated, $\varepsilon(i)$, as follows:

$$\varepsilon(i) \;=\; (r(i) - r(i-1) \times \hat{a}\,(i)) \times (1 - \hat{a}\,(i)^2)^{-\,0.5} \tag{A.7}$$

3) Center on zero mean to obtain uncorrelated white noise residuals:

$$\bar{x}_{noise}(i) = \varepsilon(i) - \sum_{i=2}^{n} \varepsilon(i)/(n-1) \tag{A.8}$$

4) Resample (with replacement) from these white noise residuals.

$$\{\bar{x}_{noise}(i)\} \rightarrow \{(\bar{x}^*_{noise}(i)\} \tag{A.9}$$

5) Invert step 2 to obtain resampled correlated residuals, $r^*(i)$:

$$r^*(i) = \hat{a}(i) \times r^*(i-1) + \bar{x}^*_{noise}(i) \times (1 - \hat{a}(i)^2)^{-0.5} \tag{A.10}$$

6) Add the best-fit gradient from step 1 to resampled correlated residuals in step 5:

$$\bar{x}^*(i) = B_0 + B_1 \times t(i) + V \times r^*(i) \tag{A.11}$$

7) Calculate GLS regression for $\bar{x}^*(i)$ using τ, V and $C(i, j)$ obtained from step 1. The gradient and intercept from this final fit are added to the probability distributions for these parameters. Return to step 4 and resample again until the probability distributions converge.

The GLS code was verified by artificially setting the persistence time, τ, to 0.0001 to check that GLS and OLS confidence intervals converged. Additionally, classical GLS and GLS bootstrap produce identical results for AR(1) test data.

A.3) Mean Difference Algorithm

This is the algorithm for method II, section 3.2.3 in the main text. This algorithm is virtually identical to Algorithm (A.1) above, except that a zero gradient is imposed on the regression.

1) Define two intervals of interest for comparison (for example 1.8-1.0 Ga and 3.8-2.5 Ga), and choose either the organic or carbonate data set, $\{t(i), x(i)\}_{i=1,\ldots,n}$. For the data points in each interval do the following:

2) Make an initial guess for the persistence time τ and the variability in the noise V. Calculate the covariance matrix, C, as before:

$$C(i, j) = V^2 \times \exp(-|t(i) - t(j)|/\tau) \tag{A.12}$$

3) Calculate the GLS regression where the only fitted parameter is the mean level (gradient specified as zero).

$$\mathbf{B} = (\mathbf{T}^T\mathbf{C}^{-1}\mathbf{T})^{-1}\mathbf{T}^T\mathbf{C}^{-1}\mathbf{x} \tag{A.13}$$

Here $\mathbf{B} = (B_0)$ is the fitted mean level, $\mathbf{x} = (x(1) \ldots x(n))^T$ and $\mathbf{T} = (1 \ldots 1)^T$

4) Use the best fit mean-level, B_0, to update the variability term (A.14) and then the covariance matrix (A.12).

$$V = \sum_{k=1}^{n} (x(k) - B_0)^2/(n-2) \tag{A.14}$$

5) Find the correlated residuals:

$$r(i) = (x(i) - B_0)/V \tag{A.15}$$

6) Update the persistence time τ by substituting the correlated residuals into equation (3.3) in the main text, setting $x_{noise}(i) = r(i)$.

7) Calculate the GLS regression with the original data using the new fitted parameters, τ, V and $C(i, j)$. Repeat steps 3-7 until convergence that is B_0 unchanging after each iteration.

8) Given the best-fit mean-level, correlated residuals and variability term, perform the following:

a) Define the autocorrelation parameter:

$$\hat{a}(i) = \exp(-[t(i) - t(i-1)]/\tau)$$

b) Calculate uncorrelated residuals

$$\varepsilon(i) = (r(i) - r(i-1) \times \hat{a}(i)) \times (1 - \hat{a}(i)^2)^{-0.5}$$

TABLE B.1

Change in f_{org} over Earth history excluding Phanerozoic data

	Change in f_{org} (w. 95%)	Change in f_{org} relative to 3.6 Ga (w. 95% confidence interval)	P value
OLS	0.085 [0.004,0.128]	1.55 [1.19,2.06]	0.00006
Classical GLS	0.083 [-0.002,0.175]	1.66 [0.30,3.58]	0.06
Bootstrap AR(1) GLS	0.084 [-0.009,0.176]	1.57 [0.96,2.53]	0.08

Data were filtered and 10 my bins were applied.

c) Center on zero mean to obtain uncorrelated white noise residuals:

$$\bar{x}_{noise}(i) = \varepsilon(i) - \sum_{i=2}^{n} \varepsilon(i)/(n-1)$$

d) Resample uncorrelated residuals

$$\{\bar{x}_{noise}(i)\} \rightarrow \{\bar{x}^*_{noise}(i)\}$$

e) Invert step (8b) to obtain resampled correlated residuals, $r^*(i)$:

$$r^*(i) = \hat{a}(i) \times r^*(i-1) + \bar{x}^*_{noise}(i) \times (1 - \hat{a}(i)^2)^{-0.5}$$

f) Add the best-fit mean-level from step 7 to the resampled correlated residuals in step (8e):

$$\bar{x}^*(i) = B_0 + V \times r^*(i)$$

g) Calculate the GLS regression for $\bar{x}^*(i)$ using the τ, V and $C(i, j)$ obtained from step 7. The mean-level from this final fit is added to the probability distribution for this parameter. Return to step (8d) and resample again to build up the probability distribution.

A.4) Modifications to Algorithms to Account for Timescale Uncertainty

Here we briefly summarize the modifications made to methods I, II and III in the section 3.2.2 (algorithms A.1 and A.2 above) to account for timescale uncertainty. See section 4.2.2 in the main text for an explanation of timescale uncertainty.

I) To adapt OLS regression to timescale uncertainty we took a heuristic approach. The Gaussian distributions for each regression parameter were sampled for each time-uncertain realization to generate an overall probability distribution for each parameter.

II) Gaussian noise was added to time values prior to step 1 and carbon isotope values were re-ordered accordingly. Gaussian distributions obtained at step 5 are sampled, and the algorithm was repeated for different time-uncertainty realizations to generate an overall uncertainty distribution for each parameter.

III) Gaussian noise was added to time values prior to step 1 and carbon isotope values were re-ordered accordingly. At step 7, rather than return to step 4 to resample carbon isotope values, we instead returned to step 1, added Gaussian noise to the original time values, and repeat the entire algorithm to generate probability distributions for regression parameters (this is computationally intensive).

For mean difference methods (section 3.2.3, Algorithm A.3 above), the algorithms were modified to account for time-scale uncertainty in a way identical to how the regression algorithm was modified. Gaussian noise was added to the time values at the start of the algorithm, and for each iteration these were resampled and the entire algorithm was repeated.

APPENDIX B: ADDITIONAL RESULTS

B.1) Sensitivity of Regression Results to Data Selection and Filtering Assumptions

Table B.1 gives the change in f_{org} over Earth history from parametric analysis methods if the Phanerozoic data are excluded. Table B.2 compares the change in f_{org} over Earth history from both parametric and non-parametric smoothing methods for filtered (all results in main text) and unfiltered data. Data filtering is described in section 2 in the main text. Generally speaking, the effect of filtering is to reduce the mean

TABLE B.2

Comparison of mean change in f_{org} over Earth history using filtered and unfiltered data

Method	Change in f_{org} (w. 95%)		Change in f_{org} relative to 3.6 Ga (w. 95% conf.)		P value	
	Filtered	Unfiltered	Filtered	Unfiltered	Filtered	Unfiltered
LOWESS	0.096 [0.056,0.132]	0.10 [0.059,0.147]	1.64 [1.32,2.04]	1.74 [1.34,2.32]	<0.0001	<0.0001
Kernel regression	0.089 [0.048,0.132]	0.082 [0.033,0.126]	1.57 [1.26,2.03]	1.52 [1.17,1.96]	<0.0001	0.02
Kalman smoother	0.081 [0.038,0.126]	0.091 [0.045,0.0138]	1.49 [1.20,1.93]	1.59 [1.25,2.13]	<0.0001	<0.0001
OLS	0.078 [0.039,0.119]	0.100 [0.057,0.146]	1.50 [1.19,1.99]	1.75 [1.31,2.54]	<0.0001	<0.0001
Classical GLS	0.073 [0.012,0.139]	0.082 [0.006,0.160]	1.50 [1.06,2.43]	1.50 [1.02,3.10]	0.01	0.04
Bootstrap GLS	0.072 [0.008,0.139]	0.083 [0.005,0.155]	1.45 [1.04,2.08]	1.59 [1.02,2.43]	0.03	0.04

Filtered results have already been reported in the main text (table 2), but are repeated here for ease of comparison. Both non-parametric smoothing results (rows 1-3) and parametric regression results (rows 4-6) are reported.

change in f_{org} over Earth history. This is unsurprising since by filtering we are excluding extreme values from unrepresentative and/or non-primary samples (for example, lacustrine settings, BIFs *et cetera*). Confidence intervals from unfiltered data are generally larger, and the change in f_{org} over Earth history is even less well constrained, which is also unsurprising for the same reasons.

B.2) Sensitivity of Mean-Difference Results to Filtering

Table B.3 summarizes the mean-difference result for unfiltered data using parametric methods. These results are analogous to table 3 in the main text. In general there are only minor difference between the unfiltered results below and the filtered results in the main text.

B.3) Mean Difference Results from Smoothing Analysis and Comparison with Parametric Results

Table B.4 compares the (filtered data) mean-difference results for non-parametric smoothing methods and parametric regression techniques. The non-parametric smoothing analysis was performed using a white noise model with zero timescale uncertainty, 10 my binning and 10,000 iterations bootstrapping. These results are summarized in section 4.3 in the main text.

B.4) Sensitivity of Results to Diagenetic Modification of Organics

The effects of non-random diagenetic modification of organic carbon were briefly investigated. Des Marais and others (1992) corrected organic carbon data using H/C ratios as a proxy for diagenetic modification; older samples generally show more evidence for diagenetic modification and so this correction results in an increasing trend in $\delta^{13}C_{org}$ compared to the uncorrected data. H/C ratios are unavailable for the majority of new data, and so the effects of diagenetic modification were investigated by artificially adding a trend to the organic carbon isotope series of -2 permil over Earth history. This had the effect of marginally increasing the change in f_{org} over Earth history given the simple carbon cycle model (for example factor of 1.76 increase with confidence interval 1.47 to 2.13 compared to a factor of 1.64 increase confidence interval 1.32 to 2.04 for uncorrected LOWESS smoothing). Similarly, using the updated carbon cycle model with standard parameters and the -2 permil correction for diagenetic changes does not modify the f_{org} increase over Earth history greatly (for example factor of 1.23 increase with confidence interval 0.3 to 2.61 compared to a factor of 1.15 increase with confidence interval 0.29 to 2.44 for uncorrected data). However it should be noted that this *ad hoc* correction does not fully replicate the corrections made for diagenetic modifications in Des Marais and others (1992) where extremely low H/C ratio samples are discarded, in addition to correcting those that remained.

Appendix C: Sensitivity Analysis for Complex Carbon Cycle Model

This section reports the sensitivity of the complex carbon cycle results in section 5 to different assumptions about parameter ranges. If authigenic carbonates are ignored ($\alpha = 0$) then decreasing OCC over time (decreasing $\beta(t)$) will result in an increase in f_{org} if there is a non-zero isotopic gradient in the oceans, ΔS (Bjerrum and Canfield, 2004). However the magnitude of the f_{org} increase is slight for $\Delta S = -2‰$.

If the OCC fraction is ignored ($\beta = 0$) and preservation of authigenics is imperfect relative to allogenic carbonates ($\lambda < 1$) then decreasing the authigenic fraction over time will also result in an increasing f_{org}. This can be understood as follows: if $\lambda = 1$ then f_{org} is unchanged compared to the simple two-sink model since there is effectively no isotopically distinct carbonate sink that isn't being captured by the well-mixed isotope record. However if $\lambda < 1$ and the authigenic fraction was larger in the past, then the true $\delta^{13}C$ of all carbonates that were buried in the past must have been lower than the isotope record suggests (the further back in time we go, the greater the amount of isotopically light carbonates that aren't captured in the geological record). This implies that f_{org} was lower in the past than the simple two-sink model would suggest, and has in fact been increasing over time. The magnitude of this effect is small for λ close to 1, but increases substantially for small λ.

When both OCC and authigenic carbonates are considered together, their effects on the history of f_{org} are more complex and counterintuitive. In particular, if there is some non-zero but fixed fraction of authigenic carbonates, then increasing β can in fact increase f_{org} contrary to what was described above. This is true even if the ocean isotope gradient is zero. The reason for this behavior is that there is no large change in the apparent carbonate record, $\delta^{13}C^{AP}_{carb}$, over Earth history. Hence as the OCC fraction is increased, $\delta^{13}C_{carb}$ must increase to maintain the equality in equation (5.2) in the main text. In other words holding the authigenic fraction fixed whilst decreasing OCC implies the ratio of authigenic carbon to allogenic carbonates is changing, which in turn leads to an increase in f_{org}. A similar counterintuitive result arises if OCC is fixed at some non-zero constant and the authigenic fraction is decreasing (it is possible to get a

TABLE B.3

Summary of the mean-difference result for unfiltered data using parametric methods

A. Late Archean (2.8-2.5) vs Early Archean (3.8-2.8 Ga)								
	Mean level 2.8-2.5 Ga		Mean level 3.8-2.8 Ga		Mean level difference [95% confidence]		P-value	
	t-test	GLSa	t-test	GLSa	t-test	GLSa	t-test	GLSa
$\delta^{13}C_{org}$	-33.8±1.5	-31.1±3.2	-25.5±1.3	-23.8±1.6	-8.3 [-12.2,-4.6]	-7.3 [-14.2,-0.15]	<0.0001	0.046
$\delta^{13}C_{carb}$	-1.59±0.56	-1.9±0.5	-0.94±0.43	*-0.91*±0.38	-0.65 [-2.04,0.75]	-1.0 [-2.3,0.27]	0.36	0.12
f_{org}	0.106±0.016	0.106±0.020	0.165±0.017	0.179±0.019	-0.060 [-0.107,0.014]	-0.0073 [-0.126,-0.017]	0.01	0.01
B. Middle Proterozoic (1.8-1.0 Ga) vs Archean (3.8-2.5 Ga)								
	Mean level 1.8-1.0 Ga		Mean level 3.8-2.5 Ga		Mean level difference [95% confidence]		P-value	
	t-test	GLSa	t-test	GLSa	t-test	GLSa	t-test	GLSa
$\delta^{13}C_{org}$	-28.1±0.4	-28.1±0.8	-29.8±1.1	-27.1±1.7	1.64 [-0.74,4.02]	-1.01 [-4.60,2.64]	0.18	0.59
$\delta^{13}C_{carb}$	0.04±0.21	0.09±0.21	-1.31±0.36	-1.51±0.35	1.35 [0.52,2.17]	1.60 [0.82,2.41]	0.0015	<0.0001
f_{org}	0.179±0.007	0.180±0.008	0.130±0.012	0.137±0.015	0.049 [0.022,0.076]	0.044 [0.010,0.076]	0.00078	0.01
C. Middle Proterozoic (1.8-1.0) vs Early Archean (3.8-2.8 Ga)								
	Mean level 1.8-1.0 Ga		Mean level 3.8-2.8 Ga		Mean level difference [95% confidence]		P-value	
	t-test	GLSa	t-test	GLSa	t-test	GLSa	t-test	GLSa
$\delta^{13}C_{org}$	-28.1±0.4	-28.2±0.84	-25.5±1.3	-23.8±1.6	-2.62[-5.22,0.01]	-4.46 [-7.83,-0.94]	0.05	0.01
$\delta^{13}C_{carb}$	0.04±0.21	0.09±0.21	-0.94±0.43	-0.91±0.40	0.98 [-0.04,1.92]	1.00 [0.15,1.93]	0.04	0.02
f_{org}	0.179±0.007	0.180±0.008	0.165±0.017	0.180±0.019	0.014 [-0.023,0.049]	0.001 [-0.040,0.041]	0.46	0.98
D. Phanerozoic (0.54-0 Ga) vs Middle Proterozoic (1.8-1.0 Ga)								
	Mean level 0.54-0 Ga		Mean level 1.8-1.0 Ga		Mean level difference [95% confidence]		P-value	
	t-test	GLSa	t-test	GLSa	t-test	GLSa	t-test	GLSa
$\delta^{13}C_{org}$	-27.8±0.3	-27.5±0.6	-28.1±0.4	-28.2±0.8	0.31 [-0.70,1.33]	0.70 [-1.38,2.68]	0.55	0.49
$\delta^{13}C_{carb}$	1.75±0.19	1.72±0.19	0.04±0.21	0.10±0.21	1.71 [1.16,2.27]	1.63 [1.08,2.17]	<0.0001	<0.0001
f_{org}	0.228±0.006	0.230±0.007	0.179±0.007	0.181±0.008	0.049 [0.032,0.066]	0.050 [0.029,0.070]	<0.0001	<0.0001

TABLE B.4

Mean difference results from smoothing analysis and comparison with parametric results

A. Late Archean (2.8-2.5) vs Early Archean (3.8-2.8 Ga)						
	Mean level difference in f_{org} [95% confidence]			P value		
Parametric methods	t-test	GLSa	GLSb	t-test	GLSa	GLSb
	-0.028 [-0.06,0.009]	-0.030 [-0.069,0.011]	-0.023[-0.12,0. 12]	0.14	0.15	0.56
Nonparametric, LOWESS	-0.010 [-0.043,0.024]			0.57		
Nonparametric, Kernel	-0.010 [-0.047,0.028]			0.58		
B. Middle Proterozoic (1.8-1.0 Ga) vs Archean (3.8-2.5 Ga)						
	Mean level difference in f_{org} [95% confidence]			P value		
Parametric methods	t-test	GLSa	GLSb	t-test	GLSa	GLSb
	0.031 [0.008,0.053]	0.026 [0.0007,0.051]	0.034 [0.007,0.060]	0.0074	0.045	0.014
Nonparametric, LOWESS	0.028 [0.002,0.054]			0.035		
Nonparametric, Kernel	0.026 [-0.002,0.056]			0.067		
C. Middle Proterozoic (1.8-1.0) vs Early Archean (3.8-2.8 Ga)						
	Mean level difference in f_{org} [95% confidence]			P value		
Parametric methods	t-test	GLSa	GLSb	t-test	GLSa	GLSb
	0.015 [-0.017,0.047]	0.012 [-0.020,0.046]	0.024 [-0.069,0.162]	0.35	0.47	0.71
Nonparametric, LOWESS	0.026 [-0.01,0.055]			0.10		
Nonparametric, Kernel	0.023 [-0.01,0.059]			0.18		
D. Phanerozoic (0.54-0 Ga) vs Middle Proterozoic (1.8-1.0 Ga)						
	Mean level difference in f_{org} [95% confidence]			P value		
Parametric methods	t-test	GLSa	GLSb	t-test	GLSa	GLSb
	0.049 [0.032,0.066]	0.049 [0.030,0.067]	0.036 [-0.008,0.078]	<0.0001	<0.0001	0.10
Nonparametric, LOWESS	0.044 [0.025,0.062]			<0.0001		
Nonparametric, Kernel	0.044 [0.025,0.063]			<0.0001		

decrease in f_{org}). Whether such changes in fractional carbonate burial are physically plausible is unclear, but they demonstrate the possible ranges in behavior if carbon cycle parameters are unconstrained. They also help explain some of the counterintuitive results in the sensitivity analysis (table 4, main text). For instance the increase in f_{org} over Earth history is larger in row 5 than row 6 despite the only change being that the decrease in β is larger for row 6. This seemingly backwards result can be explained by the considerations above.

References

Bekker, A., and Holland, H. D., 2012, Oxygen overshoot and recovery during the early Paleoproterozoic: Earth and Planetary Science Letters, v. 317–318, p. 295–304, http://dx.doi.org/10.1016/j.epsl.2011.12.012

Bekker, A., Holmden, C., Beukes, N. J., Kenig, F., Eglingtone, B., and Patterson, W. P., 2008, Fractionation between inorganic and organic carbon during the Lomagundi (2.22-2.1 Ga) carbon isotope excursion: Earth and Planetary Science Letters, v. 271, n. 1–4, p. 278–291, http://dx.doi.org/10.1016/j.epsl.2008.04.021

Berner, R. A., 2004, The Phanerozoic Carbon Cycle: CO_2 and O_2: Oxford, Oxford University Press, 158 p.

Betts, J. N., and Holland, H. D., 1991, The oxygen content of ocean bottom waters, the burial efficiency of organic carbon, and the regulation of atmospheric oxygen: Palaeogeography, Palaeoclimatology, Palaeoecology, v. 97, n. 1–2, p. 5–18, http://dx.doi.org/10.1016/0031-0182(91)90178-T

Bjerrum, C. J., and Canfield, D. E., 2004, New insights into the burial history of organic carbon on the early Earth: Geochemistry, Geophysics, Geosystems, v. 5, n. 8, http://dx.doi.org/10.1029/2004GC000713

Broecker, W. S., 1970, A boundary condition on the evolution of atmospheric oxygen: Journal of Geophysical Research-Oceans and Atmospheres, v. 75, n. 8, p. 3553–3557, http://dx.doi.org/10.1029/JC075i018p03553

Campbell, I. H., and Allen, C. M., 2008, Formation of supercontinents linked to increases in atmospheric oxygen: Nature Geoscience, v. 1, p. 554–558, http://dx.doi.org/10.1038/ngeo259

Canfield, D. E., 2004, The evolution of the Earth surface sulfur reservoir: American Journal of Science, v. 304, n. 10, p. 839–861, http://dx.doi.org/10.2475/ajs.304.10.839

——— 2014, Proterozoic Atmospheric Oxygen, *in* Farquhar, J., editor, The Atmosphere-History: Treatise on Geochemistry, Second Edition, v. 6, p. 197–216, http://dx.doi.org/10.1016/B978-0-08-095975-7.01308-5

Catling, D. C., 2014, The Great Oxidation Event Transition, *in* Farquhar, J., editor, The Atmosphere-History: Treatise on Geochemistry, Second Edition, v. 6, p. 177–195, http://dx.doi.org/10.1016/B978-0-08-095975-7.01307-3

Catling, D. C., Zahnle, K. J., and McKay, C. P., 2001, Biogenic methane, hydrogen escape, and the irreversible oxidation of early Earth: Science, v. 293, n. 5531, p. 839–843, http://dx.doi.org/10.1126/science.1061976

Catling, D. C., Glein, C. R., Zahnle, K. J., and McKay, C. P., 2005, Why O_2 is required by complex life on habitable planets and the concept of planetary "oxygenation time": Astrobiology, v. 5, n. 3, p. 415–438, http://dx.doi.org/10.1089/ast.2005.5.415

Claire, M. W., Catling, D. C., and Zahnle, K. J., 2006, Biogeochemical modeling of the rise of oxygen: Geobiology, v. 4, n. 4, p. 239–269, http://dx.doi.org/10.1111/j.1472-4669.2006.00084.x

Cleveland, W. S., 1979, Robust Locally Weighted Regression and Smoothing Scatterplots: Journal of the American Statistical Association, v. 74, n. 368, p. 829–836, http://dx.doi.org/10.1080/01621459.1979.10481038

Dahl, T. W., Hammarlund, E. U., Anbar, A. D., Bond, D. P. G., Gill, B. C., Gordon, G. W., Knoll, A. H., Nielsen, A. T., Schovsbo, N. H., and Canfield, D. E., 2010, Devonian rise in atmospheric oxygen correlated to the radiations of terrestrial plants and large predatory fish: Proceedings of the National Academy of Sciences of the United States of America, v. 107, n. 42, p. 17911–17915, http://dx.doi.org/10.1073/pnas.1011287107

Des Marais, D. J., 1994, Tectonic control of the crustal organic carbon reservoir during the Precambrian: Chemical Geology, v. 114, n. 3–4, p. 303–314, http://dx.doi.org/10.1016/0009-2541(94)90060-4

——— 1997, Isotopic evolution of the biogeochemical carbon cycle during the Proterozoic Eon: Organic Geochemistry, v. 27, n. 5–6, p. 185–193, http://dx.doi.org/10.1016/S0146-6380(97)00061-2

Des Marais, D. J., Strauss, H., Summons, R. E., and Hayes, J. M., 1992, Carbon isotope evidence for the stepwise oxidation of the Proterozoic environment: Nature, v. 359, p. 605–609, http://dx.doi.org/10.1038/359605a0

Des Marais, D. J., Harwit, M. O., Jucks, K. W., Kasting, J. F., Lin, D. N. C., Lunine, J. I., Schneider, J., Seager, S., Traub, W. A., and Woolf, N. J., 2002, Remote sensing of planetary properties and biosignatures on extrasolar terrestrial planets: Astrobiology, v. 2, n. 2, p. 153–181, http://dx.doi.org/10.1089/15311070260192246

Fallick, A., Melezhik, V., and Simonson, B. M., 2008, The ancient anoxic biosphere was not as we know it, *in* Dobretsov, N., Kolchanov, N., Rozanov, A., and Zavarzin, G., editors, Biosphere Origin and Evolution: New York, Springer, p. 169–188.

Fan, J., and Yao, Q., 2003, Nonlinear time series : nonparametric and parametric methods: Springer series in statistics: New York, Springer, 576 p.

Farquhar, J., Peters, M., Johnston, D. T., Strauss, H., Masterson, A., Wiechert, U., and Kaufman, A. J., 2007, Isotopic evidence for Mesoarchaean anoxia and changing atmospheric sulphur chemistry: Nature, v. 449, p. 706–709, http://dx.doi.org/10.1038/nature06202

Farquhar, J., Zerkle, A. L., and Bekker, A., 2014, Geologic and Geochemical Constraints on Earth's Early Atmosphere, *in* Farquhar, J., editor, The Atmosphere-History: Treatise on Geochemistry, Second Edition, v. 6, p. 91–138, http://dx.doi.org/10.1016/B978-0-08-095975-7.01304-8

Faure, G., and Mensing, T. M., 2004, Isotopes : principles and applications: Hoboken, New Jersey, Wiley, 928 p.

Flament, N., Coltice, N., and Rey, P. F., 2008, A case for late-Archaean continental emergence from thermal evolution models and hypsometry: Earth and Planetary Science Letters, v. 275, n. 3–4, p. 326–336, http://dx.doi.org/10.1016/j.epsl.2008.08.029

Gaillard, F., Scaillet, B., and Arndt, N. T., 2011, Atmospheric oxygenation caused by a change in volcanic degassing pressure: Nature, v. 478, p. 229–233, http://dx.doi.org/10.1038/nature10460

Hartnett, H. E., Keil, R. G., Hedges, J. I., and Devol, A. H., 1998, Influence of oxygen exposure time on organic carbon preservation in continental margin sediments: Nature, v. 391, p. 572–575, http://dx.doi.org/10.1038/35351

Hayes, J. M., 1994, Global methanotrophy at the Archean-Proterozoic transition, *in* Bengston, S., editor, Early Life on Earth: New York, New York, Columbia University Press, Nobel Symposium, v. 84, p. 220–236.

Hayes, J. M., and Waldbauer, J. R., 2006, The carbon cycle and associated redox processes through time: Philosophical Transactions of the Royal Society B., v. 361, p. 931–950, http://dx.doi.org/10.1098/rstb.2006.1840

Hayes, J. M., Strauss, H., and Kaufman, A. J., 1999, The abundance of ^{13}C in marine organic matter and isotopic fractionation in the global biogeochemical cycle of carbon during the past 800 Ma: Chemical Geology, v. 161, n. 1–3, p. 103–125, http://dx.doi.org/10.1016/S0009-2541(99)00083-2

Hedges, J. I., and Keil, R. G., 1995, Sedimentary organic matter preservation: an assessment and speculative synthesis: Marine Chemistry, v. 49, n. 2–3, p. 81–115, http://dx.doi.org/10.1016/0304-4203(95)00008-F

Hinrichs, K. U., 2002, Microbial fixation of methane carbon at 2.7 Ga: Was an anaerobic mechanism possible?: Geochemistry, Geophysics, Geosystems, v. 3, n. 7, p. 1–10, http://dx.doi.org/10.1029/2001GC000286

Holland, H. D., 1978, The Chemistry of the Atmosphere and Oceans: New York, Wiley, 369 p.

——— 1984, The Chemical Evolution of the Atmosphere and Oceans: Princeton, Princeton University Press, 598 p.

——— 1994, Early Proterozoic atmospheric change, Early life on Earth: New York, Columbia University Press, p. 237–244.

——— 2002, Volcanic gases, black smokers, and the Great Oxidation Event: Geochimica et Cosmochimica Acta, v. 66, n. 21, p. 3811–3826, http://dx.doi.org/10.1016/S0016-7037(02)00950-X

——— 2009, Why the atmosphere became oxygenated: A proposal: Geochimica et Cosmochimica Acta, v. 73, n. 18, p. 5241–5255, http://dx.doi.org/10.1016/j.gca.2009.05.070

Holser, W. T., Schidlowski, M., Mackenzie, F. T., and Maynard, J. B., 1988, Geochemical cycles of carbon and sulfur, *in* Gregor, C. B., Garrels, R. M., Mackenzie, F. T., and Maynard, J. B., editors, Chemical Cycles in the Evolution of the Earth: New York, Wiley, p. 105–173.

Jiang, G., Wang, X., Shi, X., Zhang, S., Xiao, S., and Dong, J., 2010, Organic carbon isotope constraints on the dissolved organic carbon (DOC) reservoir at the Cryogenian–Ediacaran transition: Earth and Planetary Science Letters, v. 299, n. 1–2, p. 159–168, http://dx.doi.org/10.1016/j.epsl.2010.08.031

Kasting, J. F., 2013, What caused the rise of atmospheric O_2?: Chemical Geology, v. 362, p. 13–25, http://dx.doi.org/10.1016/j.chemgeo.2013.05.039

Kasting, J. F., Eggler, D. H., and Raeburn, S. P., 1993, Mantle redox evolution and the oxidation state of the Archean atmosphere: The Journal of Geology, v. 101, n. 2, p. 245–257, http://dx.doi.org/10.1086/648219

Knoll, A. H., Hayes, J. M., Kaufman, A. J., Swett, K., and Lambert, I. B., 1986, Secular variation in carbon isotope ratios from Upper Proterozoic successions of Svalbard and East Greenland: Nature, v. 321, p. 832–838, http://dx.doi.org/10.1038/321832a0

Kump, L. R., and Barley, M. E., 2007, Increased subaerial volcanism and the rise of atmospheric oxygen 2.5 billion years ago: Nature, v. 448, p. 1033–1036, http://dx.doi.org/10.1038/nature06058

Kump, L. R., Bralower, T. J., and Ridgwell, A., 2009, Ocean Acidification in Deep Time: Oceanography, v. 22, n. 4, p. 94–107, http://dx.doi.org/10.5670/oceanog.2009.100

Laakso, T. A., and Schrag, D. P., 2014, Regulation of atmospheric oxygen during the Proterozoic: Earth and Planetary Science Letters, v. 388, p. 81–91, http://dx.doi.org/10.1016/j.epsl.2013.11.049

Lenton, T. M., Boyle, R. A., Poulton, S. W., Shields-Zhou, G. A., and Butterfield, N. J., 2014, Co-evolution of eukaryotes and ocean oxygenation in the Neoproterozoic era: Nature Geoscience, v. 7, p. 257–265, http://dx.doi.org/10.1038/ngeo2108

Li, Y.-H., 2000, A compendium of geochemistry: From solar nebula to the human brain: Princeton, New Jersey, Princeton University Press, 440 p.

Logan, G. A., Hayes, J., Hieshima, G. B., and Summons, R. E., 1995, Terminal Proterozoic reorganization of biogeochemical cycles: Nature, v. 376, p. 53–56, http://dx.doi.org/10.1038/376053a0

Lyons, T. W., Reinhard, C. T., and Planavsky, N. J., 2014, The rise of oxygen in Earth's early ocean and atmosphere: Nature, v. 506, p. 307–315, http://dx.doi.org/10.1038/nature13068

Mattey, D. P., 1987, Carbon isotopes in the mantle: Terra Cognita, v. 7, p. 31–37.

Melezhik, V. A., Young, G. M., Eriksson, P. G., Altermann, W., Kump, L. R., and Lepland, A., 2013, Huronian-age glaciation, *in* Melezhik, V. A., Prave, A. R., Hanski, E. J., Fallick, A. E., Lepland, A., Kump, L. R., and Strauss, H., editors, Reading the Archive of Earth's Oxygenation, v. 3: Berlin, Springer, p. 1059–1109.

Mudelsee, M., 2010, Climate time series analysis : classical statistical and bootstrap methods: Atmospheric and Oceanographic Sciences Library, v. 42, 474 p., http://dx.doi.org/10.1007/978-90-481-9482-7

Nakamura, K., and Kato, Y., 2004, Carbonatization of oceanic crust by the seafloor hydrothermal activity and its significance as a CO_2 sink in the Early Archean: Geochimica et Cosmochimica Acta, v. 68, n. 22, p. 4595–4618, http://dx.doi.org/10.1016/j.gca.2004.05.023

Och, L. M., and Shields-Zhou, G. A., 2012, The Neoproterozoic oxygenation event: Environmental perturbations and biogeochemical cycling: Earth-Science Reviews, v. 110, n. 1–4, p. 26–57, http://dx.doi.org/10.1016/j.earscirev.2011.09.004

Pavlov, A. A., and Kasting, J. F., 2002, Mass-independent fractionation of sulfur isotopes in Archean sediments: strong evidence for an anoxic Archean atmosphere: Astrobiology, v. 2, n. 1, p. 27–41, http://dx.doi.org/10.1089/153110702753621321

Pearson, D. G., Canil, D., and Shirey, S. S., 2003, Mantle samples included in volcanic rocks: Xenoliths and diamonds, *in* Carlson, R. W., editor, The Mantle and Core: Treatise on Geochemistry, v. 2, p. 171–276, http://dx.doi.org/10.1016/B0-08-043751-6/02005-3

Planavsky, N. J., Rouxel, O. J., Bekker, A., Lalonde, S. V., Konhauser, K. O., Reinhard, C. T., and Lyons, T. W., 2010, The evolution of the marine phosphate reservoir: Nature, v. 467, p. 1088–1090, http://dx.doi.org/10.1038/nature09485

Planavsky, N. J., Asael, D., Hofmann, A., Reinhard, C. T., Lalonde, S. V., Knudsen, A., Wang, X., Ossa, F. O., Pecoits, E., Smith, A. J., Beukes, N. J., Bekker, A., Johnson, T. M., Konhauser, K. O., Lyons, T. W., and Rouxel, O. J., 2014, Evidence for oxygenic photosynthesis half a billion years before the Great Oxidation Event: Nature Geoscience, v. 7, p. 283–286, http://dx.doi.org/10.1038/ngeo2122

Rothman, D. H., 2015, Earth's carbon cycle: a mathematical perspective: Bulletin of the American Mathematical Society, v. 52, n. 1, p. 47–64 (published electronically on September 17, 2014).

Rothman, D. H., Hayes, J. M., and Summons, R. E., 2003, Dynamics of the Neoproterozoic carbon cycle: Proceedings of the National Academy of Sciences of the United States of America, v. 100, n. 14, p. 8124–8129, http://dx.doi.org/10.1073/pnas.0832439100

Schidlowski, M., 1988, A 3,800-million-year isotopic record of life from carbon in sedimentary rocks: Nature, v. 333, p. 313–318, http://dx.doi.org/10.1038/333313a0

Schidlowski, M., Eichmann, R., and Junge, C. E., 1976, Carbon isotope geochemistry of Precambrian Lomagundi carbonate province, Rhodesia: Geochimica et Cosmochimica Acta, v. 40, n. 4, p. 449–455, http://dx.doi.org/10.1016/0016-7037(76)90010-7

Schidlowski, M., Appel, P. W., Eichmann, R., and Junge, C. E., 1979, Carbon isotope geochemistry of the 3.7×10^9-yr-old Isua sediments, West Greenland: implications for the Archaean carbon and oxygen cycles: Geochimica et Cosmochimica Acta, v. 43, n. 2, p. 189–199, http://dx.doi.org/10.1016/0016-7037(79)90238-2

Schopf, J. W., and Klein, C., 1992, The Proterozoic Biosphere: A Multidisciplinary Study: Cambridge, Cambridge University Press, 1348 p.

Schrag, D. P., Higgins, J. A., Macdonald, F. A., and Johnston, D. T., 2013, Authigenic Carbonate and the History of the Global Carbon Cycle: Science, v. 339, n. 6119, p. 540–543, http://dx.doi.org/10.1126/science.1229578

Shields-Zhou, G., and Och, L., 2011, The case for a Neoproterozoic Oxygenation Event: Geochemical evidence and biological consequences: GSA Today, v. 21, p. 4–11, http://dx.doi.org/10.1130/GSATG102A.1

Shields, G., and Veizer, J., 2002, Precambrian marine carbonate isotope database: Version 1.1: Geochemistry, Geophysics, Geosystems, v. 3, n. 6, http://dx.doi.org/10.1029/2001GC000266

Shirey, S. B., Cartigny, P., Frost, D. J., Keshav, S., Nestola, F., Nimis, P., Pearson, D. G., Sobolev, N. V., and Walter, M. J., 2013, Diamonds and the Geology of Mantle Carbon, *in* Hazen, R. M., Jones, A. P., and Baross, J. A., editors, Carbon in Earth: Reviews in Mineralogy and Geochemistry, v. 75, p. 355–421, http://dx.doi.org/10.2138/rmg.2013.75.12

Shumway, R. H., and Stoffer, D. S., 2011, Time series analysis and its applications : with R examples: Springer texts in statistics: New York, Springer, 596 p.

Simon, D., 2001, Kalman Filtering: Embedded Systems Programming, v. 14, n. 6, p. 72–79.

Sleep, N., 2005, Dioxygen over geological time: Metal ions in biological systems, v. 43, p. 49–73, http://dx.doi.org/10.1201/9780824751999.ch3

Sleep, N. H., and Zahnle, K., 2001, Carbon dioxide cycling and implications for climate on ancient Earth: Journal of Geophysical Research-Planets, v. 106, n. E1, p. 1373–1399, http://dx.doi.org/10.1029/2000JE001247

Stüeken, E. E., Catling, D. C., and Buick, R., 2012, Contributions to late Archaean sulphur cycling by life on land: Nature Geoscience, v. 5, p. 722–725, http://dx.doi.org/10.1038/ngeo1585

Sun, X. L., and Turchyn, A. V., 2014, Significant contribution of authigenic carbonate to marine carbon burial: Nature Geoscience, v. 7, p. 201–204, http://dx.doi.org/10.1038/ngeo2070

Sundquist, E. T., 1991, Steady-State and Non-Steady-State Carbonate Silicate Controls on Atmospheric CO_2: Quaternary Science Reviews, v. 10, n. 2–3, p. 283–296, http://dx.doi.org/10.1016/0277-3791(91)90026-Q

Thomazo, C., Ader, M., Farquhar, J., and Philippot, P., 2009, Methanotrophs regulated atmospheric sulfur isotope anomalies during the Mesoarchean (Tumbiana Formation, Western Australia): Earth and Planetary Science Letters, v. 279, n. 1–2, p. 65–75, http://dx.doi.org/10.1016/j.epsl.2008.12.036

Yang, X., Gaillard, F., and Scaillet, B., 2014, A relatively reduced Hadean continental crust and implications for the early atmosphere and crustal rheology: Earth and Planetary Science Letters, v. 393, p. 210–219, http://dx.doi.org/10.1016/j.epsl.2014.02.056

Young, G. M., Brunn, V. V., Gold, D. J., and Minter, W., 1998, Earth's oldest reported glaciation: Physical and chemical evidence from the Archean Mozaan Group ($\sim$ 2.9 Ga) of South Africa: The Journal of Geology, v. 106, n. 5, p. 523–538, http://dx.doi.org/10.1086/516039

Zahnle, K. J., Catling, D. C., and Claire, M. W., 2013, The rise of oxygen and the hydrogen hourglass: Chemical Geology, v. 362, p. 26–34, http://dx.doi.org/10.1016/j.chemgeo.2013.08.004