Robust multi-proxy data integration, using late Cretaceous paleotemperature records as a case study

Lineke Woelders*[a,b], Johan Vellekoop[a], Gert Jan Weltje[a], Lennart de Nooijer[c], Gert-Jan Reichart[c,d], Francien Peterse[d], Philippe Claeys[e], Robert P. Speijer[a]

[a]Division of Geology, Department of Earth and Environmental Sciences, KU Leuven, Celestijnenlaan 200e, 3001 Leuven, lineke.woelders@colorado.edu

[b]Institute of Arctic and Alpine Research, University of Colorado, Boulder, CO, 4001 Discovery Drive, 80303, United States of America

[c]Department of Ocean Systems, NIOZ-Royal Netherlands Institute for Sea Research and Utrecht University, Den Burg, The Netherlands

[d]Department of Earth Sciences, Faculty of Geosciences, Utrecht University, Princetonlaan 8A, 3584 CB, Utrecht, the Netherlands

[e]Analytical, Environmental and Geo-Chemistry, Vrije Universiteit Brussel, Pleinlaan 2, B-1050, Brussels, Belgium

*Corresponding author. Contact information:

e-mail: lineke.woelders@colorado.edu

Highlights

- An approach to statistical integration of multi-proxy data is presented
- Minimum uncertainties of individual records are provided
- A warming event of 3.9±1.1 ˚C occurred ~450 to ~100 kyr before the K-Pg boundary
- Reconstructions are consistent with other SST records, suggesting a global signal

Keywords

Late Cretaceous

Paleotemperature

Multi-proxy

Data integration

Abstract

In paleoclimate studies, multiple temperature records are often compared and combined to evaluate temperature trends. Yet, no standardized approach for integrating proxy-derived paleotemperature records exists. In addition, paleotemperature data are often reported without uncertainty estimates (prediction errors), and raw data are not always available. This complicates the quantification of, for example, temperature trends and the magnitude of warming events. Here we propose a robust quantitative approach for multi-proxy analysis in paleoclimate studies. To demonstrate this, we study the latest Maastrichtian warming event (LMWE) in the ODP 174AX Bass River core (New Jersey), and integrate five independent paleotemperature proxies covering the last million years of the Cretaceous. Our integrated temperature reconstruction suggests that, after a climatically stable period, a latest Cretaceous warming of 3.9±1.1 ˚C occurred between ~450 and 100 kyr before the K-Pg boundary. The error on this reconstructed temperature should be considered the absolute minimum error, as poorly constrained or unknown uncertainties cannot be fully propagated. The warming event was followed by a gradual cooling to pre-warming conditions towards the end of the Cretaceous. Furthermore, the record suggests multiple warming pulses during the LMWE. The results of this integrated approach are consistent with other latest Cretaceous temperature records, suggesting that the trend described here represents a global signal.

1. Introduction

Latest Cretaceous (late Maastrichtian) paleotemperature records suggest a global warming of 2-8˚C between ~450 and ~100 kyr before the K-Pg boundary (Li and Keller, 1998; Barrera and Savin, 1999; Olsson et al., 2002; Wilf et al., 2003; Westerhold et al., 2011; Tobin et al., 2012; Petersen et al., 2016; Thibault et al., 2016; Vellekoop et al., 2016; Barnet et al., 2017), possibly as a result of Deccan Traps volcanism (Chenet et al., 2007; Schoene et al., 2015; Henehan et al., 2016). Yet, despite these reports, the exact magnitude of the late Maastrichtian warming event (LMWE) remains poorly constrained, mainly because several sources of uncertainty are not, or are poorly, addressed.

Firstly, many, especially older, paleotemperature studies provide proxy records without an analytical error and without providing the raw data in the form of a table or supplementary data file. This introduces a source of uncertainty when using these records in further studies. Though these uncertainty estimates have been included in recent publications on late Cretaceous temperature evolution (Tobin et al., 2012, 2014; Petersen et al., 2016), their absence in older studies limits the accuracy of estimates of the magnitude of the LMWE.

Calibrations used to reconstruct past temperatures introduce a second source of uncertainty. Although they often include a regression error (e.g. Bemis et al., 1998; Lear et al., 2002; Martin et al., 2002; Anand et al., 2003; Kim et al., 2010), this does not provide a realistic estimate of their predictive capabilities as obtained by cross-validation techniques. If the raw data used in temperature calibrations are unavailable, it is impossible to correctly propagate the prediction errors associated with these calibrations, prohibiting accurate estimation of the magnitude of the LMWE. Furthermore, lack of access to the raw data does not permit improvement of calibrations by combining different data sets, or by redesigning calibration equations as a result of new insights.

The third source of uncertainty is of a more fundamental nature. The use of modern-day calibrations to reconstruct climatic and environmental changes in the paleo-domain introduces uncertainties whose magnitudes are poorly constrained (or even completely unknown), owing to the fact that parameters of proxy calibrations for temperature estimation cannot be verified for deep-time applications. Modern-day proxy calibrations could be based on species that were nonexistent in the late Cretaceous, and it is not known whether species in this time interval responded similarly to temperature changes as their modern-day counterparts. Detailed information on paleo-sea-water composition (salinity, oxygen isotopes, Mg/Ca ratios) is similarly unavailable. Further uncertainty on the magnitude of warming arises from proxy records (Li and Keller, 1998; Barrera and Savin, 1999; Olsson et al., 2002; Tobin et al., 2012) that may have suffered from unknown diagenetic effects (Pearson et al., 2001), which could have altered the original signal. And lastly, each single proxy may also be influenced by variables other than temperature, which prohibits reliable comparison of temperatures derived from different proxies (Lawrence and Woodard, 2017). For example, regional hydrological change as well as the waxing and waning of ice sheets, possibly even in the latest Cretaceous greenhouse climate

113 (Miller et al., 2005), may have influenced late Maastrichtian marine $\delta^{18}O$ values and thereby
114 reconstructed temperatures.
115 In addition to these three sources of uncertainty, that prohibit an accurate reconstruction of
116 the LWME, previous late Maastrichtian multi-proxy studies compare temperature records from
117 different locations, rather than comparing multiple proxy records from a single location (Wilf et
118 al., 2003; Woelders et al., 2017). As a result, observed differences between temperature records
119 in these studies could also be caused by, for instance, local diagenetic processes instead of actual
120 (local) climate signals. Furthermore, owing to problems of age constraints (Olsson et al., 2002;
121 Wilf et al., 2003; Woelders et al., 2017), it cannot be excluded that seemingly age-equivalent
122 signals are actually diachronous.
123 Finally, the low temporal resolution (commonly <10 data points in the last million years of
124 the Maastrichtian (Barrera and Savin, 1999; Tobin et al., 2012, 2014; Petersen et al., 2016)) and
125 low signal-to-noise ratios (Li and Keller, 1998; Barrera and Savin, 1999; Wilf et al., 2003;
126 Westerhold et al., 2011; Petersen et al., 2016) also limit detailed reconstruction of the LMWE.
127 The above sources of uncertainty, and additional challenges posed by low resolution
128 datasets and low signal-to-noise ratios, hamper a robust determination of the magnitude of the
129 LMWE, which in turn complicates the assessment of the environmental and ecological
130 implications of late Maastrichtian climate change on the K-Pg mass extinction.
131 The risks associated with relying on single proxy records may be reduced by using multiple
132 paleo-temperature proxy records. Data in multi-proxy studies are commonly displayed side by
133 side to highlight consistency between datasets, or to illustrate patterns of variability unique to
134 each dataset. This approach can, however, be improved upon substantially by statistically
135 integrating temperature records based on different proxies. This enables quantification of the
136 signal shared by the individual datasets (cf. Bloemsma et al., 2012), and allows for more accurate
137 and robust quantification of temperature trends and anomalies. The latter is extremely relevant
138 for climate modeling purposes, as accurate estimates of past sea surface temperatures (SSTs) are
139 a prerequisite for paleoclimate model forcing (e.g. Sloan et al., 2001) and model evaluation (e.g.
140 Otto‐Bliesner et al., 2017).

Here we introduce a formal approach to multi-proxy analysis based on data integration, to permit objective assessment of the significance of temperature changes. Integration of reconstructed temperature records based on different datasets using different proxies, and thus different temperature calibrations, requires quantification of the uncertainties of each, since the uncertainty directly translates to the weight of evidence of the data and temperature transfer functions. Following this approach, we present an integrated multi-proxy temperature reconstruction of the late Maastrichtian, spanning approximately the last million years of the Cretaceous, by combining benthic and planktic foraminiferal $\delta^{18}O$ and Mg/Ca and the organic geochemical SST proxy TEX_{86} records for the Bass River core (ODP Leg 174AX, New Jersey, Figure 1) (Miller et al., 1998; Olsson et al., 2002).

The shelf sequence of Bass River (paleodepth ~100 m (Olsson et al., 2002)) comprises an uppermost Maastrichtian record with exceptionally well-preserved benthic and planktic foraminifera (Figure 2) and biomarker lipids, enabling a multi-proxy temperature reconstruction. The LMWE has previously been identified in the Bass River core, based on stable oxygen isotopes (Olsson et al., 2002; Esmeray-Senlet et al., 2015), indicating that this sedimentary archive is ideal for a case study aimed at testing robust multi-proxy paleo-temperature reconstruction. The glauconite-rich sediments suggest low rates of sediment accumulation, but there is no evidence of the presence of hiatuses in this interval (Miller et al., 1998). Furthermore, despite the limited distance (~100 km) of the Bass River site to the paleo-coastline (Esmeray-Senlet et al., 2015), the site is directly connected to the open ocean and likely experienced little terrestrial influx (Zachos et al., 2006). This suggests that the site is likely to have recorded the effects of global climate change.

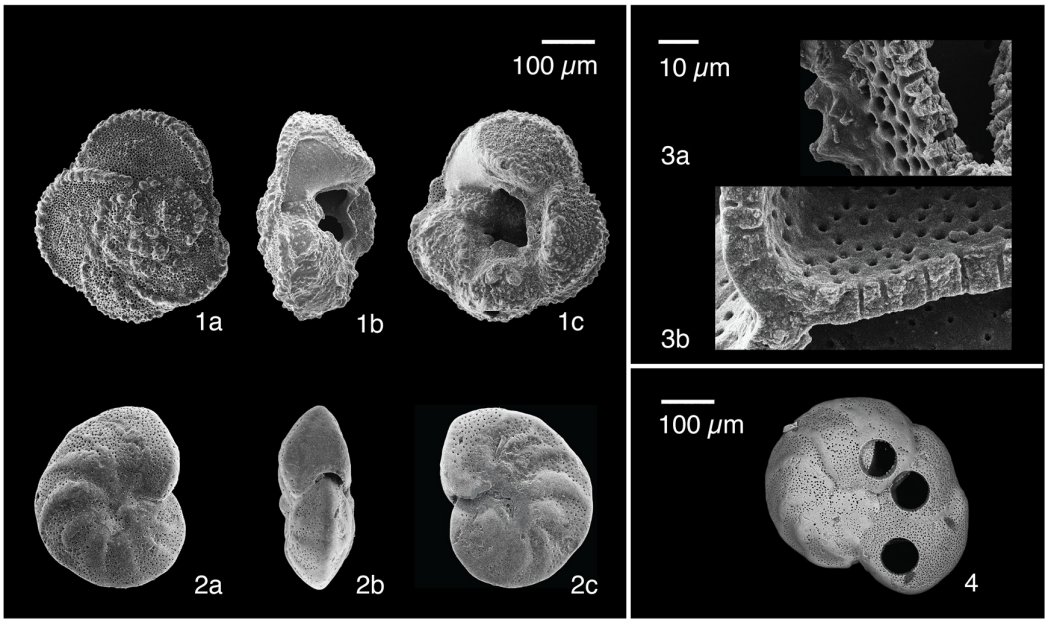
165

166 **Figure 1** Geological map of the New Jersey shelf area with the location of ODP Leg 174 AX Bass

167 River site in New Jersey, USA, indicated. Modified after Esmeray-Senlet (2015).

168

169



170

171 **Figure 2** SEM images of *Anomalinoides midwayensis* and *Globotruncana* sp. 1a-c: *Globotruncana*

172 sp. 2a-c: *Anomalinoides midwayensis*. 3a: Detail of broken *Globotruncana* sp. test. 3b: Detail of

173 broken *Anomalinoides midwayensis* test. 4: *Anomalinoides midwayensis* with laser holes.

2. Material and methods

2.1. Stable oxygen isotope analysis

Samples were washed over a sieve with a mesh of 63 µm and ultrasonically cleaned. Only specimens from the size fraction of 125-630 µm were analyzed. For benthic $\delta^{18}O$ analysis, duplicate measurements were performed on 7-9 excellently preserved specimens of *Anomalinoides midwayensis* (Figure 2) isolated from each of 37 analyzed samples. For planktic $\delta^{18}O$ analysis, 6-9 specimens of excellently preserved *Globotruncana* spp. (Figure 2) were measured from samples containing sufficient specimens. *Globotruncana* spp. was not sufficiently abundant for adequate $\delta^{18}O$ analysis in the lowermost samples of the record which is why no data could be generated for this part of the record. Duplicate measurements on *Globotruncana* spp. could not be performed because of the limited number of well-preserved specimens in each sample. Oxygen stable isotopes were measured with Isotopic Ratio Mass Spectrometry (IRMS) at the Stable Isotope Laboratory of the Analytical, Environmental and Geochemistry research group, Free University of Brussels (VUB). The setup consists of a Perspective IRMS coupled to a NuCarb automated carbonate sampling device (Nu Instruments). Two samples of the in-house standard MAR-2(2), which has been calibrated against the international standard NBS-19 were measured every 8 samples to correct for instrumental drift. Reported values of MAR-2(2) are 0.13 ‰ VPDB for $\delta^{18}O$ and 3.41 ‰ VPDB for $\delta^{13}C$. Analytical errors were calculated by the sum of squares of the standard deviations from the individual sample results and the measured MAR-2(2). The weighted mean and corresponding confidence intervals of the benthic dataset were calculated using the method described in Text S1. Data points without analytical error or without a duplicate measurement were assigned an average confidence interval derived from the complete dataset. As no duplicate measurements are available for the planktic dataset, the confidence interval on the planktic data was assumed to be identical to the average confidence interval calculated for the benthic foraminifera.

The temperature transfer function used for benthic foraminifera is that published for *Cibicidoides* by Marchitto et al. (2014). This function is suitable for relatively warm water conditions and is based on epibenthic *Cibicidoides* species, which is here considered applicable to *Anomalinoides midwayensis*, an epibenthic taxon morphologically similar to *Cibicidoides.* The

extinct planktic *Globotruncana* spp. is presumably a non-symbiotic (D'Hondt and Zachos, 1998) subsurface to thermocline dweller (Abramovich et al., 2003). The temperature transfer function used for *Globotruncana* spp. is the function for non-symbiotic surface dwelling *Globigerina bulloides* (13 chambered shell) by Bemis et al. (1998). See also Text S2. In these calculations, an ice-free world $\delta^{18}O_{sw}$ (δ_w) value of -1.2‰ was assumed (Shackleton and Kennett, 1975).

To estimate the uncertainty of the predicted temperatures, the standard rules of error propagation were used, using the analytical errors and the uncertainties on the variables in the temperature equations. It should be noted that uncertainty introduced by e.g. unknown seasonality effects, diagenetic effects (despite the careful selection of well-preserved specimens) and by combining several *Globotruncana* species into *Globotruncana* spp., could have contributed significantly to the total uncertainty on the reconstructed temperatures (Fraass and Lowery, 2017). These uncertainties were, however, not taken into account, since their quantification is impossible.

2.2. Mg/Ca paleothermometry

Mg/Ca ratios in foraminiferal calcite were used to derive bottom water and sea surface temperatures. Prior to analysis, all foraminiferal tests were cleaned by rinsing with ethanol and ultrasonication, to remove adhering clay and other detrital material. All foraminifera were ultrasonically treated for 30 seconds in ethanol. For benthic Mg/Ca analysis, 8-12 specimens of *A. midwayensis* were selected for 38 samples. For planktic Mg/Ca analysis, 6-12 specimens of *Globotruncana* spp. (Figure 2) were selected for 14 samples. *Globotruncana* spp. was not sufficiently abundant in the lowermost samples of the record for adequate Mg/Ca analysis, which is why no data were generated for this part of the record.

We ablated the selected specimens using a NWR193UC (New Wave) that uses an ArF excimer laser with short pulse width and deep-ultraviolet wavelength (193nm) (e.g. Reichart et al., 2003). More information on the ablation procedure can be found in Text S3.

For planktic foraminifera, before calculating temperatures from measured Mg/Ca values, a pH correction was applied (using the estimated pH of past seawater, pH_{sw}). This was not done for benthic foraminifera (Text S3). Furthermore, modern-day temperature calibrations based on Mg/Ca values cannot be directly applied to Mg/Ca values of foraminifera that lived in oceans

232 with a considerably different Mg/Ca. Therefore, the measured Mg/Ca values were transformed to
233 modern-day values before the calibrations were applied. This transformation was done by
234 applying a correction for the difference between past Mg/Ca_{sw} and present seawater Mg/Ca
235 ($Mg/Ca_{sw,present}$) (Text S3). For this transformation, a Mg/Ca_{sw} value of 1.25 mol/mol during the
236 latest Cretaceous as reconstructed by Stanley and Hardie (1998) was assumed. The assumed
237 Mg/Ca_{sw} and pH_{sw} are associated with uncertainties, potentially influencing the reconstructed
238 temperatures. Despite the fact that these uncertainties can be considerable (Text S3), they only
239 result in a linear transformation of absolute temperature reconstructions. Relative temperature
240 changes are not affected by these uncertainties.

241 The temperature transfer function used for benthic foraminifera is the function for
242 *Cibicidoides* from Lear et al. (2002) which is suitable for relatively warm water conditions.
243 Epibenthic *Cibicidoides* morphologically resembles extinct epibenthic taxon *Anomalinoides*
244 *midwayensis* used in this study. For extinct planktic *Globotruncana* spp., we use the Mg/Ca (Lea et
245 al., 1999) transfer function based on non-symbiont bearing *Globigerina bulloides*. See also Text
246 S3.

247 To estimate uncertainties for the predicted temperatures, the standard rules of error
248 propagation were used, using the analytical errors and the uncertainties on the variables in the
249 temperature equations. See also Text S3. As with stable oxygen isotope analysis, uncertainties
250 resulting from e.g. unknown seasonality effects, undetected diagenetic overprint or combining
251 the individual *Globotruncana* species in the *Globotruncana* spp. group could not be quantified.

252 **2.3. TEX_{86} paleothermometry**

253 For TEX_{86} analyses, freeze-dried, powdered samples (~10 g dry mass) were extracted with
254 an accelerated solvent extractor using a 9:1 (vol/vol) dichloromethane (DCM):methanol solvent
255 mixture. The obtained extracts were separated over an activated Al_2O_3 column, using 9:1
256 (vol/vol) hexane: DCM, ethyl acetate (100%), 95:5 (vol/vol) DCM:MeOH, and 1:1 (vol/vol) DCM:
257 methanol, into apolar, ethylacetate, tetraether, and polar fractions, respectively. The tetraether
258 fractions, containing the glycerol dialkyl glycerol tetraethers (GDGTs) used for the TEX_{86} index
259 were analyzed in duplicate by high performance liquid chromatography/ atmospheric pressure
260 positive ion chemical ionization-mass spectrometry (HPLC/APCI-MS) using an Agilent 1260

Infinity ultra HPLC coupled to an Agilent 6130 single quadrupole MS, following the elution scheme of Hopmans et al. (2016).

The TEX_{86}^{H} calibration from Kim et al. (2010) is the most commonly used transfer function for deep-time (pre-Neogene), extrapolar greenhouse conditions (Schouten et al., 2013). Therefore, in this study, TEX_{86} index values were calculated for the duplicate measurements and in principal converted into SSTs using the TEX_{86}^{H} calibration from Kim et al. (2010). Later (section 4.3) we discuss this choice of the empirical TEX_{86}^{H} calibration for our TEX_{86} data from the data-sparse, latest Cretaceous greenhouse over the statistically better constrained BAYSPAR calibration (Tierney and Tingley, 2014). Additional information on the quality control of the data can be found in Text S4. The weighted mean of the duplicate temperature estimate and the corresponding confidence intervals were calculated using the method described in Text S1. The uncertainty on the reconstructed temperatures is ± 2.5°C.

2.4. Calculation of the weighted mean and inverted Chi² fitting

We introduce a formal approach to multi-proxy analysis based on data integration, which permits an objective assessment of the significance of temperature changes. In order to integrate the five proxy records, we standardize each based on uncertainties of the individual data points as well as their scatter around the weighted mean of the individual records (Inverted chi-squared fitting and Z score analysis, see Text S1). The uncertainties of the individual data points are obtained by propagating analytical errors as well as the regression errors on the calibrations used. As such, these uncertainties are the absolute minimal uncertainties associated with the data points, as poorly constrained or unknown uncertainties cannot be fully propagated. The Z score analysis is needed because late Maastrichtian temperature has been reconstructed using different methods, and an integrated estimate can only be obtained by centering the individual temperature records on zero to eliminate systematic errors that are inherent to paleo-temperature estimates (Z score analysis, Text S1). After the quantitative integration of temperature records, the range of the mean standardized temperature record is restored. The uncertainty around the restored mean standardized temperature is restored by an inverse scaling operation (see Text S1). The magnitude of the LMWE is estimated by determining the difference between the weighted mean of the restored standardized temperatures in the

relatively stable interval below the warming (-3.11 to -1.45 m) and the weighted mean of the data points of the restored standardized temperatures with higher temperatures in the warming interval.

As intense bioturbation resulted in mixing of Paleocene sediments with the uppermost 15 cm of the Maastrichtian (Olsson et al., 1997), measurements from this interval have not been used to calculate the weighted mean of the proxy records.

3. Results

3.1. Latest Cretaceous climate change

The late Maastrichtian temperature reconstructions from Bass River show similar trends (Figure 3a and 3b). All proxies indicate an interval with relatively stable temperatures from ~3.11 m up to ~1.45 m below the K-Pg boundary. Between 1.42 and 1.13 m below the boundary, a warming relative to the stable interval is indicated by all temperature proxies. The simultaneous occurrence of the warming interval in the individual temperature records suggests that bottom and (sub)surface waters warmed simultaneously. The approach of data integration as described in Methods (2.4) results in an estimate of late Maastrichtian warming of 3.3±2.8 °C (light-gray area in Figure 4).

It is impossible to pinpoint the exact timing of the onset of the LMWE because of the lack of detailed biostratigraphy and the presence of a core gap (Miller et al., 1998) (Figure 3). Given these limitations, the timing of the warming interval recorded in Bass River (between ~450 and ~100 kyr before the K-Pg boundary, with peak warming at ~250 kyr before the K-Pg boundary) was estimated based on paleotemperature records with better age constraints (Li and Keller, 1998; Barrera and Savin, 1999; Wilf et al., 2003; Westerhold et al., 2011; Tobin et al., 2012; Petersen et al., 2016; Vellekoop et al., 2016; Barnet et al., 2017; Woelders et al., 2017).

Above the warming peak between -1.42 and -1.13 m, all proxies suggest a cooler interval before the K-Pg boundary. However, temperatures remain relatively high up to 0.50 m below the boundary. Although the absence of stratigraphic markers in this part of the core prevents an accurate age estimate, the onset of this cooling likely corresponds to an age of ~100 kyr before the K-Pg boundary (Barnet et al., 2017). Between ~-0.75 to -0.50 m below the boundary, the

planktic proxy-based temperature reconstructions suggest a second, less intense, warming event (Figure 3b), although not statistically significant. Data can be found in Supplementary Data Files Dataset S1-S3.

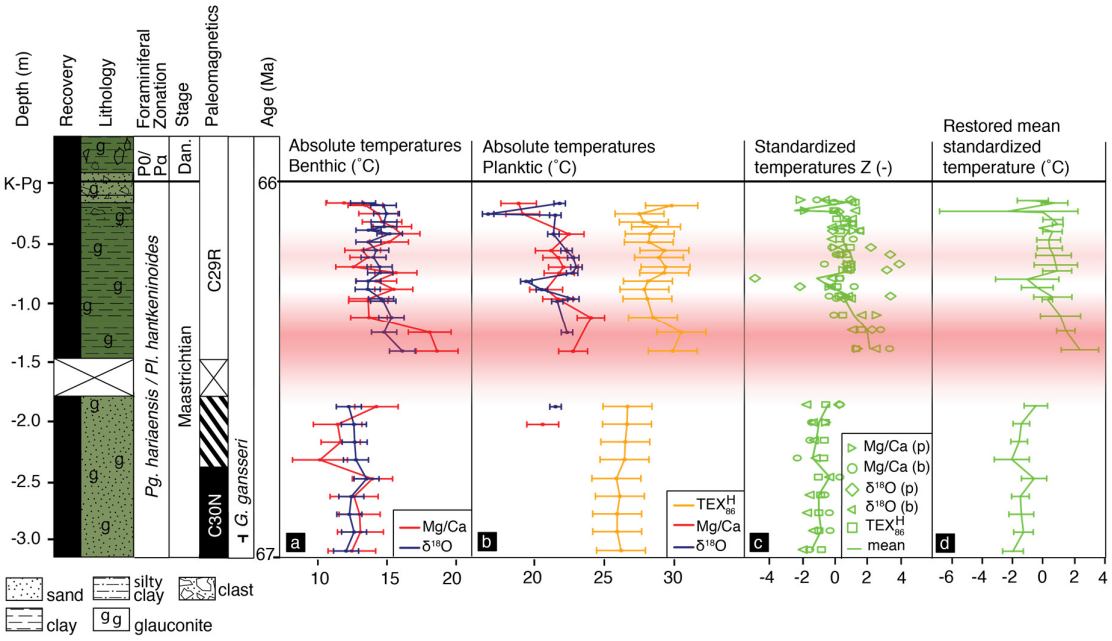


Figure 3 Multi-proxy late Maastrichtian temperature reconstructions at Bass River. Absolute temperature reconstructions based on (a) benthic and (b) planktic proxy records. (c) Standardized temperatures of individual temperature records as well as the mean standardized temperature. (d) Restored mean standardized temperature. Stratigraphy and age model based on Olsson et al. (2002) and Esmeray-Senlet et al. (2015). Red zones tentatively indicate warming. The error bars on the temperature reconstructions based on TEX_{86}^{H}, as well as on the benthic and the planktic foraminifera were calculated using the standard rules for error propagation, using the weighted measurement errors combined with the calibration errors. These uncertainties are the absolute minimal uncertainties associated with the data points, as poorly constrained or unknown uncertainties cannot be fully propagated.

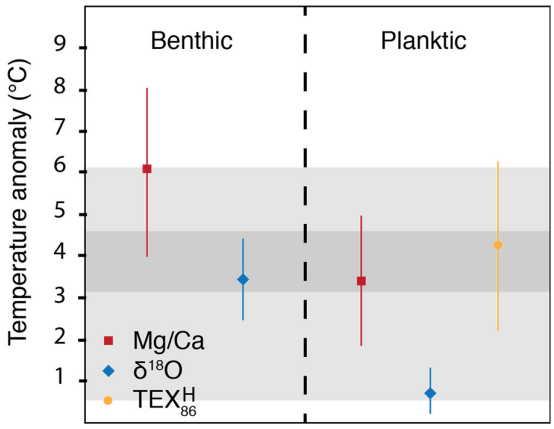


Figure 4 Magnitude of the temperature anomaly during the LMWE for the different proxy-based temperature records of Bass River. Dark gray area: weighted mean average of the LMWE from benthic and planktic Mg/Ca, benthic $\delta^{18}O$ and TEX_{86}-based temperature reconstructions (alternative approach, see section 4.1). Light gray area: integrated multi-proxy estimation of the LMWE (section 2.4).

4. Discussion

4.1. Unknown and unquantified uncertainties

In this study, the error budget for each reconstructed temperature record was set up to be as inclusive and realistic as possible. Yet, it should be noted that despite this effort, it is impossible to include all possible sources of error and uncertainty, since some of these sources are unknown or poorly constrained.

For example, the input of low salinity waters from runoff on the near-shore location of the studied site could have changed under influence of a changing climate, and this could in turn have affected stable oxygen isotopic values. Similarly, unknown diagenetic effects could have influenced proxy values. Yet, since there is no evidence for changes in fresh water inputs near the studied site, nor for diagenetic effects, it is impossible to quantify this kind of unknown uncertainty.

But even if there is some (poor) constraint on a source of uncertainty, it is not recommended to include this source in the uncertainty budget. An example of poorly constrained

uncertainty possibly affecting the outcome of the reconstructions in this study is the assumed constant atmospheric CO_2 concentration over the studied interval. A stable atmospheric CO_2 level, and thus seawater pH, throughout the late Maastrichtian was assumed in this study. CO_2 and concurrently seawater pH may, however, have changed during this interval, which may possibly have influenced proxy values.

But few estimates of changing pCO_2 during the LMWE are available, and these estimates are mostly highly uncertain and/or based on records with low temporal resolution. Nordt et al. (2003), for instance, estimated a change in atmospheric CO_2 of ~600 to ~1400 ppm during the LMWE using oxygen and carbon isotopic ratios (based on only one measurement value in the warming interval). Henehan et al. (2016) estimate that atmospheric CO_2 may have changed from ~600 ppm to max. ~700 ppm, or could have changed as much as from ~600 ppm to ~1300 ppm during the warming interval. An increase in pCO_2 from 600 to 700 ppm would result in a decrease in seawater pH from ~8.1 to ~8.0 (0.1 pH units) (Zeebe, 2001). A ~600 to ~1300 increase would result in a decrease in seawater pH from ~8.1 to ~7.7 (0.4 pH units).

Benthic foraminiferal Mg/Ca does not seem sensitive to changes in pH (Supplementary Information). Some authors show an insignificant change in Mg/Ca over the applied range in pH's (e.g. Kısakürek et al., 2008). Other authors report a modest effect of pH on planktic foraminiferal Mg/Ca in some species (Lea et al., 1999; Russell et al., 2004). Assuming that the equation for *O. universa* from Russell et al. (2004) represents the maximum effect of pH on Mg incorporation, a pH variability of 0.1 unit would result in a maximum variability in Mg/Ca of 0.6 mmol/mol, irrespective of the absolute Mg/Ca values. With a planktic Mg/Ca of ~4.6, the reconstructed temperature would be associated with an additional maximum uncertainty of ~1 ^{0}C (Lea et al., 1999).

The relationships between $\delta^{18}O$, atmospheric CO_2 and sea water carbonate chemistry are reasonably well established (Zeebe, 2001), and potential variations are more pronounced. If atmospheric CO_2 would have changed from 600 to 700 ppm during the LMWE, corresponding to a change in ocean pH of 0.1 pH unit, this would result in a modest apparent shift in $\delta^{18}O$-based seawater temperature of 0.7 ^{0}C (Zeebe, 2001). An increase from ~600 to ~1400 ppm (0.4 pH units) would result in a relatively large apparent temperature shift of 2.7 ^{0}C.

As there always are poorly constrained factors influencing the outcome of any reconstruction, this would imply that reconstructing any paleoenvironmental variable would simply be impossible, as the uncertainty would become much larger than the reconstructed signal. It is therefore recommended to only include well-constrained uncertainties in the error budget, and to be aware of the fact that the calculated uncertainty on a reconstructed variable is always merely a minimum value. Error budgets may improve if these poorly constrained factors become better understood, permitting them to be included in the approach advocated here.

The use of multiple proxies in paleo-reconstructions is strongly recommended, if only because different proxies respond differently to potential unknown or unquantified factors. If individual reconstructions in a multi-proxy reconstruction are in agreement, this implies that they are most likely not severely influenced by unknown or unquantified factors. Multi-proxy studies, therefore, also contribute to a better understanding of deviations among individual proxy records.

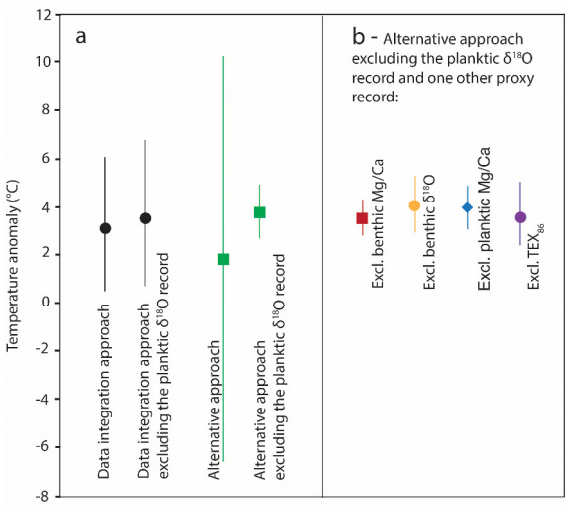
4.2. Magnitude of late Maastrichtian warming at Bass River

Our approach of data integration results in an estimate of late Maastrichtian warming of 3.3±2.8 °C (Light gray zone in Figure 4; Figure 5a), where the error on this reconstructed temperature is the absolute minimum error, as poorly constrained or unknown uncertainties cannot be fully propagated. There are several reasons for this relatively large uncertainty. Firstly, the stratigraphic position of the temperature changes in the warming interval differs between the individual temperature records (Figure 3). This means, for instance, that the maximum temperature may occur at slightly different depths for each individual record. Such differences may have been caused by secondary processes such as reworking, diagenesis and leaching, which may have distorted the original signals of the different individual proxies. When the different records are then combined using the data integration approach, where data from the same stratigraphic position are combined, this may result in large uncertainty. Secondly, not all individual proxy records, and corresponding late Maastrichtian warming magnitudes, are in agreement. For instance, the rather modest planktic $\delta^{18}O$-based late Maastrichtian warming magnitude of 0.8±0.6 °C is not supported by the other proxy records (Figure 3). This is likely the result of an overall lack of pre-warming planktic data in the record. When the data from deviating

415 records are included in the integration, this results in a large uncertainty. Thirdly, including a
416 record that has large uncertainty, such as the TEX_{86}-based temperature record (Figure 3), results
417 in a higher uncertainty for the integrated temperature record. Finally, this approach can only be
418 applied if proxy data of the exact same age (depth) are available. Although this is the case for the
419 Bass River core, this may not be the case for most multi-proxy records.
420 To circumvent this problem and reduce aforementioned uncertainties, the magnitude of the
421 LMWE can also be estimated for each individual record by calculating the difference between the
422 weighted mean of the reconstructed temperatures in the pre-warming interval (-3.11 to -1.45 m)
423 and the maximum temperature in the warming interval (Figure 4). In this approach, the
424 maximum temperature does not necessarily occur at the same depth for every record (Figure 3).
425 In the benthic Mg/Ca record, maximum warming occurs at -1.42 and -1.26 m depth, in the
426 planktic Mg/Ca record at -1.13 m depth, in the benthic $\delta^{18}O$ record at -1.42, -1.26 and -1.13 m
427 depth, and in the planktic $\delta^{18}O$ record and the TEX_{86} record at -1.26 m depth. Subsequently, the
428 weighted mean of the calculated warming magnitude of the individual records can be calculated
429 in an alternative approach. The application of this alternative approach, even when data from the
430 same depth are available, is justified by the fact that aforementioned secondary processes may
431 have distorted the original signals of the different individual proxies. It is stressed, however, that
432 this alternative approach should be used with more caution than the robust technique discussed
433 above, and its performance should be judged in the light of the outcome of the robust technique.
434 When this alternative approach is applied to the entire proxy dataset of Bass River, a
435 warming of 1.9±8.3 °C is indicated (Figure 5a). The large uncertainty of this reconstructed
436 warming magnitude does not allow us to establish whether or not a warming occurred during
437 this interval.
438 Excluding strongly deviating records from data integration during this step of the
439 alternative approach reduces uncertainty significantly. For instance, since the pre-warming
440 interval for the planktic $\delta^{18}O$-based temperature record consists of only one data point, due to
441 the rarity of the investigated planktic foraminiferal species below this level, accurately
442 reconstructing a temperature anomaly for this proxy record is not feasible. Moreover, other,
443 unconstrained (i.e. without uncertainty estimate) planktic $\delta^{18}O$ records from Bass River from this
444 interval indicate a 0.5 to 1.0‰ shift in planktic $\delta^{18}O$ values during this interval, corresponding to

445 a ~2.5 to ~4.5 °C temperature increase (Olsson et al., 2002; Esmeray-Senlet et al., 2015).
446 Disregarding the entire planktic $\delta^{18}O$ record is therefore considered a reasonable decision in this
447 alternative approach. In Figure 5a, it is demonstrated that eliminating the planktic $\delta^{18}O$ record
448 strongly decreases the uncertainty on the estimated late Maastrichtian warming magnitude.
449 Figure 5b furthermore shows that eliminating an additional proxy record from the alternative
450 data integration process does not reduce uncertainty significantly, indicating that the other proxy
451 records are in agreement.

452 Although this pragmatic, alternative approach is less robust than the formal integrated
453 method, which does not change significantly after omitting the planktic $\delta^{18}O$ record (Figure 5a), it
454 allows for a qualitative assessment of the individual temperature reconstructions. The weighted
455 mean of the four proxy records combined (without the planktic $\delta^{18}O$ record) suggests a late
456 Maastrichtian warming magnitude of 3.9±1.1 °C (Figure 5a; dark gray zone in Figure 4). This
457 estimate is in good agreement with the formal, robust, integrated estimate of the LMWE of
458 3.3±2.8 °C, although with considerably smaller uncertainty. In addition, the pragmatic alternative
459 approach allows for the integration of temperature records that are not derived from the exact
460 same stratigraphic depth and/or location, which is useful in case these are not available.
461



462

463 **Figure 5** (a) Magnitude of LMWE (°C) reconstructed using the formal approach as described in
464 Methods (2.4) and the alternative approach described in section 4.1, with and without inclusion

of the planktic $\delta^{18}O$ record. (b) Reconstruction using the alternative approach, excluding one additional proxy record from the analysis.

4.3. Evaluating the robustness of absolute and relative reconstructed temperatures

Although all proxy records display similar trends, the reconstructed absolute bottom and surface water temperatures differ from each other (Figure 3). This observed gradient is in line with the paleohydrographic reconstruction for Bass River by Esmeray-Senlet et al. (2015). Yet, it is possible that the application of modern transfer functions to extinct organisms in deep time resulted in deviating reconstructed absolute temperatures, which may explain the difference between the absolute TEX_{86}^{H} and planktic foraminiferal temperature curves.

The transfer functions that are used to reconstruct temperatures from proxy data are not calibrated for deep-time time intervals, with environmental conditions that are different from those of today, or for extinct (foraminiferal) species. This introduces an unconstrained uncertainty in the reconstructed temperatures, which is not necessarily a simple linear transformation of the calculated temperatures as is the case with the used pH and Mg/Ca_{sw} corrections (see Text S3). The uncertainties regarding the transfer functions used thus affect absolute as well as relative temperatures. Moreover, the uncertainties of individual transfer functions can differ from each other. Using different transfer functions for the same proxy data can substantially impact reconstructed absolute temperatures.

For example, the extinct, non-symbiont bearing *Globotruncana* spp. (D'Hondt and Zachos, 1998) used in the Bass River analysis presumably was a subsurface to intermediate water dweller (Abramovich et al., 2003). Therefore, the use of a Mg/Ca-based transfer function for the surface dweller *Globigerina bulloides*, as done here, is not the most obvious choice, especially since transfer functions based on deeper dwelling taxa such as *Globorotalia truncatulinoides* are also available (e.g. Anand et al., 2003). Yet, temperatures reconstructed using the functions based on *Globorotalia truncatulinoides* yield temperatures that are too high to be realistic, > 10 °C higher than planktic $\delta^{18}O$-based reconstructed temperatures and > 5 °C higher than TEX_{86}-based reconstructed temperatures. Although it should be noted that this offset could partially be caused

493 by the chosen pH and Mg/Ca_{sw} values, it is unlikely that an offset of this magnitude is solely
494 caused by the choice of these parameters.

495 The $\delta^{18}O$-based temperature transfer function used for *Globotruncana* spp. is the function
496 for non-symbiotic surface dwelling *Globigerina bulloides* (13 chambered shell) by Bemis et al.
497 (1998), since no transfer functions that are explicitly designed for subsurface or thermocline
498 dwelling planktic foraminifera are available. However, this choice of transfer function most likely
499 results in absolute reconstructed temperatures that are offset.

500 For TEX_{86}-based temperature reconstructions, a spatially-varying relationship between
501 TEX_{86} and SSTs was recently suggested (Kim et al., 2010; Tierney and Tingley, 2014). This
502 resulted in the development of BAYSPAR, a Bayesian regression approach to the TEX_{86}–SST
503 calibration that explicitly allows for model parameters to smoothly vary as a function of space
504 (Tierney and Tingley, 2014). This is an excellent alternative approach for reconstructing
505 temperatures from TEX_{86} values in more recent times (i.e. Neogene and Quaternary), where
506 TEX_{86} data are both spatially as well as temporally abundant. However, in deep-time, such as the
507 latest Cretaceous, data are sparse, which reduces the value of the approach embodied in
508 BAYSPAR. The decision to reconstruct paleotemperatures based on the relation between
509 measured TEX_{86}^{H} and the corresponding range of temperatures in the modern-day data results in
510 a large uncertainty of temperatures reconstructed using this method for deep-time intervals
511 compared to other methods (Figure S6). In addition, TEX_{86}^{H} is a non-linear model, whereas
512 BAYSPAR is a linear model. For the modern ocean, the two models result in comparable SSTs.
513 However, for the high end of the TEX-SST relationship, BAYSPAR results in higher SSTs than with
514 TEX_{86}^{H} (Figure S6).

515 Recent mesocosm experiments indicate that the TEX_{86} index does not systematically
516 respond to temperature change for all Thaumarchaeal strains (Elling et al., 2015). Instead, the
517 change in temperature is better captured by the Ring Index, which, in contrast to TEX_{86}, includes
518 GDGT-0 and crenarchaeol. The ratio of GDGT-0 versus TEX86 for the global core top calibration
519 of Kim et al. (2010; cf. Cramwinckel et al. (2018)) shows a strongly non-linear relation, indicating
520 that at high temperatures the response of the GDGTs included in the TEX_{86} index to temperature
521 change is limited compared to that of GDGT-0 and crenarchaeol. This suggests that the flattening
522 of the relation between TEX_{86} and SST at high temperatures is not only a consequence of

regression dilution (Tierney and Tingley, 2014) but also a true biophysical mechanism. Hence, we prefer to use a non-linear function to reconstruct SSTs for our latest Cretaceous dataset. Nevertheless, absolute SST values obtained from the extrapolated part of the calibration should always be interpreted with caution. However, trends, timing, and the direction of changes in GDGT-based SST records are reliable and independent of the calibration used.

The effect of using different transfer functions on absolute temperature reconstructions as well as on temperature anomalies (the LMWE) was tested by applying several different transfer functions to each proxy dataset. For an overview of the differences in absolute temperature reconstructions and temperature changes, using different calibrations and transfer functions, amongst others using the BAYSPAR calibration (Tierney and Tingley, 2014) versus the TEX_{86}^{H} calibration, see Figure S2 to S6. The use of different transfer functions indeed often results in significantly deviating absolute temperatures. Yet, the calculated late Maastrichtian temperature anomalies (calculated using the alternative approach, see section 4.1) using different transfer functions are similar (Figure S2 to S6). We therefore argue that calculated temperature changes are far less sensitive to the choice of transfer function, and therefore more reliable than absolute temperatures in deep-time studies.

4.4. Intercomparison of the individual proxy records at Bass River

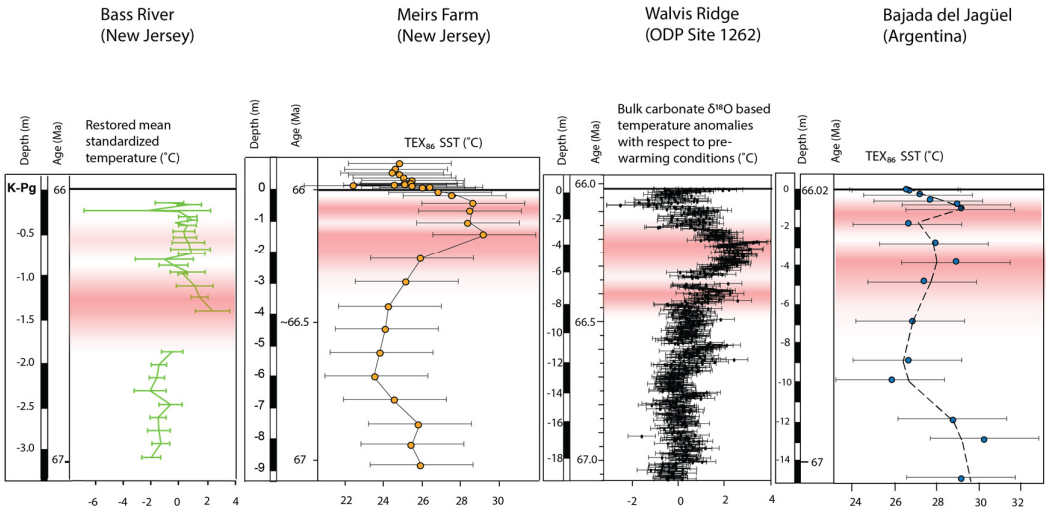
Comparing the calculated magnitude of warming from individual records in a multi-proxy study can provide insight into whether or not these proxies were likely influenced by other factors besides temperature. For instance, large reductions in the size of ice sheets results in more negative seawater $\delta^{18}O$ values. Therefore, $\delta^{18}O$ records during periods of major ice melt would yield higher reconstructed temperatures compared to other temperature reconstructions. If substantial ice sheets had been present on Antarctica during the latest Cretaceous, as suggested by Miller et al. (2005), they likely would have disappeared or significantly been reduced during the LMWE. The input of ice sheet melt water would in turn have led to lower seawater $\delta^{18}O$ values, which would result in an overestimation of the LMWE reconstructed from foraminiferal $\delta^{18}O$-based temperatures. Yet, the amplitude of the warming reflected by $\delta^{18}O$-based temperature reconstructions is similar to that of the other proxies in the same interval (Figure 3, Figure 4). Hence, our data do not suggest that a significant ice sheet was present on

552 Antarctica during the latest Cretaceous, although we cannot exclude the possibility of a minor
553 ice-sheet melting signal, as this may well be obscured by the effects of local circulation patterns
554 and/or hydrological changes.

555 **4.5. Comparison to other relatively high resolution late Maastrichtian records**

556 Other available late Maastrichtian temperature records with uncertainty estimates, and/or
557 records with relatively high resolution, are the TEX_{86}-based temperature records obtained from
558 the New Jersey Shelf (Meirs Farm, inner shelf (Vellekoop et al., 2016)) and from Argentina
559 (Bajada del Jagüel, mid-shelf (Woelders et al., 2017)). SSTs of these records are calculated using
560 the TEX_{86}^{H}-equation by Kim et al. (2010) and are reported to have an uncertainty of ±2.5 °C (Kim
561 et al., 2010). The high resolution temperature record from Walvis Ridge (ODP Site 1262, upper
562 abyssal (Birch et al., 2016; Woelders et al., 2017)) is based on bulk stable oxygen isotopes. The
563 calibration uncertainty on these reconstructed temperatures, derived using the equation by
564 Epstein et al. (1953), is ±0.6 °C (which seems to be an excessively small uncertainty).

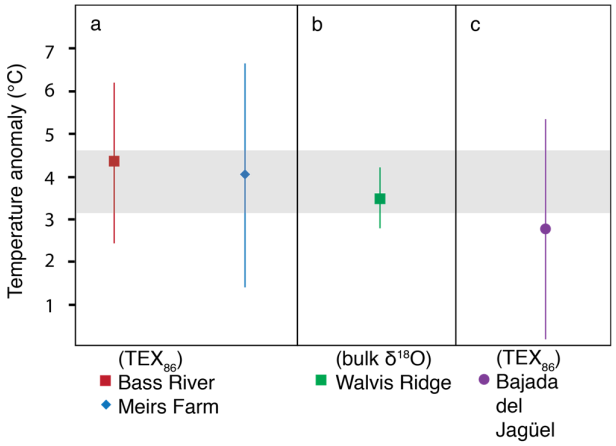
565 These late Maastrichtian temperature records appear to be similar to those from the Bass
566 River record: a relatively stable pre-warming interval is followed by a warming interval that
567 seemingly consists of several smaller warming events. Towards the K-Pg boundary, a cooling is
568 observed (Figure 6). In addition, the estimated magnitude of the LMWE at Bass River can be
569 compared to the magnitude of warming at other locations following the alternative approach
570 described in section 4.1. The estimated magnitude of warming derived from the individual
571 records from these different locations situated in different depositional environments is very
572 similar (Figure 7). Although these (single) proxy records often display a considerably larger
573 uncertainty than our integrated reconstruction, these observations strongly suggest that the
574 LMWE was a global phenomenon.
575



576

577 **Figure 6** Late Maastrichtian temperature records from Bass River (this study), Meirs Farm
578 (Vellekoop et al., 2016), Walvis Ridge (Birch et al., 2016; Woelders et al., 2017) and Bajada del
579 Jagüel (Woelders et al., 2017).

580



581

582 **Figure 7** (a) Magnitude of the LMWE estimated from TEX_{86}-based temperature records from two
583 locations at the New Jersey Shelf (Bass River (this study) and Meirs Farm (Vellekoop et al.,
584 2016)). The magnitude of the LMWE was estimated by comparing the weighted mean of the
585 TEX_{86}-based temperatures of the pre-warming interval with the highest temperature in the
586 warming interval (alternative approach, see section 4.1). (b) Magnitude of the LMWE estimated
587 from a bulk stable oxygen isotope based temperature record from ODP Site 1262 (Walvis Ridge)
588 (Birch et al., 2016). (c) Magnitude of the LMWE estimated from a TEX_{86}-based temperature
589 record from Bajada del Jagüel (Woelders et al., 2017). Light gray area: weighted mean average of

590 the LMWE from benthic and planktic Mg/Ca, benthic $\delta^{18}O$ and TEX_{86}-based temperature
591 reconstructions from Bass River (this study).

592 **5. Conclusions and recommendations**

593 This study provides different strategies to integrate multi-proxy temperature data. Our
594 robust integrated approach results in an estimated 3.3±2.8 °C warming during the late
595 Maastrichtian at Bass River. A more pragmatic approach resulted in an estimation of late
596 Maastrichtian warming of 3.9±1.1 °C. In this approach, anomalous records are excluded from the
597 data integration and only the difference between the pre-warming interval and the highest
598 temperature in the warming interval are compared regardless of when this warming occurred
599 exactly. The error on the reconstructed magnitude of warming is the absolute minimum error as
600 poorly constrained or unknown uncertainties cannot be fully propagated.

601 While our estimates are in agreement with other well-constrained, high resolution
602 temperature records of the late Maastrichtian, our study also shows that combining multiple
603 temperature reconstructions and their uncertainty estimates enables quantitative estimation of
604 past climate change and can reduce uncertainty. It furthermore facilitates a qualitative
605 assessment of the weight of evidence of individual proxy records. In order to be able to do this,
606 however, it is necessary to accurately quantify uncertainties of individual temperature
607 reconstructions.

608 Our approach of integrating multiple temperature records derived by different proxies
609 relies on temperature *trends* rather than on *absolute* temperature estimates, as temperature
610 trends are more robust than absolute reconstructed paleotemperatures, especially in deep-time
611 and over relatively short time intervals. We show that in using $\delta^{18}O$, Mg/Ca and TEX_{86} and
612 present-day transfer functions to obtain temperature trends over relatively short time intervals
613 in the late Cretaceous, differences in these parameters between the present-day and the late
614 Cretaceous can be neglected. In line with this, calculated absolute temperatures in this study
615 should primarily be regarded as an indication for potential absolute temperatures. Yet, it should
616 be noted that rapidly changing $\delta^{18}O_{sw}$, Mg/Ca_{sw} and pH_{sw} values within the short interval from
617 which a trend is extracted, can, in fact, influence reconstructed absolute temperatures differently

throughout this interval. Thereby, these unconstrained factors can also introduce unconstrained uncertainties on the relative temperature estimates.

The uncertainty quantifications in this study suggest that in order to improve multi-proxy based reconstructions of absolute temperatures, the uncertainties within temperature transfer functions must be properly estimated and reduced where possible. To be able to do so more effectively, the three sources of uncertainty as mentioned in the Introduction should all be addressed. Firstly, it is recommended that all studies reconstructing paleo-temperatures include raw measurement data with analytical errors (and replicate analyses if possible). Secondly, raw data on which current as well as future calibrations are based should be made available publicly. This allows calibrations to be improved as insights grow and it allows for the development of calibrations based on larger datasets than the original set. Finally, unknown and unquantified uncertainties that may influence reconstructed temperatures, such as atmospheric CO_2 and local hydrological processes, need to be better constrained. This would make it possible to set up comprehensive error budgets, and greatly improve deep-time multi-proxy based temperature reconstructions.

Acknowledgements

We would like to thank Wim Boer, Stef van Steenberghe and Niels de Winter for their laboratory assistance. We would like to thank Jessica Tierney, Marit Brommer, two anonymous reviewers and the editor for their valuable comments and suggestions, which greatly improved the manuscript. Funding of this project was provided by the Research Foundation Flanders (FWO) to R.P.S. and P.C. (G.0B85.13). NWO grant 834.11.006 enabled the purchase of the UHPLC-MS system used for GDGT analysis at Utrecht University. This research was partially funded by the program of the Netherlands Earth System Science Center (NESSC).

Authorship

L.W. contributed to the conception and design of the study, acquired data, analyzed and interpreted the data and drafted the manuscript. J.V. contributed to the conception and design of the study, acquired data, analyzed and interpreted the data and drafted the manuscript. G.J.W. contributed to the conception and design of the study, analyzed and interpreted the data and

revised the manuscript critically. L.d.N, G.J.R., F.P. and P.C. contributed to data acquisition and revised the manuscript critically. R.P.S. contributed to the conception and design of the study and revised the manuscript critically. All authors approve of this version to be submitted.

662 Abramovich, S., Keller, G., Stüben, D., Berner, Z., 2003. Characterization of late Campanian and
663 Maastrichtian planktonic foraminiferal depth habitats and vital activities based on stable
664 isotopes. Palaeogeogr. Palaeoclimatol. Palaeoecol. 202, 1–29. doi:10.1016/S0031-
665 0182(03)00572-8
666 Anand, P., Elderfield, H., Conte, M.H., 2003. Calibration of Mg/Ca thermometry in planktonic
667 foraminifera from a sediment trap time series. Paleoceanography 18, 1050.
668 doi:10.1029/2002PA000846
669 Barnet, J.S.K., Littler, K., Kroon, D., Leng, M.J., Westerhold, T., Röhl, U., Zachos, J.C., 2017. A new
670 high-resolution chronology for the late Maastrichtian warming event: Establishing robust
671 temporal links with the onset of Deccan volcanism. Geology 46, 147–150.
672 Barrera, E., Savin, S.M., 1999. Evolution of late Campanian-Maastrichtian marine climates and
673 oceans, in: Barrera, E., Johnson, E. (Eds.), Evolution of the Cretaceous Ocean-Climate
674 System. Geological Society of America Special Paper 332. pp. 245–282.
675 Bemis, B.E., Spero, H.J., Bijma, J., Lea, D.W., 1998. Reevaluation of the oxygen isotopic composition
676 of planktonic foraminifera: Experimental results and revised paleotemperature equations.
677 Paleoceanography 13, 150–160.
678 Birch, H.S., Coxall, H.K., Pearson, P.N., Kroon, D., Schmidt, D.N., 2016. Partial collapse of the marine
679 carbon pump after the Cretaceous-Paleogene boundary. Geology 44, 287–290.
680 doi:10.1130/G37581.1
681 Bloemsma, M.R., Zabel, M., Stuut, J.B.W., Tjallingii, R., Collins, J.A., Weltje, G.J., 2012. Modelling the
682 joint variability of grain size and chemical composition in sediments. Sediment. Geol. 280,
683 135–148. doi:10.1016/j.sedgeo.2012.04.009
684 Chenet, A.-L., Quidelleur, X., Fluteau, F., Courtillot, V., Bajpai, S., 2007. 40K–40Ar dating of the
685 Main Deccan large igneous province: Further evidence of KTB age and short duration. Earth
686 Planet. Sci. Lett. 263, 1–15. doi:10.1016/j.epsl.2007.07.011
687 Cramwinckel, M.J., Huber, M., Kocken, I.J., Agnini, C., Bijl, P.K., Bohaty, S.M., Frieling, J., Goldner, A.,
688 Hilgen, F.J., Kip, E.L., Peterse, F., van der Ploeg, R., Röhl, U., Schouten, S., Sluijs, A., 2018.
689 Synchronous tropical and polar temperature evolution in the Eocene. Nature 559, 382–386.
690 doi:10.1038/s41586-018-0272-2
691 D'Hondt, S., Zachos, J.C., 1998. Cretaceous foraminifera and the evolutionary history of planktic

692 photosymbiosis. Paleobiology 24, 512–523.
693 Elling, F.J., Könneke, M., Mußmann, M., Greve, A., Hinrichs, K.-U., 2015. Influence of temperature,
694 pH, and salinity on membrane lipid composition and TEX86 of marine planktonic
695 thaumarchaeal isolates. Geochim. Cosmochim. Acta 171, 238–255.
696 Epstein, S., Buchsbaum, H., Lowenstam, H.A., Urey, H.C., 1953. Variation of O18 content of waters
697 from natural sources. Bull. Geol. Soc. Am. 64, 1315–1326.
698 Esmeray-Senlet, S., Wright, J.D., Olsson, R.K., Miller, K.G., Browning, J. V., Quan, T.M., 2015.
699 Evidence for reduced export productivity following the Cretaceous/Paleogene mass
700 extinction. Paleoceanography 30, 718–738. doi:10.1002/2014PA002724
701 Fraass, A.J., Lowery, C.M., 2017. Defining uncertainty and error in planktic foraminiferal isotope
702 measurements. Paleoceanography 32, 104–122. doi:10.1002/2016PA003035
703 Henehan, M.J., Hull, P.M., Penman, D.E., Rae, J.W.B., Schmidt, D.N., 2016. Biogeochemical
704 significance of pelagic ecosystem function: an end-Cretaceous case study. Philos. Trans. R.
705 Soc. B 371. doi:10.1098/rstb.2015.0510
706 Hopmans, E.C., Schouten, S., Sinninghe Damsté, J.S., 2016. The effect of improved chromatography
707 on GDGT-based palaeoproxies. Org. Geochem. 93, 1–6.
708 doi:10.1016/j.orggeochem.2015.12.006
709 Kim, J.H., van der Meer, J., Schouten, S., Helmke, P., Willmott, V., Sangiorgi, F., Koç, N., Hopmans,
710 E.C., Sinninghe Damsté, J.S., 2010. New indices and calibrations derived from the
711 distribution of crenarchaeal isoprenoid tetraether lipids: Implications for past sea surface
712 temperature reconstructions. Geochim. Cosmochim. Acta 74, 4639–4654.
713 doi:10.1016/j.gca.2010.05.027
714 Kısakürek, B., Eisenhauer, A., Böhm, F., Garbe-Schönberg, D., Erez, J., 2008. Controls on shell
715 Mg/Ca and Sr/Ca in cultured planktonic foraminiferan Globigerinoides ruber (white). Earth
716 Planet. Sci. Lett. 273, 260–269. doi:10.1016/j.epsl.2008.06.026
717 Lawrence, K.T., Woodard, S.C., 2017. Past sea surface temperatures as measured by different
718 proxies - a cautionary tale from the Late Pliocene. Paleoceanography.
719 doi:10.1002/2017PA003101
720 Lea, D.W., Mashiotta, T.A., Spero, H.J., 1999. Controls on magnesium and strontium uptake in
721 planktonic foraminifera determined by live culturing. Geochim. Cosmochim. Acta 63, 2369–

2379.

Lear, C.H., Rosenthal, Y., Slowey, N., 2002. Benthic foraminiferal Mg/Ca-paleothermometry: A revised core-top calibration. Geochim. Cosmochim. Acta 66, 3375–3387.

Li, L., Keller, G., 1998. Abrupt deep-sea warming at the end of the Cretaceous. Geology 26, 995–998.

Marchitto, T.M., Curry, W.B., Lynch-Stieglitz, J., Bryan, S.P., Cobb, K.M., Lund, D.C., 2014. Improved oxygen isotope temperature calibrations for cosmopolitan benthic foraminifera. Geochim. Cosmochim. Acta 130, 1–11.

Martin, P.A., Lea, D.W., Rosenthal, Y., Shackleton, N.J., Sarnthein, M., Papenfuss, T., 2002. Quaternary deep sea temperature histories derived from benthic foraminiferal Mg/Ca. Earth Planet. Sci. Lett. 198, 193–209.

Miller, K.G., Sugarman, P.J., Browning, J.V., Olsson, R.K., Pekar, S.F., Reilly, T.J., Cramer, B.S., Aubry, M.-P., Lawrence, R.P., Curran, J., Stewart, M., Metzger, J.M., Uptegrove, J., Bukry, D., Burckle, L.H., Wright, J.D., Feigenson, M.D., Brenner, G.J., Dalton, R.F., 1998. Bass River site, in: Miller, K.G., Sugarman, P.J., Browning, J.V., Olsson, R.K., Pekar, S.F., Reilly, T.J., Cramer, B.S., Aubry, M.-P., Lawrence, R.P., Curran, J., Stewart, M., Metzger, J.M., Uptegrove, J., Bukry, D., Burckle, L.H., Wright, J.D., Feigenson, M.D., Brenner, G.J., Dalton, R.F. (Eds.), Proceedings of the Ocean Drilling Program, Initial Reports, 174AX. College Station, pp. 5–43. doi:10.2973/odp.proc.ir.174ax.101.1998

Miller, K.G., Wright, J.D., Browning, J.V., 2005. Visions of ice sheets in a greenhouse world. Mar. Geol. 217, 215–231.

Nordt, L., Atchley, S., Dworkin, S., 2003. Terrestrial evidence for two greenhouse events in the latest Cretaceous. GSA Today 13, 4–9.

Olsson, R.K., Miller, K.G., Browning, J. V, Habib, D., Sugarman, P.J., 1997. Ejecta layer at the Cretaceous-Tertiary boundary, Bass River, New Jersey (Ocean Drilling Program Leg 174AX). Geology 25, 759–762.

Olsson, R.K., Miller, K.G., Browning, J. V, Wright, J.D., Cramer, B.S., 2002. Sequence stratigraphy and sea-level change across the Cretaceous-Tertiary boundary on the New Jersey passive margin, in: Koeberl, C., Macleod, K.G. (Eds.), Catastrophic Events and Mass Extinctions: Impacts and beyond. Geological Society of America Special Paper 356. pp. 97–108.

Otto-Bliesner, B.L., Jahn, A., Feng, R., Brady, E.C., Hu, A., Löfverström, M., 2017. Amplified North Atlantic warming in the late Pliocene by changes in Arctic gateways. Geophys. Res. Lett. 44, 957–964.

Pearson, P.N., Ditchfield, P.W., Singano, J., Harcourt-Brown, K.G., Nicholas, C.J., Olsson, R.K., Shackleton, N.J., Hall, M.A., 2001. Warm tropical sea surface temperatures in the late Cretaceous and Eocene epochs. Nature 413, 481–487. doi:10.1038/35097000

Petersen, S. V, Dutton, A., Lohmann, K.C., 2016. End-Cretaceous extinction in Antarctica linked to both Deccan volcanism and meteorite impact via climate change. Nat. Commun. 7, 12079.

Reichart, G.J., Jorissen, F., Anschutz, P., Mason, P.R.D., 2003. Single foraminiferal test chemistry records the marine environment. Geology 31, 355–358.

Russell, A.D., Hönish, B., Spero, H., Lea, D.W., 2004. Seawater carbonate chemistry, processes and elements during experiments with planktonic foraminifera Orbulina universa. Pangaea 68, 4347–4361.

Schoene, B., Samperton, K.M., Eddy, M.P., Keller, G., Adatte, T., Bowring, S.A., Khadri, S.F.R., Gertsch, B., 2015. U-Pb geochronology of the Deccan Traps and relation to the end-Cretaceous mass extinction. Science 347, 9–12.

Schouten, S., Hopmans, E.C., Sinninghe Damsté, J.S., 2013. The organic geochemistry of glycerol dialkyl glycerol tetraether lipids: A review. Org. Geochem. 54, 19–61.

Shackleton, N.J., Kennett, J.P., 1975. Paleotemperature history of the Cenozoic and the initiation of Antarctic glaciation: oxygen and carbon isotope analyses in DSDP Sites 277, 279, and 281. Initial Reports Deep Sea Drill. Proj. 29, 743–755.

Sloan, L.C., Huber, M., Crowley, T.J., Sewall, J.O., Baum, S., 2001. Effect of sea surface temperature configuration on model simulations of “equable” climate in the early eocene. Palaeogeogr. Palaeoclimatol. Palaeoecol. 167, 321–335. doi:10.1016/S0031-0182(00)00245-5

Stanley, S.M., Hardie, L., 1998. Secular oscillations in the carbonate mineralogy of reef-building and sediment-producing organisms driven by tectonically forced shifts in seawater chemistry. Palaeogeogr. Palaeoclimatol. Palaeoecol. 144, 3–19.

Thibault, N., Harlou, R., Schovsbo, N.H., Stemmerik, L., Surlyk, F., 2016. Late Cretaceous (late Campanian–Maastrichtian) sea-surface temperature record of the Boreal Chalk Sea. Clim. Past 12, 1–10. doi:10.5194/cp-12-1-2016

782 Tierney, J.E., Tingley, M.P., 2014. A Bayesian, spatially-varying calibration model for the TEX86
783 proxy. Geochim. Cosmochim. Acta 127, 83–106. doi:10.1016/j.gca.2013.11.026
784 Tobin, T.S., Ward, P.D., Steig, E.J., Olivero, E.B., Hilburn, I.A., Mitchell, R.N., Diamond, M.R., Raub,
785 T.D., Kirschvink, J.L., 2012. Extinction patterns, δ18O trends, and magnetostratigraphy from
786 a southern high-latitude Cretaceous–Paleogene section: Links with Deccan volcanism.
787 Palaeogeogr. Palaeoclimatol. Palaeoecol. 350–352, 180–188.
788 doi:10.1016/j.palaeo.2012.06.029
789 Tobin, T.S., Wilson, G.P., Eiler, J.M., Hartman, J.H., 2014. Environmental change across a terrestrial
790 Cretaceous-Paleogene boundary section in eastern Montana, USA, constrained by carbonate
791 clumped isotope paleothermometry. Geology 42, 351–354. doi:10.1130/G35262.1
792 Vellekoop, J., Esmeray-Senlet, S., Miller, K.G., Browning, J. V, Sluijs, A., van de Schootbrugge, B.,
793 Sinninghe Damsté, J.S., Brinkhuis, H., 2016. Evidence for Cretaceous-Paleogene boundary
794 bolide "impact winter" conditions from New Jersey, USA. Geology 44, 619–622.
795 doi:10.1130/G37961.1
796 Westerhold, T., Röhl, U., Donner, B., McCarren, H.K., Zachos, J.C., 2011. A complete high-resolution
797 Paleocene benthic stable isotope record for the central Pacific (ODP Site 1209).
798 Paleoceanography 26, PA2216. doi:10.1029/2010PA002092
799 Wilf, P., Johnson, K.R., Huber, B.T., 2003. Correlated terrestrial and marine evidence for global
800 climate changes before mass extinction at the Cretaceous-Paleogene boundary. Proc. Natl.
801 Acad. Sci. U. S. A. 100, 599–604. doi:10.1073/pnas.0234701100
802 Woelders, L., Vellekoop, J., Kroon, D., Smit, J., Casadio, S., Prámparo, M.B., Dinarès-Turell, J.,
803 Peterse, F., Sluijs, A., Speijer, R.P., 2017. Latest Cretaceous climatic and environmental
804 change in the South Atlantic region. Paleoceanography. doi:10.1002/2016PA003007
805 Zachos, J.C., Schouten, S., Bohaty, S., Quattlebaum, T., Sluijs, A., Brinkhuis, H., Gibbs, S.J., Bralower,
806 T.J., 2006. Extreme warming of mid-latitude coastal ocean during the Paleocene-Eocene
807 Thermal Maximum: Inferences from TEX86 and isotope data. Geology 34, 737–740.
808 doi:10.1130/G22522.1
809 Zeebe, R.E., 2001. Seawater pH and isotopic paleotemperatures of Cretaceous oceans.
810 Palaeogeogr. Palaeoclimatol. Palaeoecol. 170, 49–57.
811

Supplementary Information

Text S1. Calculation of the weighted mean and inverted Chi² fitting

We have a series of measured x values x_i, with standard deviations Δx_i, which are assumed to be identical to a weighted mean value x_M within bounds defined by statistical uncertainty. If the measurement errors follow a normal distribution, the probability associated with the above assumption may be evaluated by means of the goodness of fit statistic χ^2, defined as Equation (1):

$$\chi^2_{(n-1)} = \sum_{i=1}^{n} \frac{(x_i - x_M)^2}{\Delta x_i^2} \qquad \text{Equation (1)}$$

which follows a χ^2 distribution with n-1 degrees of freedom (Press et al., 1989; Taylor, 1982).

The weighted mean x_M of a series of measurements may be calculated under the assumption of normally distributed measurement errors by taking the derivative of the above equation with respect to x_M (Taylor, 1982). This derivative is by definition equal to 0 at the χ^2 minimum:

$$\frac{\delta \chi^2}{\delta x_M} = \sum_{i=1}^{n} \frac{2\, x_M}{\Delta x_i^2} - \sum_{i=1}^{n} \frac{2\, x_i}{\Delta x_i^2} = 0 \qquad \text{Equation (2)}$$

Solving this equation provides the following expression for the weighted mean x_M

$$x_M = \frac{\sum_{i=1}^{n} \frac{x_i}{\Delta x_i^2}}{\sum_{i=1}^{n} \frac{1}{\Delta x_i^2}} \qquad \text{Equation (3)}$$

The standard deviation of x_M, the so-called "internal" error of the mean, is equal to:

$$\Delta x_{int} = \frac{1}{\sqrt{\sum_{i=1}^{n} \frac{1}{\Delta x_i^2}}} \qquad \text{Equation (4)}$$

A good fit of the data to the weighted mean requires that $\chi^2 \leq \chi^2_{exp}$, where the latter is the expectation value (for n-1 degrees of freedom and probability = 0.5). If the value of χ^2 is larger and the assumption of normally distributed errors is believed to be correct, it may be concluded that the fit is poor. Moreover, it may be argued that the standard deviations of the data points, and hence, the "internal" error of the mean have been underestimated. A revised estimate of the standard deviation of x_M, the so-called "external" error of the mean, may be calculated, which is defined as the measurement error for which χ^2 is equal to the expectation value:

$$\Delta x_{ext} = \Delta x_{int} \sqrt{\frac{\chi^2}{\chi^2_{exp}}}$$ Equation (5)

A good fit of the data to the weighted mean is thus ensured by regarding the maximum of these two values as the "true" error of the weighted mean:

$$\Delta x_M = \max(\Delta x_{int}, \Delta x_{ext})$$ Equation (6)

As a result of this strategy, the error of the weighted mean reflects the combined effects of random statistical variation as well as systematic errors that are not accounted for by data-processing techniques. The presence of such unresolved systematic errors may be suspected in cases where "internal" errors are consistently smaller than "external" errors. It must be stressed that the use of the χ^2 statistic as a tool to assess the magnitude of uncertainties in a series of measurements precludes testing of the goodness of fit, because the above technique relies on the assumption that the fit is good (Press et al., 1989). It is therefore also known as inverted chi-squared fitting.

For the purpose of multi-proxy analysis, we will transform the data points of each proxy record into Z scores:

$$Z_i = \frac{x_i - x_M}{\Delta x_M}$$ Equation (7)

Because Z scores are realizations from a standard normal distribution, we may combine the values of different proxy records corresponding to the same stratigraphic interval, and subject them to statistical analysis.

The Z scores corresponding to the same stratigraphic interval may be averaged, and their sample standard deviations calculated (in case there are at least three records available). The significance of the averaged Z score may be assessed with a one-sample t-test (Davis, 2002) with a population value equal to zero by definition. Hence, we test if the temperature is equal to the mean temperature.

The magnitude of temperature change may be estimated from the averaged Z scores by imposing the inverse transformation:

$$\bar{X} = \bar{Z} \sum_{j=1}^{k} w_j$$ Equation (8)

Where the weight factors are defined as:

$$w_j = \frac{1}{\Delta x_{M,j}}$$ Equation (9)

The variable $\bar{X}$ represents the paleotemperature scaled to the average temperature of the entire record and may be used to estimate temperature changes.

Text S2. Stable oxygen isotope paleothermometry

Raw data can be found in Dataset S1. The temperature transfer function used for benthic foraminifera is the transfer function for *Cibicidoides* by Marchitto et al. (2014) (Equation (10)).

$$\delta_c(\pm\sigma_{ct}) - \delta_w + 0.27 = b(\pm\sigma_b)T + a(\pm\sigma_a)T^2 + c(\pm\sigma_{ct})$$ Equation (10)

in which a=0.0011, b=-0.245 and c=3.58. Furthermore, σ_a=0.0002, σ_b=0.005 and σ_c=0.02. A ice-free world $\delta^{18}O_{sw}$ (δ_w) value of -1.2‰ was assumed (Shackleton and Kennett, 1975). The value for δ_c is the weighted mean of the measured $\delta^{18}O_{test}$ value, which has a weighted measurement error σ_{ct}. Equation (10) was rewritten to Equation (11) to directly obtain temperature.

$$T = \frac{-b-\sqrt{b^2-4\times a\times (c-(\delta_c-\delta_w+0.27))}}{2a}$$ Equation (11)

in which uncertainties on the variables are as mentioned in Equation (10). This equation is only valid for absolute temperatures between 0.6 to 25.6 °C, so this equation has only one feasible solution for each $\delta^{18}O$ value.

The temperature transfer function used for *Globotruncana* spp. (a non-symbiotic (D'Hondt and Zachos, 1998) subsurface to thermocline dweller (Abramovich et al., 2003)) is the function for non-symbiotic surface dwelling *Globigerina bulloides* (13 chambered shell) by Bemis et al. (1998) (Equation (12)).

$$T = a(\pm\sigma_a) - b(\pm\sigma_b) \times (\delta_c(\pm\sigma_{ct}) - \delta_w + c)$$ Equation (12)

in which a=13.6, b=4.77 and c=-0.27. Furthermore, σ_a=0.4 and σ_b=0.27 an ice-free world $\delta^{18}O_{sw}$ (δ_w) value of -1.2‰ was assumed (Shackleton and Kennett, 1975). The value for δ_c is the weighted mean of the measured $\delta^{18}O_{test}$ value, which has a weighted measurement error σ_{ct}.

Text S3. Mg/Ca paleothermometry

Raw data can be found in Dataset S2. Ablation was performed at a repetition rate of 6Hz with an energy density of 1J/cm^2. A circular spot with a diameter of 80µm was selected on a foraminiferal chamber (60 µm in case of small chamber sizes) and the foraminifer was ablated

during 60 seconds or until the ablation crater penetrated the test wall. The aerosol of small particles was transported with a flow rate of 0.8 l/min of helium to a quadrupole ICPMS (Thermo Scientific iCAP-Q) using a smoother for increased signal stability. The helium flow was mixed with approximately 0.3l/min nebulizer argon gas and 5ml/min N_2 gas. Monitored masses included ^{11}B, ^{23}Na, ^{24}Mg, ^{25}Mg, ^{27}Al, ^{39}K, ^{43}Ca, ^{44}Ca, ^{57}Fe, ^{88}Sr, ^{137}Ba and ^{238}U. Potential contamination or diagenesis of the outer or inner layer of calcite was excluded by monitoring the Al and Fe signals. At the start of each series, we analyzed SRM NIST612 and NIST610 glass standard in triplicate that were used for calibration using ^{43}Ca as an internal standard (values based on Jochum et al. (2012)) using an energy density of 5±0.1 J/cm^2. Switching between energy densities of 1 on the carbonates and 5 J/cm^2 on the glass standard has been shown not to affect the analysis (Dueñas-Bohórquez et al., 2009). The Iceland spar, NCHS (NIOZ Calcite House Standard), JCp-1 (Giant clam) and MACS-3 (Synthetic Calcium Carbonate) were analyzed at the start of each series and after every ~10 samples to monitor drift of the measured signal over the course of the analyses. All element to calcium ratios were calculated with the Thermo Qtegra software. The matrix matched standards showed an average relative standard deviation (SD) for Mg/Ca of 2.7%. Measurements with ^{24}Mg values deviating more than twice the standard deviation from the initial dataset of the average values were regarded as outliers and discarded from the dataset. In total, 442 single-chamber Mg/Ca values were obtained for benthic foraminifera and 114 for planktic foraminifera.

For planktic foraminifera, a pH correction is recommended, since a more dominant pH control on Mg/Ca_{test} values is expected for the Maastrichtian, with a relatively low pH_{sw} (Evans et al., 2016b; Kısakürek et al., 2008). The pH correction was done using the Mg/Ca ratio of Evans et al. (2016b) assuming a pH_{sw} of 7.7 for the late Maastrichtian (Tyrrell and Zeebe, 2004) (Equation (13)).

$$Mg/Ca_{ratio} = -0.70 \times pH + 6.7 \qquad \text{Equation (13)}$$

in which Mg/Ca_{ratio} is the Mg/Ca ratio with respect to pH=8.02.

Since benthic foraminiferal Mg/Ca values do not appear to be sensitive to changes in pH, we consider it unnecessary to apply a correction for pH_{sw} on the Mg/Ca_{test} values of benthic foraminifera (Allison et al., 2010; Dueñas-Bohórquez et al., 2011).

Modern-day temperature calibrations based on Mg/Ca values cannot directly be applied to Mg/Ca values of foraminifera that lived in oceans with a considerably different Mg/Ca. Therefore, the measured Mg/Ca values were transformed to modern-day values before the calibrations were applied. This transformation was done by applying a correction for the difference between past Mg/Ca_{sw} and present seawater Mg/Ca ($Mg/Ca_{sw,present}$). For this transformation, a Mg/Ca_{sw} value of 1.25 mol/mol during the latest Cretaceous as reconstructed by Stanley and Hardie

(1998) was used as recommended by Evans et al. (2012) in the nonlinear correction for secular change in Mg/Ca_{sw} for benthic foraminifera (Evans and Müller, 2012) (Equation (14)).

$$Mg/Ca_{test} = \frac{Mg/Ca_{sw}{}^{H}}{Mg/Ca_{sw,present}{}^{H}} \times B\ exp^{AT} \qquad \text{Equation (14)}$$

In this correction function, B and A can be derived from the modern day transfer function. Assuming H=0.31 yields average temperatures that are similar to average temperatures found using the benthic $\delta^{18}O$ calibration, and lies within the range for H of 0.17 to 0.73 found for benthic foraminifer *C. wuellerstorfi* (Evans et al., 2016b). Equation (14) can be rewritten to Equation (15):

$$\log(Mg/Ca_{test}) = AT + \log\left(\frac{Mg/Ca_{sw}{}^{H}}{Mg/Ca_{sw,present}{}^{H}} \times B\right) \qquad \text{Equation (15)}$$

and, subsequently, to Equation (16):

$$T = \frac{\log(Mg/Ca_{test}) - \log\left(\frac{Mg/Ca_{sw}{}^{H}}{Mg/Ca_{sw,present}{}^{H}} \times B\right)}{A} \qquad \text{Equation (16)}$$

Although for planktic foraminifera, preferably, the specific correction method for transforming Mg/Ca_{sw} to present seawater Mg/Ca ($Mg/Ca_{sw,present}$) of Evans et al. (2016a) should be used, temperature reconstructions using this correction do not yield realistic absolute temperature estimates (ranging from 56 to 67 °C). This can have several reasons. Firstly, the Mg/Ca_{test} ranges that we measured in the late Maastrichtian are outside the range for which the correction method can be applied (Evans et al., 2016a). Secondly, the used transfer functions are not applicable to *Globotruncana* spp., which went extinct at the K-Pg boundary. Instead, a more linear response of foraminiferal Mg/Ca to seawater Mg/Ca (Mewes et al., 2015; Raitzsch et al., 2010) may be applicable. Thirdly, it is possible that late Cretaceous Mg/Ca_{sw} was higher than modeled by Stanley and Hardie (1998). Model evidence for the Paleogene suggests Mg/Ca seawater values most likely were below 2.5 mol/mol (Evans and Müller, 2012), where values as low as 2.2 mol/mol have been suggested for late Cretaceous/early Cenozoic Mg/Ca seawater (Gothmann et al., 2015). Given these uncertainties, we decided to apply the same correction function to the planktic Mg/Ca_{test} values data as we did to the benthic Mg/Ca_{test} values (Evans and Müller, 2012) (Equation (14), Equation (15) and Equation (16)) using H=0.17, as this value yields average temperatures similar to average temperatures found using the planktic $\delta^{18}O$ calibration.

The assumed Mg/Ca_{sw} and pH_{sw} that are used in these transformations do have uncertainties, which influence the reconstructed temperatures. For Mg/Ca_{sw}, an uncertainty window of ±0.5 mol/mol is generally assumed (e.g. Evans and Müller, 2012) around a Mg/Ca_{sw} of

1.25 mol/mol (Stanley and Hardie, 1998) or even as large as 2.2 mol/mol (Gothmann et al., 2015). This range in Mg/Ca_{sw} yields an uncertainty of ±1.5°C on the Mg/Ca-based reconstructed temperatures, both for benthic as well as for planktic foraminifera. In addition to this, it should be noted that the uncertainty of Mg/Ca_{sw} also depends largely on the duration of the studied time interval, as Mg/Ca_{sw} likely fluctuated on timescales of tens of millions of years (Stanley and Hardie, 1998). The longer the time interval, the larger the uncertainty. Unfortunately, this uncertainty cannot be constrained. Furthermore, the value for H in the correction for secular change in Mg/Ca_{sw} for benthic foraminifera has a large range (Evans et al., 2016a), which yields temperatures that are 1.5°C cooler to 5.5°C warmer than the benthic Mg/Ca-based temperatures calculated in this study. Assuming a similar range of H for planktic foraminifera yields reconstructed temperatures up to 9°C higher than the planktic Mg/Ca-based temperatures calculated in this study. Finally, assuming past seawater pH to be 7.7 ±0.1 yields an uncertainty of ±0.5°C on the reconstructed temperatures based on planktic Mg/Ca_{test} values.

Despite the fact that these uncertainties are, or can be, considerable, they only result in a linear transformation of absolute temperature reconstructions (Equation (13) and Equation (16)). Relative temperature changes are not affected by these uncertainties.

The temperature transfer function used for benthic foraminifera is the function for *Cibicidoides* by Lear et al. (2002) (Equation (17)) which is suitable for relatively warm water conditions and is based on *Cibicidoides* species, as epibenthic *Cibicidoides* morphologically resembles extinct species *Anomalinoides midwayensis* used in this study.

$$Mg/Ca_{test}(\pm \sigma_{ct}) = a(\pm \sigma_a) \times e^{b(\pm \sigma_b) \times T} \quad \text{Equation (17)}$$

in which a=0.867 and b=0.109. Furthermore, σ_a=0.049 and σ_b=0.007. Mg/Ca_{test} is the corrected (see above), weighted mean value for Mg/Ca_{test}, with accompanying weighted standard error σ_{ct}. Equation (17) was rewritten to Equation (18) to directly obtain temperature.

$$T = \frac{\log(Mg/Ca_{test}) - \log(a)}{b} \quad \text{Equation (18)}$$

in which the errors on the variables are as mentioned in Equation 8.

For extinct planktic *Globotruncana* spp., we use the Mg/Ca (Lea et al., 1999) (Equation (19) rewritten to Equation (20)) transfer function based on non-symbiont bearing *Globigerina bulloides*.

$$Mg/Ca_{test}(\pm\sigma_{ct}) = a(\pm\sigma_a) \times e^{b(\pm\sigma_b) \times T} \quad \text{Equation (19)}$$

in which a=0.51 and b=0.104. Furthermore, σ_a=0.1 and σ_b=0.01. Mg/Ca_{test} is the corrected (see above), weighted mean value for Mg/Ca_{test}, with accompanying weighted standard error σ_{ct}. Equation (19) was rewritten to Equation (20) to directly obtain temperature.

$$T = \frac{\log(Mg/Ca_{test}) - \log(a)}{b} \qquad \text{Equation (20)}$$

in which the errors on the variables are as mentioned in Equation (19).

To estimate the uncertainty of the predicted T (Equation (18) and Equation (20)), the standard rules of error propagation were used, using the analytical errors and the uncertainties on the variables in the temperature equations. For the uncertainty on the Mg/Ca term, the value of log(Mg/Ca) was calculated for all individual measurements on a sample. Then, for each sample, the average and standard deviation of the log(Mg/Ca) values was calculated. As such, the uncertainty for the log(Mg/Ca) term as a whole is estimated, because average and standard deviation (error) of the log(Mg/Ca) values is not identical to the logarithm of the mean and standard deviations of the average of the Mg/Ca values. This is in full accordance with the principles of compositional data analysis (Aitchison, 1986; Weltje and Tjallingii, 2008). The fact that log(Mg/Ca) appears in the equation for T, and not Mg/Ca, demonstrates this.

Text S4. Quality control of the TEX_{86} analysis

Raw data can be found in Dataset S3. TEX_{86} index values were calculated for the duplicate measurements and in principal converted into sea-surface temperatures using the TEX_{86}^{H} calibration from Kim et al. (2010) (Equation (21)).

$$SST = 68.4 \times \log(TEX_{86}) + 38.6 \qquad \text{Equation (21)}$$

In the samples analyzed for this study, concentrations of isoprenoidal GDGTs range from 18-122 ng/g dry-weight sediment. The overall chromatography of the GDGTs was excellent, with signal to noise ratios >3. In recent years, it has been shown that TEX_{86}-inferred temperatures can deviate significantly from actual temperatures, depending on environmental conditions and archaeal community composition (e.g. Qin et al., 2015). The reconstructed absolute temperatures should thus be interpreted with care. Here we nevertheless predominantly aim to reconstruct trends i.e. changes in SST rather than absolute temperatures. While TEX_{86} is calibrated to SST (Schouten et al., 2002), various studies suggested TEX_{86} might sometimes reflect deeper water temperatures (e.g. Huguet et al., 2007; Lopes dos Santos et al., 2010). However, the studied site was deposited in a shallow, shelf depositional environment (water depth of 110 ± 20m; (Esmeray-Senlet et al., 2015)), excluding the possibility that the recorded temperature trends would significantly differ from trends in SST.

High concentrations of Soil Organic Matter (SOM) in sediments may cause a bias in TEX_{86}-reconstructed sea-surface temperatures due to the input of terrestrial isoprenoidal GDGTs (Weijers et al., 2006). The relative contribution of SOM to marine sediments can be approximated based on the ratio of branched GDGTs, which are thought to have a primarily soil bacterial origin, to that of crenarchaeol, which is produced by marine Thaumarchaeota. This ratio can be quantified using the so-called Branched and Isoprenoid Tetraether (BIT) index (Hopmans et al., 2004). To identify whether our TEX_{86} record is biased by the input of SOM, we calculated the BIT index for all our samples. All of the 35 samples analyzed in this study have BIT-index values of <0.1, well below the recommended threshold of 0.3 (Weijers et al., 2006) (Dataset S3). This signifies that our TEX_{86}-reconstructed sea-surface temperatures are not biased by terrestrial-derived GDGTs.

A possible bias in TEX_{86} palaeothermometry may also be introduced by the input of methanogenic and methanotrophic archaeal GDGTs, leading to erroneous SST reconstructions (Blaga et al., 2009; Weijers et al., 2011; Zhang et al., 2011). Potential contributions of methanogenic and methanotrophic archaeal GDGTs can be recognized using the ratio of GDGT-0/Crenarchaeol (Blaga et al., 2009) and the Methane Index (Zhang et al., 2011), respectively. In our study, the GDGT-0/crenarchaeol ratio ranges between 0.02 – 0.35, well below the recommended threshold of 2.0 (Blaga et al., 2009), whereas the Methane Index (MI) ranges between 0.05-0.31, below the recommended threshold of 0.5 (Zhang et al., 2011) (Dataset S3). These values suggest that at the studied locality there is little to no input of GDGTs derived from methanogenic or methanotrophic archaea.

The cultivation experiments of Qin et al. (2015) have shown that oxygen limitation can also result in large deviations in reconstructed temperatures. Hence, TEX_{86} results should be interpreted with caution in records that contain indications of low-oxygen conditions or underlie oxygen minimum zones. Fortunately, the Bass River record represents a shallow marine, well-oxygenated environment (Esmeray-Senlet et al., 2015; e.g. Miller et al., 2010). This indicates that oxygen limitation played no role in the record of Bass River. Recently, Zhang et al. (2016) demonstrated that the Ring Index (RI), the weighted average of cyclopentane moieties, can be used determine if TEX_{86} temperature estimates are influenced by non-thermal factors and/or deviate from the modern global TEX_{86}-RI relationship. In our study, average TEX_{86} and average RI per sample are tightly coupled and fall within the modern TEX_{86}-RI trend (Dataset S3; Figure S1). Therefore, we are confident that the GDGT distributions of our samples are primarily a function of temperature.

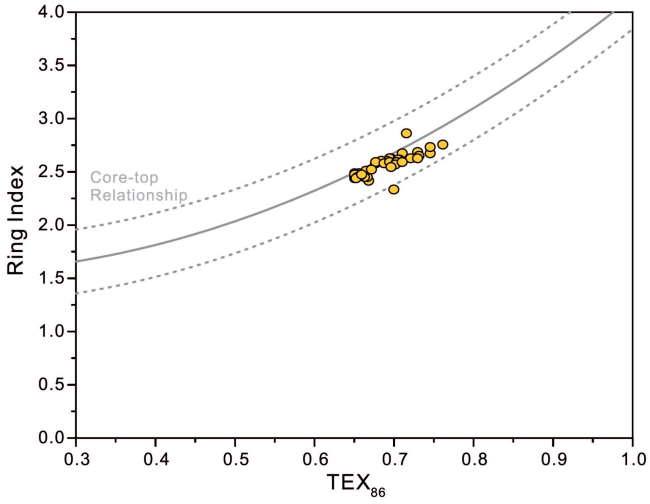


Figure S1 Ring Index plotted against TEX_{86}

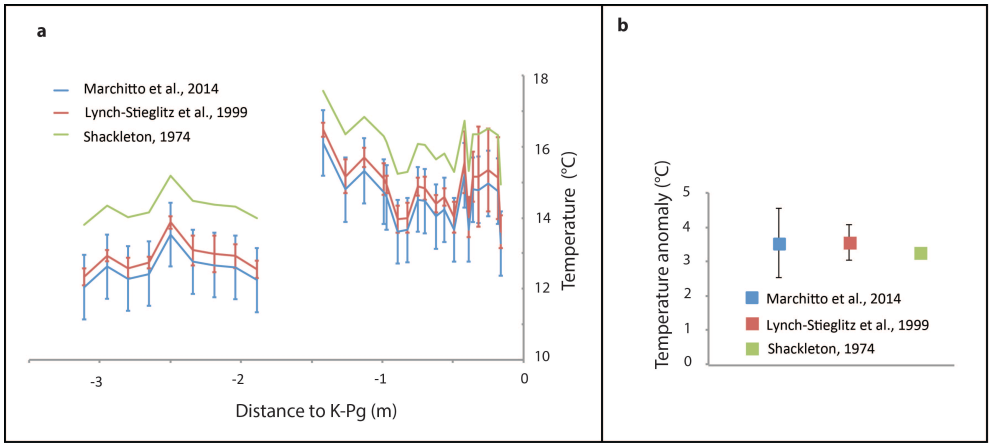


Figure S2 Absolute temperature reconstructions (a) and temperature difference between the warming interval and the pre-warming interval (b) calculated using the benthic $\delta^{18}O$-based transfer functions of Marchitto et al. (2014) and Lynch-Stieglitz (1999) for *Cibicidoides,* and of Shackleton (1974) for *Uvigerina.* Note that the transfer function of Shackleton (1974) has no uncertainty estimation.

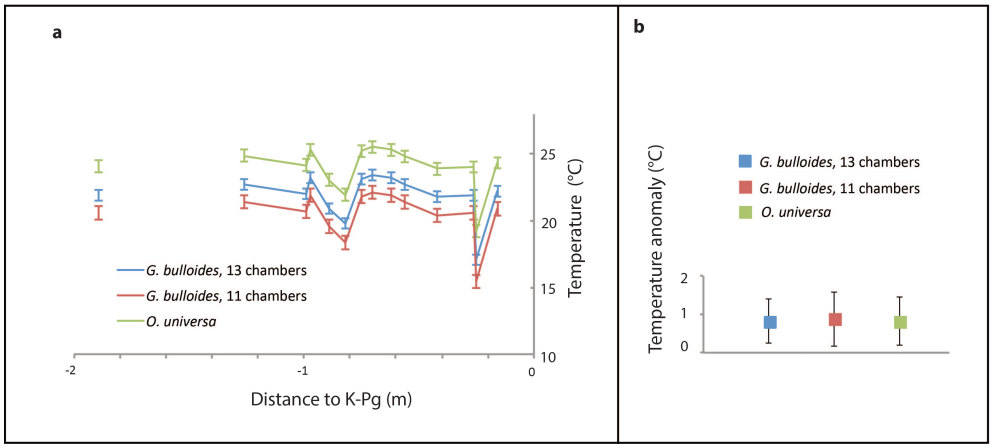


Figure S3 Absolute temperature reconstructions (a) and temperature difference between the warming interval and the pre-warming interval (b) calculated using the planktic $\delta^{18}O$-based transfer functions of Bemis et al. (1998) for *Globigerina bulloides* with 13 chambers, *G. bulloides* with 11 chambers, and *Orbulina universa.*

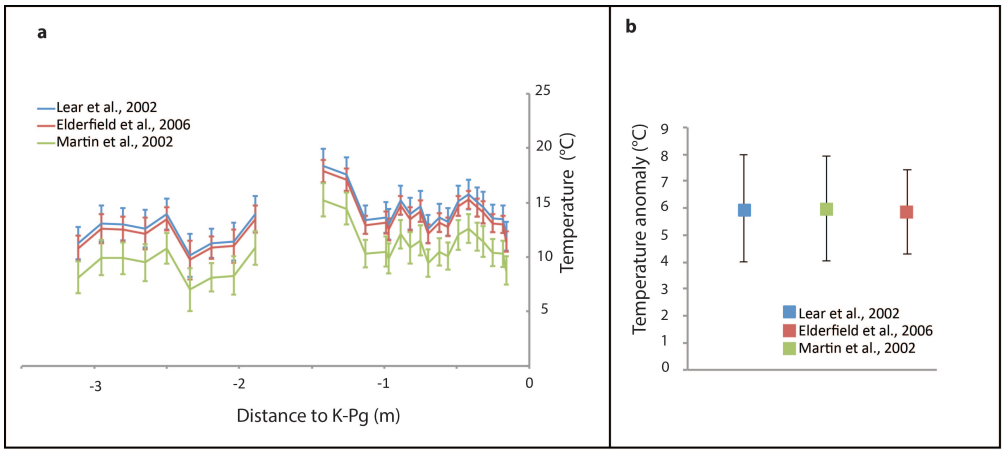


Figure S4 Absolute temperature reconstructions (a) and temperature difference between the warming interval and the pre-warming interval (b) calculated using the benthic Mg/Ca-based transfer functions of Lear et al. (2002), Elderfield et al. (2006) and Martin et al. (2002) for *Cibicidoides* spp.

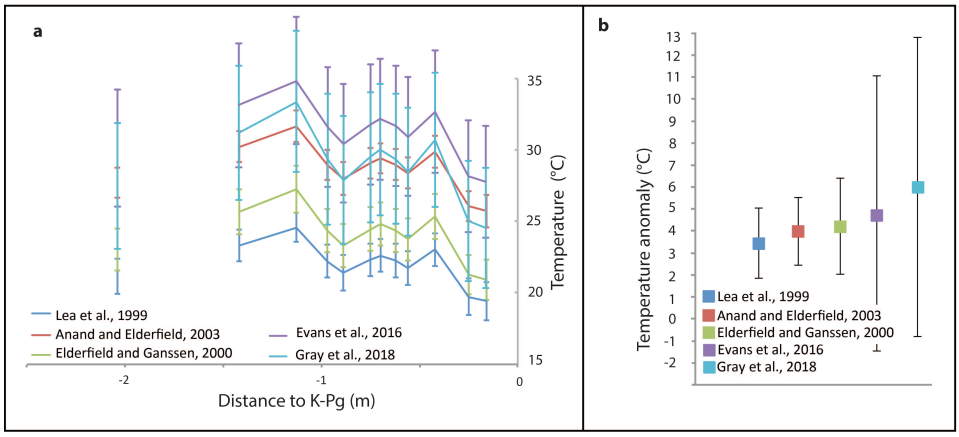


Figure S5 Absolute temperature reconstructions (a) and temperature difference between the warming interval and the pre-warming interval (b) calculated using the planktic Mg/Ca-based transfer functions of Lea et al. (1999) for *G. bulloides,* of Anand and Elderfield (2003) based on 10 planktic species, of Elderfield and Ganssen (2000) based on 8 planktic species, of Evans et al. (2016a) and of Gray et al. (2018), both based on *G. ruber*.

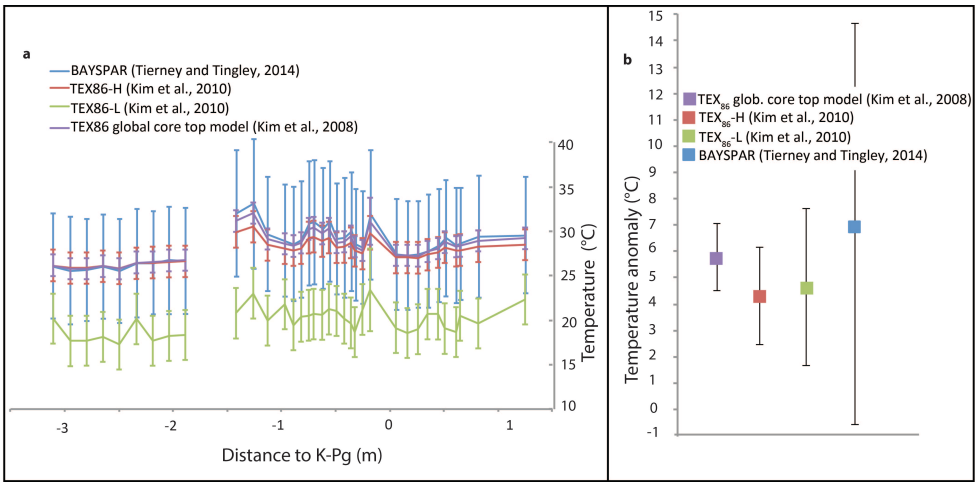


Figure S6 Absolute temperature reconstructions (a) and temperature difference between the warming interval and the pre-warming interval (b) calculated using the TEX_{86}-based core-top calibration transfer function of Kim et al. (2008), the TEX_{86}^{H} and TEX_{86}^{L} transfer functions (Kim et al., 2010) and the BAYSPAR method (Tierney and Tingley, 2014) (prior standard deviation 20, maximum distance 500, minimum number 1, number of samples 5000; conform instructions on the website bayspar.geo.arizona.edu. Error bar is 1 SD.).