Supporting Information for "A century of observed temperature change in the Indian Ocean"

J.O. Wenegrat[1], E. Bonanno[1], U. Rack[2], and G. Gebbie [3]

[1]Department of Atmospheric and Oceanic Science, University of Maryland, College Park, USA

[2]School of Earth and Environment, University of Canterbury Christchurch, NZ

[3]Woods Hole Oceanographic Institution, Woods Hole, USA

Contents of this file

1. Text S1 to S4
2. Figures S1 to S8

Introduction

This document contains supporting information for Wenegrat et al. 'A century of observed temperature change in the Indian Ocean', under review for publication in *Geophysical Research Letters*.

Text S1: Vertical profile of temperature difference

Basinwide average profiles are calculated following the method detailed in Gebbie and Huybers (2019, their supplementary information S5.4), as updated here.

The temperature difference between the historical observations and the corresponding World Ocean Atlas (WOA) value is,

$$\Delta T(r_i) = T(r_i, t_w) - T(r_i, t_h), \tag{1}$$

where $T(r_i, t_h)$ is the ith *historical* temperature observation at location, r_i, and time, t_h, and $T(r_i, t_w)$ is the WOA temperature at the same location. The observations are combined into a vector,

$$\mathbf{\Delta T} = \begin{pmatrix} \Delta T(r_1) \\ \Delta T(r_2) \\ \vdots \\ \Delta T(r_M) \end{pmatrix}. \tag{2}$$

Temperature changes at a given pressure are assumed equivalent to potential temperature changes.

Basinwide-average temperature profiles:

Our goal is to extract the decadal signal of water-mass change from the historical temperature observations

$$\overline{\mathbf{\Delta T}} = \begin{pmatrix} \overline{\Delta\theta(z_1)} \\ \overline{\Delta\theta(z_2)} \\ \vdots \\ \overline{\Delta\theta(z_K)} \end{pmatrix}, \tag{3}$$

where we have defined a grid of K depths. Given knowledge of the basinwide averages, one can make a prediction for each WOA−historical temperature difference,

$$\mathbf{\Delta T} = \mathbf{H}\overline{\mathbf{\Delta T}} + \mathbf{q}, \tag{4}$$

where $\mathbf{H}$ maps the basinwide mean onto the observational point by noting the basin of the observations and vertical linear interpolation, $\mathbf{q}$ is contamination by measurement error and signals that are not representative of the decadal-mean, basinwide-average temperature. The contamination is decomposed into three parts,

$$\mathbf{q} = \mathbf{n}_T + \mathbf{n}_S + \mathbf{n}_M, \tag{5}$$

where $\mathbf{n}_T$ is contamination by transient effects such as isopycnal heave due to internal waves or mesoscale eddies, $\mathbf{n}_S$ is due to the irregular spatial sampling of each basin, and $\mathbf{n}_M$ is measurement or calibration error of the thermometer. Note that no depth correction is made here, and temperature differences may be biased toward warming (as discussed below in section S2).

The expected size of $\mathbf{n}_T$ is related to the energy in the interannual and higher-frequency bands. We use estimates from the WOCE Global Hydrographic Climatology (Gouretski & Koltermann, 2004) to quantify this error and its spatial pattern. Errors that primarily reflect an uncertainty due to a representativity error were previously estimated in this climatology, where the magnitude of interannual temperature variability is 1.6°C at the surface, decreasing to 0.8°C below the mixed layer, and 0.02°C at 3000 meters depth. Inherent in their mapping is a horizontal lengthscale of $L^T_{xy} = 450$ km. This corresponds to a vertical lengthscale of $L^T_z = 450$ meters when applying an aspect ratio based upon mean depth and lateral extent of the ocean. Their mapping is the degree of error necessary to place the non-synoptic cruises of a 10-year time interval into a coherent picture. Estimated errors are similar to those of (Wortham & Wunsch, 2014), who also note that the spatial

scales increase as the temporal scales increase. Above 1300 meters depth, the aliased variability is typically larger than the measurement error described below.

Next we describe the second moment matrix of temporal contamination, $\mathbf{R}_{TT} =< \mathbf{n}_T(\mathbf{n}_T)^T >$. Note that $\mathbf{n}_T$ depends on the difference of contamination during the two time periods, $n_T(r_i) = \eta_T(r_i, t_w) - \eta_T(r_i, t_h)$, where $\eta_T(r,t)$ is the difference between temperature at a given time and the decadal average. The WGHC statistics give the error covariance for $\eta_T(r, t_w)$ not $n_T(r)$. This covariance matrix is reconstructed by first creating a correlation matrix,

$$\mathbf{R}_\rho = \begin{pmatrix} \rho(0) & \rho(\delta) & \rho(2\delta) & \dots \\ \rho(\delta) & \rho(0) & \rho(\delta) & \dots \\ \rho(2\delta) & \rho(\delta) & \rho(0) & \dots \\ \vdots & \vdots & \vdots & \ddots \\ & & & \rho(0) \end{pmatrix}, \tag{6}$$

where the autocorrelation function, $\rho(\delta)$, is given by a Gaussian with a horizontal lengthscale of 450 km and a vertical lengthscale of 450 meters. We derive the covariance matrix by pre- and post-multiplying the correlation matrix, $\mathbf{R}_{\eta\eta} = \boldsymbol{\sigma}_\eta \boldsymbol{\sigma}_\eta^T \circ \mathbf{R}_\rho$, where $\boldsymbol{\sigma}_\eta$ is the vector of the standard deviation of the WGHC interannual variability and $\circ$ is the Hadamard product. Here the time interval of the historical cruises is about 30 years, or three times as long as the WOCE era. Due to the red spectrum of ocean variability, the potential for aliased variability over this longer time interval is increased. To get a better constraint on T_{ratio}, we have to assume a frequency spectrum. If we assume the power density spectrum is red with a power law of f^{-2}, then we can integrate to determine the variance at frequencies greater than 1/(30 yr) and 1/(10 yr). The variance at frequencies greater than f is proportional to 1/f, so the ratio of variance greater than 1/(30 yr) to that greater than 1/(10 yr) is $T_{ratio} = 30/10 = 3$. Both the modern and

historical intervals have variability and are assumed to be statistically independent, and thus, $\mathbf{R}_{TT} = (T_{ratio} + 1)\mathbf{R}_{\eta\eta}$.

We assume that the variance due to spatial water-mass variability, i.e., $\mathbf{R}_{SS} =< \mathbf{n}_S(\mathbf{n}_S)^T >$, has a magnitude that is 20% that of the temporal variability as the local water-mass variability on interannual scales is dwarfed by heaving motions (Huang, 2015). The relevant parameter is $S_{ratio} = 0.2$. These water-mass variations are assumed to have a larger spatial scale ($L^S_{xy} = 2000$ km horizontally, $L^S_z = 1$ km vertically), as seen in an evaluation of water-mass fractions on an isobaric surface (Gebbie & Huybers, 2010). Accounting for this spatial variability has the potential to increase the final error of our estimates by taking into account biases that may occur due to the specific expedition tracks. Numerically, we calculate $\mathbf{R}_{SS}$ in two steps. We form a new $\mathbf{R}_{\eta\eta}$ correlation matrix that takes into account the water-mass lengthscales. Then we adjust the variance according to S_{ratio} via the equation, $\mathbf{R}_{SS} = S_{ratio}(T_{ratio} + 1)\mathbf{R}_{\eta\eta}$.

Finally, we assume that the measurement covariance, R_{MM}, is a matrix with the diagonal equal to the observational uncertainty, $\sigma_{obs} = 0.14$°C, squared (Roemmich et al., 2012).

We solve for the basinwide-average temperature profiles using a weighted and tapered least-squares formulation that minimizes,

$$J = \mathbf{q}^T\mathbf{R}^{-1}_{qq}\mathbf{q} + \mathbf{m}^T\mathbf{S}^{-1}\mathbf{m}, \tag{7}$$

where $\mathbf{R}_{qq}$ reflects the combined effect of the three types of errors (i.e., $\mathbf{R}_{qq} = \mathbf{R}_{TT} + \mathbf{R}_{SS} + \mathbf{R}_{MM}$). This least-squares weighting is chosen such that the solution coincides with the maximum likelihood estimate (assuming that the prior statistics are normally

distributed and appropriately defined). Only a weak prior assumption, reflected in the weighting matrix, $\mathbf{S}$, is placed on the solution, namely that the correlation lengthscale is $L_z^{AVG} = 500$ m in the vertical, the variance is on the order of $(\sigma_S = 1°\text{C})^2$, and the expected value is $< \overline{\mathbf{\Delta T}} >= 0$. The least-squares estimate is then,

$$\overline{\tilde{\mathbf{\Delta T}}} = (\mathbf{H}^T \mathbf{R}_{qq}^{-1} \mathbf{H} + \mathbf{S}^{-1})^{-1} \mathbf{H}^T \mathbf{R}_{qq}^{-1} \mathbf{\Delta T}. \tag{8}$$

The error covariance of the estimate is,

$$\mathbf{C}_{\tilde{\Delta T}} = (\mathbf{H}^T \mathbf{R}_{qq}^{-1} \mathbf{H} + \mathbf{S}^{-1})^{-1}, \tag{9}$$

where the standard error is $\boldsymbol{\sigma}_{\tilde{\Delta T}} = \sqrt{\text{diag}(\mathbf{C}_{\tilde{\Delta T}})}$. This method also recovers the off-diagonal terms that correspond to the correlated errors among different parts of the basinwide-average.

Ocean heat content change

Ocean heat content change, $\Delta\mathcal{H}$, is a linear function of the temperature change and can be written as an inner vector product:

$$\Delta\mathcal{H} = \mathbf{h}^T \overline{\tilde{\mathbf{\Delta T}}}, \tag{10}$$

where $\mathbf{h}$ is a vector containing coefficients related to ocean heat capacity, seawater density, the representative area of the Indian Ocean, and the integration of temperature change over the vertical dimension. Here we integrate to a depth of $z\star = 700$ m so that we obtain heat content change from the sea surface to this depth. The Indian Ocean area is assumed to be equal to 15% of the global ocean area at all depths (ignoring the hypsometric effect).

The error covariance of $\Delta\mathcal{H}$ is an outer product,

$$\mathbf{C}_H = < (\mathbf{h}^T \overline{\tilde{\mathbf{\Delta T}}})(\mathbf{h}^T \overline{\tilde{\mathbf{\Delta T}}})^T >, \tag{11}$$

where <> refers to the expected value. $\mathbf{C}_H$ is a scalar like $\Delta\mathcal{H}$. Rearranging this equation, we obtain,

$$\mathbf{C}_H = \mathbf{h}^T \mathbf{C}_{\tilde{\Delta T}} \mathbf{h}, \tag{12}$$

where $\mathbf{C}_{\tilde{\Delta T}}$ is known from the calculation of the previous section. The standard error of the heat content change is the square root of $\mathbf{C}_H$. Trends are estimated from $\Delta\mathcal{H}$ assuming the historical and WOA observations are representative of their mean observation year of 1887 and 2011, respectively.

Text S2: Errors due to line-stretch

The *Gazelle* used hemp line for profiling, which can stretch under the weight of the instruments and bottom-weight. This would tend to bias the reported *Gazelle* depths shallow, introducing a *warming* bias in the modern minus historical data. The *Valdivia* and *Planet* both used wire for profiling, which is less subject to stretch.

To assess the magnitude of this error we define two temperatures using WOA data. The first, T_{rep}, is found by interpolating the WOA data to the position and reported depth of the historical observations. The second, T_{adj}, is found by interpolating the WOA data to a stretch-corrected depth. Gebbie and Huybers (2019) compared bottom depths reported by the *Challenger* with modern bottom depths, and inferred a 4% shallow-bias in the reported depths, consistent with hemp line loaded to 25% breaking strength. We use this estimate here to correct the *Gazelle* depths. From this we can define a temperature error as $T_{err} = T_{adj} - T_{rep}$ such that positive values indicate warm biases in the modern minus historical estimates.

Profiles of T_{err} are shown in figure S3. Errors are identically zero at the surface, and increase approximately linearly down to 750 m. Below this depth the errors decrease due to the weak interior temperature gradients. The maximum error in the *Gazelle* data is estimated to be 0.31 °C at 750 m depth, however this error decreases to 0.17 °C in the basin-wide mean across all cruises (where we have assumed the wire used on the *Valdivia* and *Planet* introduces no errors in reported depth).

Text S3: Aliasing of temporal variability in the historical measurements

While the focus of this work is on multidecadal temperature variability, other timescales may be aliased into the historical observations, which could affect our estimates of 20th century temperature change. The use of data from 3 separate cruises spread over the period 1874–1907 may help alleviate this—and temporal aliasing is accounted for in the estimates of the basinwide means (section S1)—however the observations from each cruise are also not distributed uniformly through the basin (figure S1) suggesting temporal variability could alias into the observed spatial structure of the temperature change.

A prominent pattern of temperature variability here is the Indian Ocean Dipole (IOD) (N. H. Saji et al., 1999). Positive IOD events are associated with anomalously cold SST in the eastern Indian Ocean, and anomalously warm SST in the west. The pattern is reversed for negative events. Figure S4 shows the SST from ERSST reanalysis averaged over the western and eastern tropical Indian Ocean, and a Dipole Mode Index constructed from the HadSST reanalysis (N. Saji & Yamagata, 2003). The *Gazelle* sampled during the transition from a negative IOD event to neutral conditions, with a weak cool anomaly in the western tropical basin. The cruise track however was largely confined to latitudes

south of 30°S, where IOD temperature anomalies are smaller (N. Saji & Yamagata, 2003). The *Planet* also sampled at the onset of a more strongly negative IOD event, with observations at low latitude where temperature anomalies are strongest. However, we note that the absolute magnitude of the temperature anomalies evident in the ERSST reanalyses were not particularly large during this period, suggesting the effect of the IOD on the *Planet* observations may be more limited than implied by the gradient-based calculation of the Dipole Mode Index. The *Valdivia* cruise was during neutral IOD conditions. Finally, we note that El-Niño-Southern Oscillation (ENSO) variability also affects Indian Ocean SST, however none of the 3 cruises appear to have been during periods of strong ENSO events (Gergis & Fowler, 2009).

Finally, to test for possible aliasing of inter-cruise variability into the zonal-mean spatial pattern (eg. figure 3) we recalculate the latitude-depth section removing one cruise at a time (figures S5, S6, and S7). From this it can be seen that, notwithstanding data gaps, the basic pattern of interior temperature change is robust to the removal of individual cruise data.

Text S4: Statistical robustness of the temperature change pattern

The historical observations are sparse, and the calculated temperature differences are noisy, such that significant averaging is required for statistical inference. However, it is also apparent that there is a striking similarity between the spatial pattern of temperature changes observed in the historical data and the late 20th century changes in the modern observational record (cf. figures 3 and 4). In the main text we therefore present

the latitude-depth section of the modern minus historical temperature changes (figure 3) with bin sizes chosen principally for visual clarity, despite the observations being too underpowered to provide meaningful statistics on this scale. The consistency of the pattern seen in the historical data with multiple independent lines of evidence, as discussed in the text, provides a measure of confidence in its physical interpretation.

However, we also provide here a more rigorous assessment of the broad pattern of 20th century temperature change highlighted in the text. To do this we bin average the modern minus historical temperature change in larger bins spanning 10° of latitude, and 500 m depth (figure S8). The resulting field is broadly similar—albeit greatly smoothed— to the less heavily averaged version in the main text (figure 3). To determine regions where the null hypothesis of zero mean temperature change can be rejected it is necessary to control the false discovery rate associated with multiple hypothesis testing. We use the method outlined by Wilks (2016), where local null hypotheses are rejected if their p-values (based on the standard t test) are smaller than a threshold value, p^*,

$$p^* = \max_{i=1,\ldots,N} \left[p_{(i)} : p_{(i)} \leq (i/N)\, \alpha_{FDR} \right], \tag{13}$$

where subscripts denote the indices of the bin p-values sorted in ascending order, N is the number of bins, and α_{FDR} controls the false-discovery rate (ie. the rate at which the local null hypothesis will be incorrectly rejected). The reader is referred to Wilks (2016) for further details of the method.

Regions where the null hypothesis cannot be rejected at $\alpha_{FDR} = 0.15$ are shown in figure S8 by the stippling. This value of the α_{FDR} is relatively high, but was found to give the best balance between hypothesis testing and retaining sufficient spatial resolution to

capture the features of interest (we note as well that the definition of significance in (13) is more stringent than applying individual significance calculations at each bin, such that almost all regions where the null hypothesis is rejected also have $p < 0.05$). The major features of the 20th century temperature change that are discussed in the text are in regions where the null hypothesis is rejected. This includes the strong warming near the surface along the Antarctic Circumpolar Current, moderate warming extending through the subtropical gyre interior, and cooling and warming on the poleward and equatorward flank of the thermocline dome, respectively.

We also note that for this same bin-averaging and value of α_{FDR}, the global null hypothesis (that the null hypothesis is true for all bins) cannot be rejected for the temperature differences calculated between the 1955-1964 climatology and the historical observations.

References

Gebbie, G., & Huybers, P. (2010, August). Total Matrix Intercomparison: A Method for Determining the Geometry of Water-Mass Pathways. *Journal of Physical Oceanography*, *40*(8), 1710–1728. Retrieved 2022-02-02, from http://journals.ametsoc.org/doi/10.1175/2010JPO4272.1 doi: 10.1175/2010JPO4272.1

Gebbie, G., & Huybers, P. (2019, January). The Little Ice Age and 20th-century deep Pacific cooling. *Science*, *363*(6422), 70–74. Retrieved 2021-07-12, from https://www.sciencemag.org/lookup/doi/10.1126/science.aar8413 doi: 10.1126/science.aar8413

Gergis, J. L., & Fowler, A. M. (2009, February). A history of ENSO events since A.D. 1525: implications for future climate change. *Climatic Change*, *92*(3-4), 343–

387. Retrieved 2021-07-17, from http://link.springer.com/10.1007/s10584-008-9476-z doi: 10.1007/s10584-008-9476-z

Gouretski, V., & Koltermann, K. (2004). *WOCE Global Hydrographic Climatology* (Tech. Rep. No. 35). Berichte des Bundesamtes für Seeschifffahrt und Hydrographie.

Huang, R. X. (2015, December). Heaving modes in the world oceans. *Climate Dynamics*, *45*(11-12), 3563–3591. Retrieved 2022-01-25, from http://link.springer.com/10.1007/s00382-015-2557-6 doi: 10.1007/s00382-015-2557-6

Roemmich, D., John Gould, W., & Gilson, J. (2012, June). 135 years of global ocean warming between the Challenger expedition and the Argo Programme. *Nature Climate Change*, *2*(6), 425–428. Retrieved 2021-07-12, from http://www.nature.com/articles/nclimate1461 doi: 10.1038/nclimate1461

Saji, N., & Yamagata, T. (2003). Possible impacts of Indian Ocean Dipole mode events on global climate. *Climate Research*, *25*, 151–169. Retrieved 2021-07-16, from http://www.int-res.com/abstracts/cr/v25/n2/p151-169/ doi: 10.3354/cr025151

Saji, N. H., Goswami, B. N., Vinayachandran, P. N., & Yamagata, T. (1999, September). A dipole mode in the tropical Indian Ocean. *Nature*, *401*(6751), 360–363. Retrieved 2021-07-16, from http://www.nature.com/articles/43854 doi: 10.1038/43854

Wilks, D. S. (2016, December). "The Stippling Shows Statistically Significant Grid Points": How Research Results are Routinely Overstated and Overinterpreted, and What to Do about It. *Bulletin of the American Meteorological Society*, *97*(12), 2263–2273. Retrieved 2021-07-22, from https://journals.ametsoc.org/doi/10.1175/BAMS-D-15-00267.1 doi: 10.1175/BAMS-D-15-00267.1

Wortham, C., & Wunsch, C. (2014, March). A Multidimensional Spectral Description of Ocean Variability. *Journal of Physical Oceanography*, *44*(3), 944–966. Retrieved 2022-02-02, from http://journals.ametsoc.org/doi/10.1175/JPO-D-13-0113.1 doi: 10.1175/JPO-D-13-0113.1

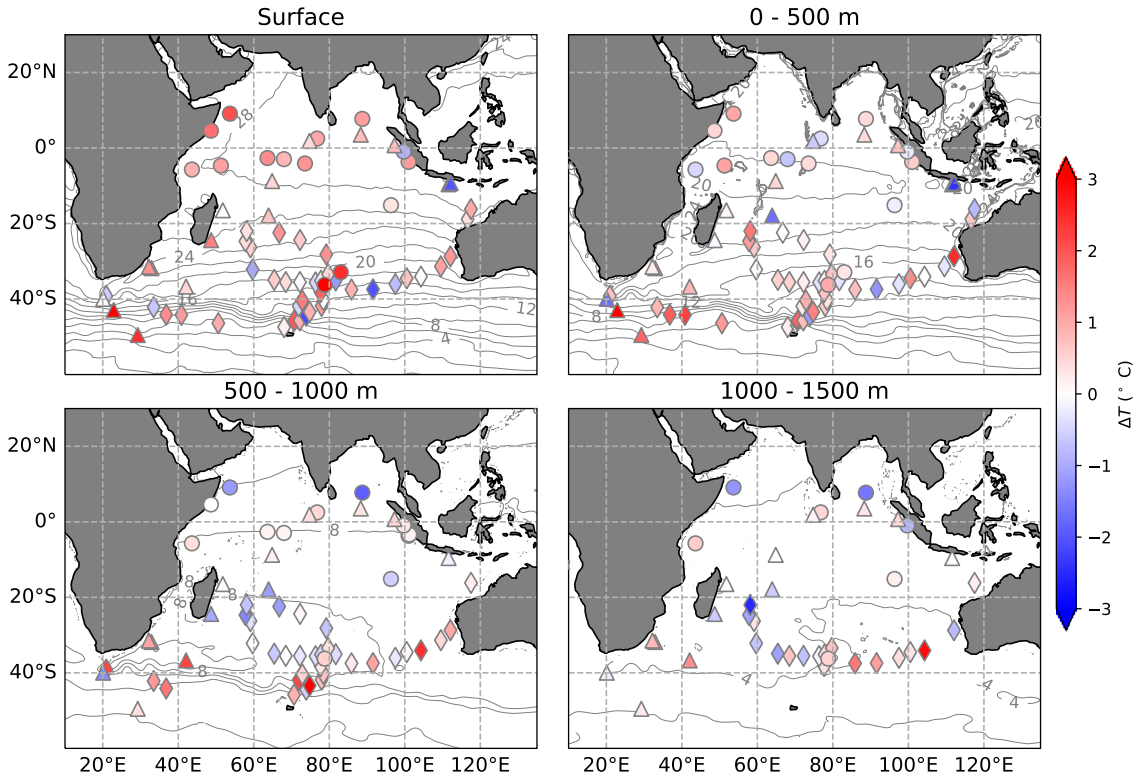

Figure S1. Depth averaged temperature change between 2005–2017 and the observations from the *Gazelle* (diamond markers), *Valdivia* (circle markers), and *Planet* (triangle markers). Depth ranges of averaging are indicated in the title of each subpanel. Modern annual average temperature values over the same depth ranges are also shown (thin contours) with a contour interval of 2°C.

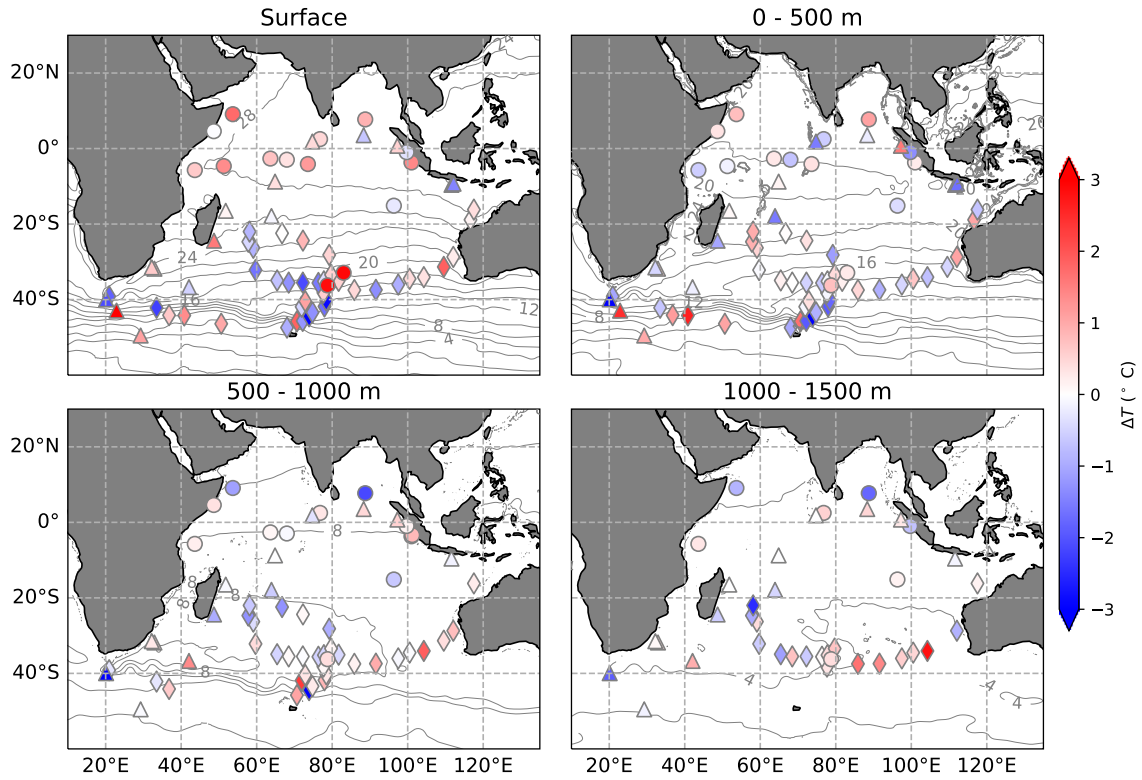

Figure S2. As in figure S1, but for the temperature differences between the 1955–1964 climatology and the historical observations.

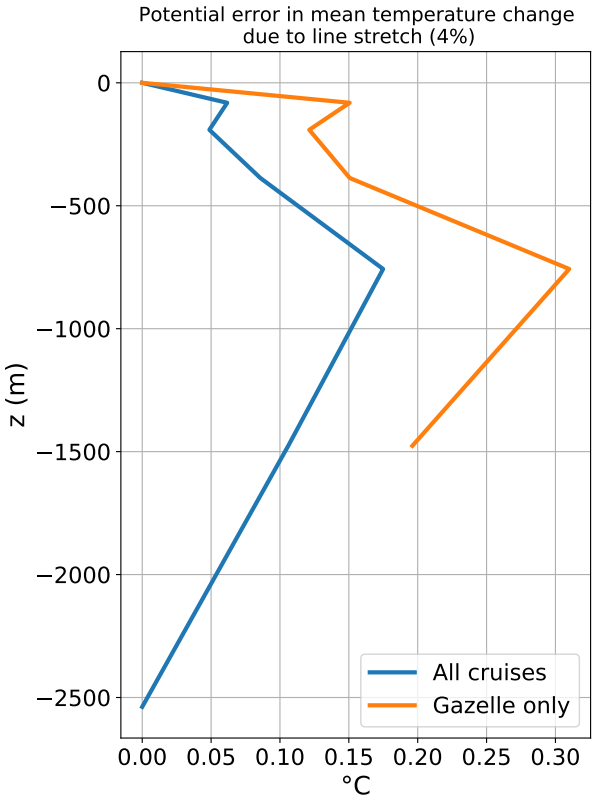

Figure S3. Depth profiles of the mean temperature bias introduced across the historical station locations by an assumed 4% shallow bias in the reported observation depths of the *Gazelle*. Positive values imply estimates of modern minus historic data are biased warm.

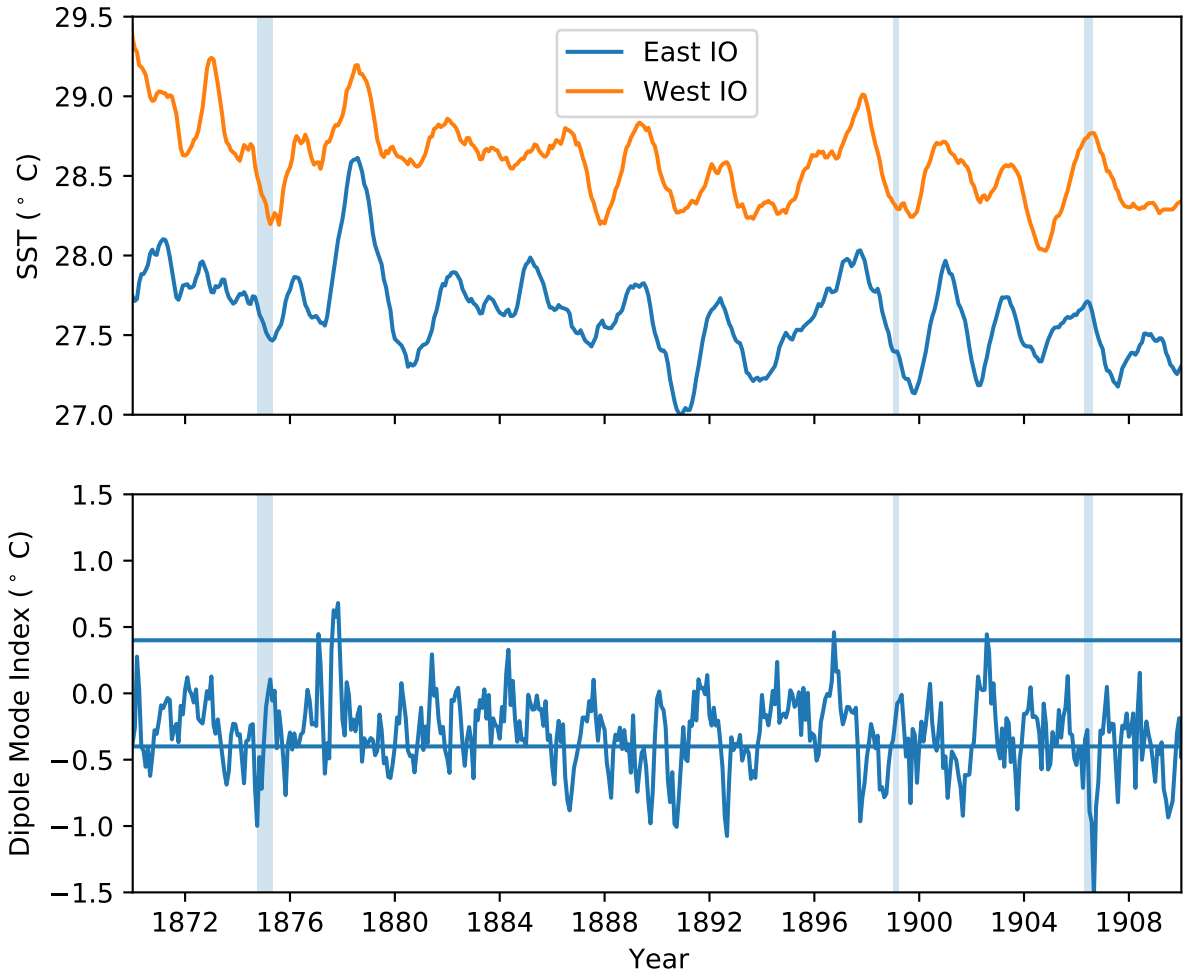

Figure S4. Top: Sea-surface temperature from ERSST averaged over the West (50°E – 70°E, 10°S – 10°N) and East (90°E – 110°E, 10°S – 0°) Indian Ocean. Bottom: The Dipole Mode Index as defined in N. Saji and Yamagata (2003, https://psl.noaa.gov/gcos_wgsp/Timeseries/DMI). In both plots the time-period of the historical cruise observations are indicated by the blue shading.

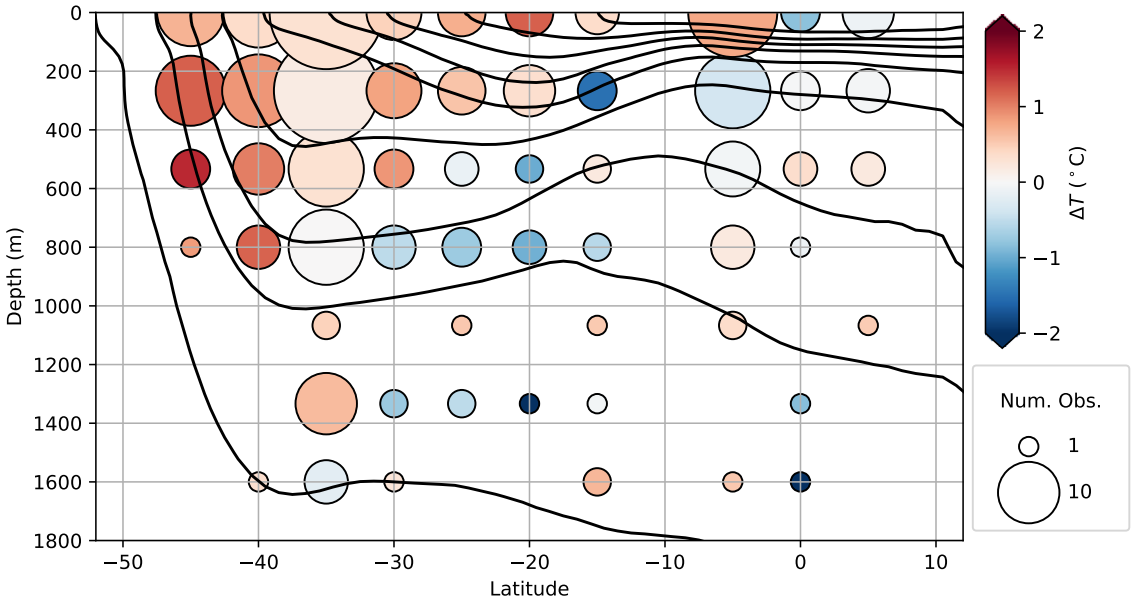

Figure S5. Latitude-depth slice of modern minus historical temperatures, as in figure 3, but without the *Planet* observations.

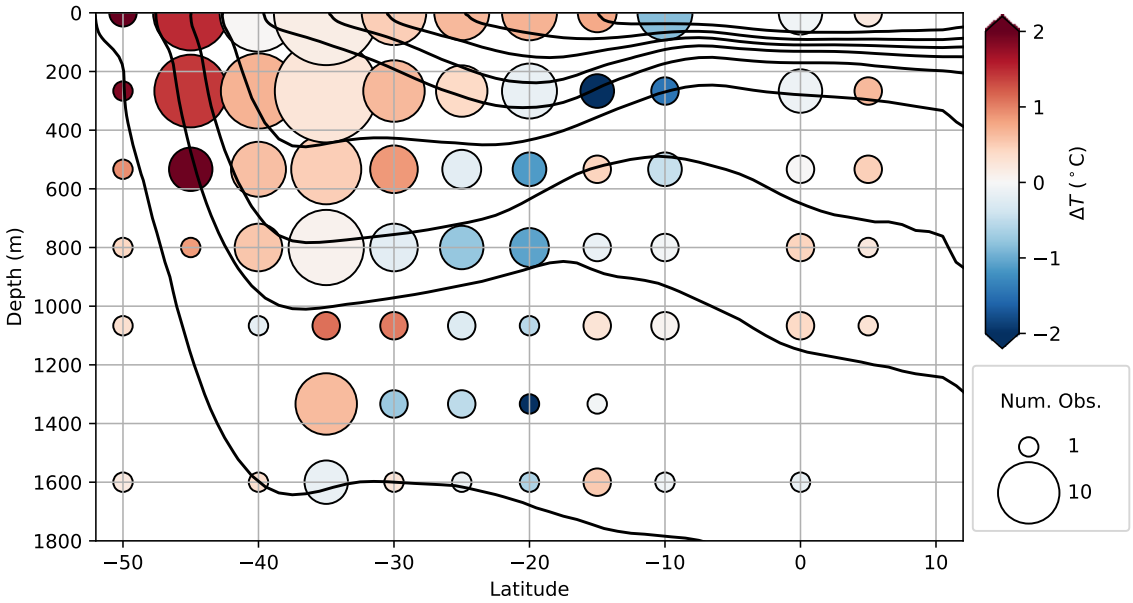

Figure S6. Latitude-depth slice of modern minus historical temperatures, as in figure 3, but without the *Valdivia* observations.

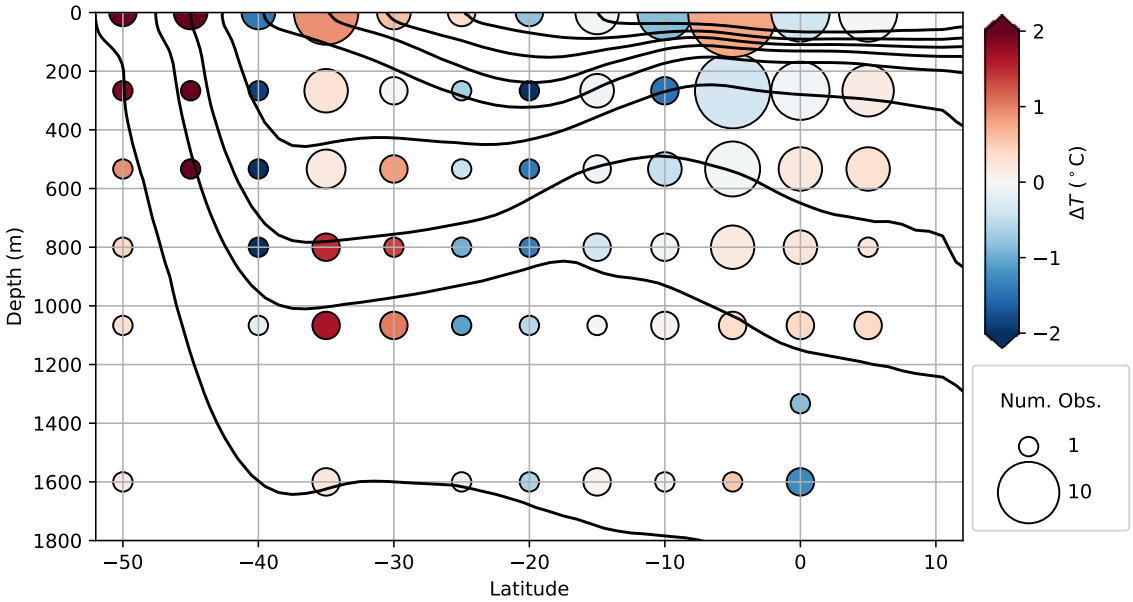

Figure S7. Latitude-depth slice of modern minus historical temperatures, as in figure 3, but without the *Gazelle* observations.

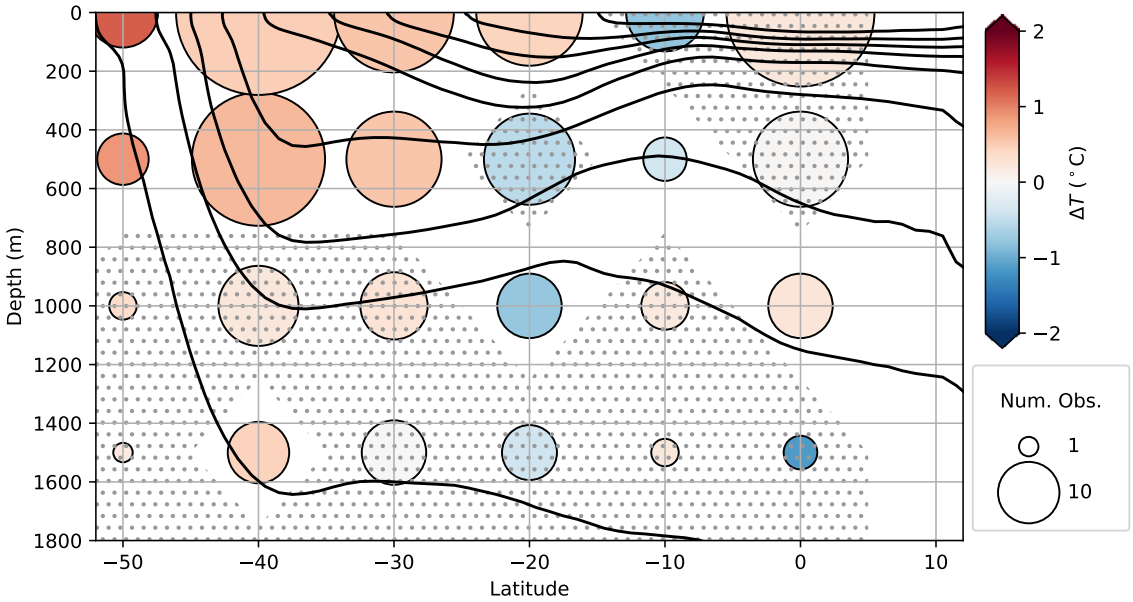

Figure S8. Latitude-depth slice of modern minus historical temperatures, as in figure 3, but averaged over larger bins. In this plot regions of stippling indicate areas where the null hypothesis of 0 mean temperature change cannot be rejected at the $\alpha_{FDR} = 0.15$ level.