On the Time Scale Structure of Climate Variability and Response

On the Time Scale Structure of Climate Variability and Response

A DISSERTATION PRESENTED
BY
CRISTIAN PROISTOSESCU
TO
THE DEPARTMENT OF EARTH AND PLANETARY SCIENCES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN THE SUBJECT OF
EARTH AND PLANETARY SCIENCES

HARVARD UNIVERSITY
CAMBRIDGE, MASSACHUSETTS
DECEMBER 2016

Thesis advisor: Professor Peter Huybers Cristian Proistosescu

On the Time Scale Structure of Climate Variability and Response

Abstract

This thesis analyzes several aspects pertaining to the time scale structure of climate variability and the temperature response to external radiative forcing. The analysis begins with an examination of the structure of temperature variability on synoptic time scales. Non-normal characteristics of temperature distributions have important implications for extreme climatic events. Mechanisms giving rise to the observed non-normality have been posited across a range of time scales. The structure of non-normality is examined in radiosonde time series of winter temperature using linear filters. These linear filters are shown to suppress non-normal variability. Observed non-normality on longer time is shown to likely be introduced at the highest resolved frequency and propagated through autocorrelation. Conversely, observations of normal distributions on synoptic time scales are shown to be an artifact of filtering. In the process, a Monte-Carlo based test for non-normality is developed for use in time series presenting autocorrelation.

On longer time scales the single most important climate quantity is, arguably, a measure of the long-term warming that results from increased greenhouse gas concentrations, termed Equilibrium Climate Sensitivity or ECS. Estimates of ECS are drawn from numerical simulations, historical data, and paleoclimate proxies. The historical estimates have a range of 1.5–3°C compared to the 2–4.5°C range drawn from simulations and proxies. These historical estimates, however, are suspected to be biased low, as they rely on an assumption of a single radiative response to warming across

all timescales. This bias is quantified using a Bayesian methodology to parse fast and slow modes in the evolution of Earth's temperature and radiation within an ensemble of 24 climate GCMs. Centennial-scale modes with stronger amplifying feedbacks ultimately contribute 44% of the long-term warming, but account for only 3% of current warming. Thus, although the GCM ensemble has a median equilibrium sensitivity of 3.4°C, historical forcing would yield a biased estimate of only 2.5°C, consistent with the observational range.

The equilibrium response is unlikely to be realized in upcoming centuries, as the radiative forcing is expected to vary on the same time scales. The temporal structure of atmospheric CO_2 following present and future carbon emissions is simulated using a response function consisting of three eigenmodes fitted to 16 models of the carbon cycle. Convolving these response functions with those of the CMIP5 models provides a comprehensive probabilistic distribution for the evolution of temperature in response to any given emission scenario. The framework allows for calculation of emission budget associated with a given peak temperature threshold and a given probability of exceeding that threshold. As expected, the uncertainty range is sensitive to the fast decadal modes of the physical system and the carbon cycle response for narrow emission profiles. However, for more realistic broad emission profiles the uncertainty range is most sensitive to the slow centennial modes of the physical system.

Due to the low degree of representation in modern observations, constraining the evolution of slow climatic modes will likely require use of paleoclimate proxies. The dominant modes of variability over the Pleistocene is associated with periodic changes in the earth's orbital configuration termed Milankovitch cycles. Such cycles have been described in pre-Pleistocene sediment records spanning warmer climates that may provide better analogues for future warming. Most of these studies, however, employ cli-

matic records set on orbitally tuned chronologies. The bias introduced in spectral power estimates from the background continuum in the presence of tuning is quantified using Monte-Carlo methods, and appropriate hypothesis test for orbital forcing is developed. The test is applied to two marine sediment δ^{18}O records spanning the Oligo-Miocene, from ODP cores 1090 and 1218. Orbital tuning is found to increase the statistical significance of a precession peak, whereas the obliquity and eccentricity peaks are no longer significant when compared to a null hypothesis of tuned background noise.

Contents

Dedicated to my parents, for their unwavering support.

Acknowledgments

Thanks go to my family, my friends. I would also like to thank the department of Earth Planetary Sciences and the Graduate School of Arts and Sciences for their support. This work would not have been possible without insightful conversations with numerous other scientists, particularly present and past members of the Huybers group. Last but not least thanks go to Peter Huybers for the mentorship provided throughout the years.

1

Identification and interpretation of non-normality in atmospheric time series

Abstract

Non-normal characteristics of geophysical time series are important determinants of extreme events and may provide insight into the underlying dynamics of a system. The structure of non-normality in winter temperature is examined through the use of linear filtering of radiosonde temperature time series. Filtering either low or high frequencies generally suppresses what is otherwise statistically significant non-normal variability in temperature. The structure of non-normality is partly attributable to geometric relations between filtering and the appearance of skewness, kurtosis, and higher order moments in time series data, and partly attributable to the presence of non-normal temperature variations at the highest resolved frequencies in the presence of atmospheric memory. A non-normal autoregressive model and a multiplicative noise model are both consistent with the observed frequency structure of non-normality. These results suggest that the generating mechanism for non-normal variations does not necessarily act at the frequencies at which greatest non-normality is observed.

1.1 Introduction

Departures from normality in temperature have important implications for the frequency of extreme events (Ruff & Neelin, 2012), and discerning the statistical characteristics of non-normality can give insight into relevant physical processes (Sura & Perron, 2010). A wide variety of mechanisms have been proposed to explain the observed non-normality (Sura & Hannachi, 2015). One class of mechanisms posits that non-normality arises on synoptic time-scales, either from the effects of state-dependent or multiplicative noise (e.g. Sura et al., 2005; Sardeshmukh & Sura, 2009; Sura & Perron, 2010), or from

additive noise with non-normal perturbations (e.g. Luxford & Woollings, 2012). The second class of mechanisms ascribes non-normality to low frequency variability, such as non-normal distributions arising as mixtures of gaussian processes, introduced through nonlinear regime shifts or non-stationarity (e.g. Hannachi, 2010).

Sardeshmukh & Sura (2009) note that the relationship of skewness and kurtosis in observations of midlatitude variability is consistent with multiplicative noise acting on synoptic time scales. Schneider et al. (2015), however, have shown that midlatitude temperature records filtered to the canonical synoptic time scales appears normal within 3 standard deviations, whereas deviations from normality are apparent when including variability at time scales longer than 15 days. Similarly, Rennert & Wallace (2009) use geopotential height data filtered to three different frequency bands to show that skewness in atmospheric variability is related to cross-frequency coupling between intermediate and low-frequency time scales.

In seeking to distinguish between these seemingly conflicting interpretations it is useful to distinguish structure indicative of physical processes from structure related to the generic effects of filtering time series. Rosenblatt (1961) noted that "[i]t appears to be part of the engineering folklore that a narrow band-pass filter applied to a stationary random input yields an output that is approximately normally distributed." The tendency of filtering to alter the perceived normality of the data has been noted for other signals (e.g., Rozanov, 1961; Mallows, 1967; Papoulis, 1972). Donohoe & Battisti (2009) describe how the asymmetry in the distribution of synoptic cyclones and anticyclones depends on the choice of temporal or spatial filtering techniques, and a similar tendency towards normality has been noted for spatial filtering in the form of gridding (Director & Bornn, 2015; Cavanaugh & Shen, 2015).

An overall tendency for filtering to make time series data appear more normal can be

explained from a Fourier perspective, whereby filtering nullifies interactions between frequencies that are necessary for representing non-normal structure (Kotulski & Sobczyk, 1981; Garth & Bresler, 1997). There remains the possibility, however, that distinct structure in the non-normality of filtered records can provide physical insight. In the following we examine radiosonde time series, first demonstrating statistically significant non-normality, then analytically describing how this non-normal structure can be expected to decay upon filtering, and finally describing more nuanced structures associated with autocorrelation using two simple numerical models. We describe non-normality using the the third and forth moments of the distribution, skewness and kurtosis. Intuitively, skewness reflects the symmetry of the distribution with positive values indicating greater weight in the upper tail, while kurtosis reflects overall weight placed in the tails relative to the center of the distribution.

1.2 Radiosonde Temperature Data

As an initial example, Fig. 1.1 shows the distribution of wintertime radiosonde temperature from the Barrow, Alaska station. Data are obtained from the Integrated Global Radiosonde Archive (Durre et al., 2006) and are subset to only 12:00 GMT soundings at 850 millibars. The 1958-2009 interval is selected for analysis because only 2% of samples are missing, and these are infilled by linear interpolation. We remove the annual cycle and its first two harmonics by subtracting least squares-fit sine waves, though results are similar when the annual cycle is retained. The sample distribution of the Barrow data involves a positive skew and is significantly non-normal ($p < 0.01$). Significance is assessed using a Kolmogorov-Smirnov (KS) test (Smirnov, 1939), where the largest deviation between the sample cumulative distribution and that of a normal distribu-

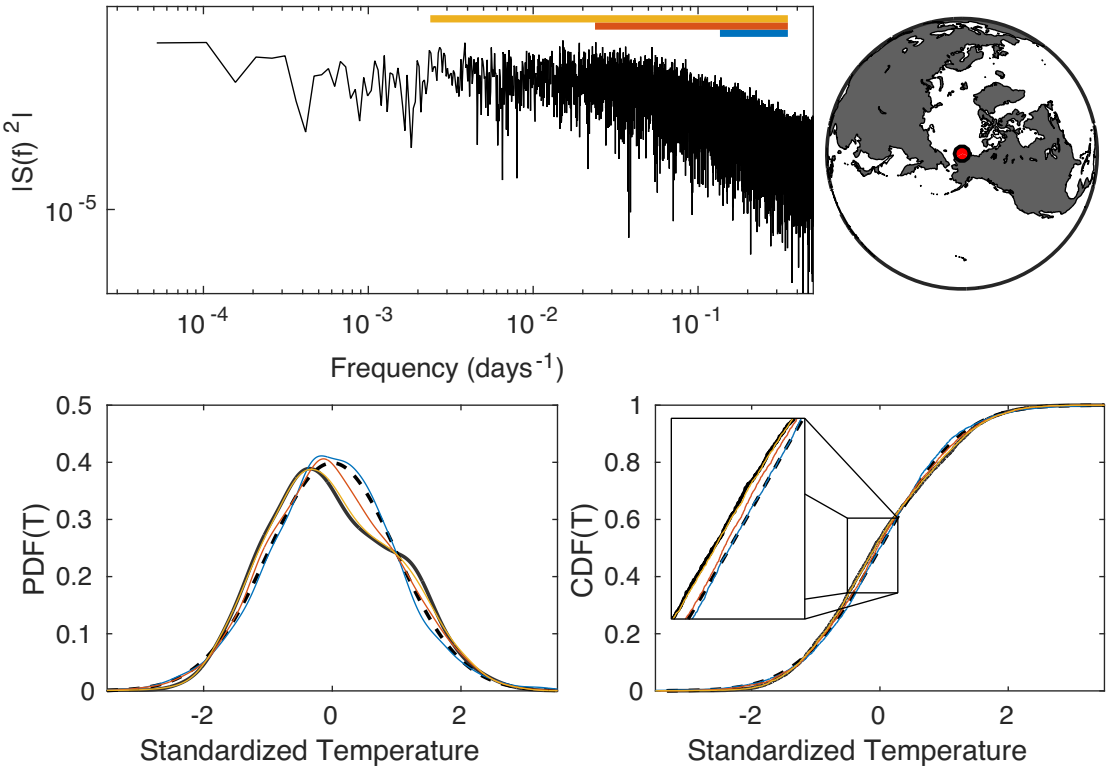

Figure 1.1: Filtering of radiosonde temperature time series and normality. **Top:** Power spectrum of a radiosonde time series of temperature at 850 mbar from Barrow, AK. The annual cycle is removed by notch-filtering the annual frequency and its first two harmonics. The time series is then progressively filtered using a top-hat band-pass filter. The admitted frequency band for each filter is depicted by color bars indicating frequencies of 1/3-1/7 days $^{-1}$ (blue), 1/3-1/40 days^{-1} (orange), and 1/3-1/400 days^{-1} (yellow). **Bottom:** The probability density function (PDF) and cumulative density function (CDF) are depicted for winter (DJF) temperatures for the full data (solid black line), the filtered time series (solid colored lines),and a standardized normal distribution (dashed line).

tion is evaluated. In order to account for autocorrelation in the time series, the null distribution is built from normally distributed surrogate data generated by phase randomizing the original time series (See Supplemental Material). Rejection of normality can appear either from features of sub-seasonal variability or interannual variability. For instance, normal variability that is subject to an interannual trend in the mean would give the appearance of non-normality (Huybers et al., 2014). Interseasonal variability,

such as increased variance in winter can also lead to increased values of kurtosis (Frisch & Sornette, 1997). Interannual contributions and interseasonal variability is suppressed in our analysis through only considering December through February winter months (DJF) and normalizing each seasonal realization to zero mean and unit variance.

One indication of the relevance of understanding the origins of deviations from normality are that they reach 0.8 and 0.3 °C for the 5th and 95th percentiles respectively and 2.3 and -0.5 °C for the 1st and 99th percentiles (See also Table 1.1 in the Supplemental Material). Such deviations can lead to considerable changes in the rates of exceedance of temperature thresholds, particularly in association with changes in mean climate (Ruff & Neelin, 2012). Non-normality in the Barrow data is also consistent with a more comprehensive study (Perron & Sura, 2013) that found statistically significant deviations from non-normality in daily reanalyses data for nine atmospheric variables, including temperature. Surface temperature from the U.S. Global Historical Climatology Network (GHCN) has also been shown to exhibit significant non-normality, as well as trends in higher order moments and quantiles (Huybers et al., 2014; Cavanaugh & Shen, 2014).

The Barrow radiosonde data becomes more consistent with a normal distribution upon filtering. For example, filtering to only retain synoptic scale anomalies at 1/3-1/15 day^{-1} frequencies alters the Barrow results to be substantially more consistent with a normal distribution ($p = 0.035$, $p = 0.15$ using a standard KS test). Filtering is performed using a simple top hat filter whereby Fourier components outside the pass-band are set to zero. This approach is not optimal from the perspective of suppressing Gibbs phenomena but is amenable to later analytic calculations. Furthermore, no appreciable differences are obtained when using more sophisticated filters, such as a forward-backward pass with a Butterworth filter.

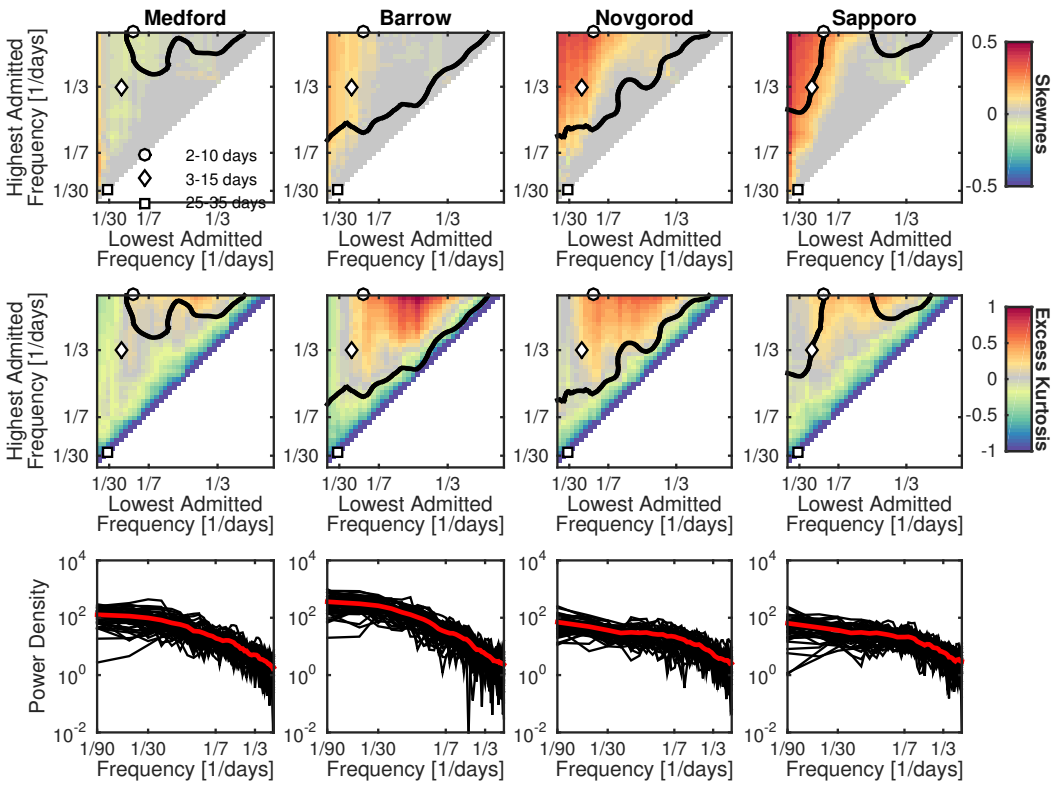

Figure 1.2: Effects of filtering on skewness (**Top**) and excess kurtosis (**Middle**) of 850 mbar DJF temperature from radiosonde stations for the full range of pass-bands. Axes denote pass band limits. Each season is treated as an independent ensemble member, standardized to zero mean and unit variance. Values of sample S and K averaged are over all seasons. Lines parallel to the diagonal have equal bandwidth for the pass-band, with the width of the pass-band increasing further away from the diagonal. Points below the solid black lines cannot be distinguished from samples drawn from a normal distribution under a Kolmogorov Smirnov (KS) test, at 95% confidence level. Markers denote choices of filter pass-band discussed in the text. **Bottom**: spectral estimates of each individual season (black) and the mean spectrum (red). Note the logarithmic frequency axis for spectral estimates.

Table 1.1: Difference in 5th, 95th, 1st, and 99th percentiles of anomalies from the mean in wintertime data from those of a normal distribution with identical mean and variance

	Medford	Barrow	Sapporo	Tateno
T_5^N	-7.9	-11.8	-6.5	-6.4
ΔT_5	0.4	0.8	0.6	0.8
T_{95}^N	7.9	11.8	6.5	6.4
ΔT_{95}	0.3	0.3	0.6	0.8
T_1^N	-11.1	-16.8	-9.2	-9.1
ΔT_1	1.1	2.3	1.0	1.9
T_{99}^N	11.1	16.8	9.2	9.1
ΔT_{99}	-0.5	-0.5	0.7	1.5

The example in Fig. 1.1 can be generalized (Fig. 1.2) to show how non-normal components vary according to filtering. In addition to Barrow (USA, 1958-2009), we examine how skewness and excess kurtosis of three other long and nearly complete records from Medford (USA, 1958-2009), Novgorod (Russia, 1966-2009), and Sapporo (Japan, 1965-2009) change as a function of high and low cut-off frequencies specified for filtering. Data are obtained and processed in a manner analogous to the Barrow example. A general pattern holds wherein skewness and kurtosis diminish with increasingly narrow filter bandwidths, but more nuanced structure is also present. Positive skewness appears when the lowest frequencies of 1/90 days^{-1} are admitted but which diminishes to near zero once low-frequency cut offs of 1/7 days^{-1} or higher are specified. Relative to the low-frequency cut off, changing the high-frequency cut off has only a weak influence on skewness, except for the presence of skewness extending out toward a frequency band

near 1/3-1/4 days^{-1}. In contrast to skewness, excess kurtosis is negative when admitting the lowest frequencies and is maximized when passing frequencies between 1/5 to 1/2 days^{-1}, roughly encompassing synoptic timescales. Excess kurtosis also becomes strongly negative when the pass band becomes extremely narrow.

Significance of non-normality in the radiosonde time series is evaluated for each filter combination using the previously described KS test. All records show statistically significant non-normality in those regions having the greatest magnitudes of skewness and excess kurtosis. Higher order moments are also present but decay more quickly under filtering and make only minor contributions to overall non-normality.

1.3 Why Filtering Tends to Generate Normality

The tendency of a time series to become more normally distributed after filtering can be understood through the use of higher order spectra (Brillinger, 1965; Garth & Bresler, 1997). The power spectrum, bispectrum and trispectrum of a process $x(t)$ can be defined in terms of the Fourier Transform $\hat{x}(f)$ as,

$$P(f_1) = \hat{x}(f_1) \cdot \hat{x}^*(f_1), \tag{1.1a}$$

$$B(f_1, f_2) = \hat{x}(f_1) \cdot \hat{x}(f_2) \cdot \hat{x}^*(f_1 + f_2), \tag{1.1b}$$

$$T(f_1, f_2, f_3) = \hat{x}(f_1) \cdot \hat{x}(f_2) \cdot \hat{x}(f_3) \cdot \hat{x}^*(f_1 + f_2 + f_3). \tag{1.1c}$$

Higher order spectra follow a relation similar to Parseval's theorem relating the power spectrum and variance, wherein the surface integral over the bispectrum is related to skewness, S, and the volume integral over the trispectrum is related to excess kurtosis,

K,

$$S = \frac{\int B(f_1, f_2) \cdot df_1 \cdot df_2}{\left(\int P(f_1) \cdot df_1\right)^{3/2}}, \tag{1.2a}$$

$$K = \frac{\int T(f_1, f_2, f_3) \cdot df_1 \cdot df_2 \cdot df_3}{\left(\int P(f_1) \cdot df_1\right)^{4/2}}. \tag{1.2b}$$

Deviations from a normal distribution are encoded in interactions between different frequency bands in the Fourier representation of time series and are modified by filtering. The specific effects of filtering $x(t)$ upon skewness and kurtosis can be obtained by replacing $\hat{x}(f)$ with a filtered version, $\hat{y}(f) = \hat{h}(f) \cdot \hat{x}(f)$, in Eqs. 1.1–1.2. Excluding a frequency f_i eliminates all interacting frequencies pairs, such that the unfiltered area of the bispectrum generally diminishes more rapidly than that of the power spectrum. For example, a low-pass filter that retains only a fraction β of unfiltered frequencies will lead to a bispectrum with only $3\beta^2/4$ of its area being unfiltered for $\beta < 2/3$. A geometric depiction of the effects of band-pass filtering upon the bispectrum is given in Fig. 1.3, and a similar depiction for the trispectrum is given in Fig. 1.4. Although the patterns are visually simple, their analytical computation involves integrals over higher order spectral volumes, which can become rather involved (Büeler et al., 2000). Analytical solution for bispectrum filtering are given in the supplemental material.

Fig. 1.5 shows the fraction of skewness and excess kurtosis admitted as a function of filtered cut-off frequencies. There is excellent agreement between these analytically computed values and that obtained from filtering a random time series containing 10^7 realizations from a Pearson distribution. The Pearson distribution is chosen because it is specifically derived as a class of models used to fit a non-normal distribution based on

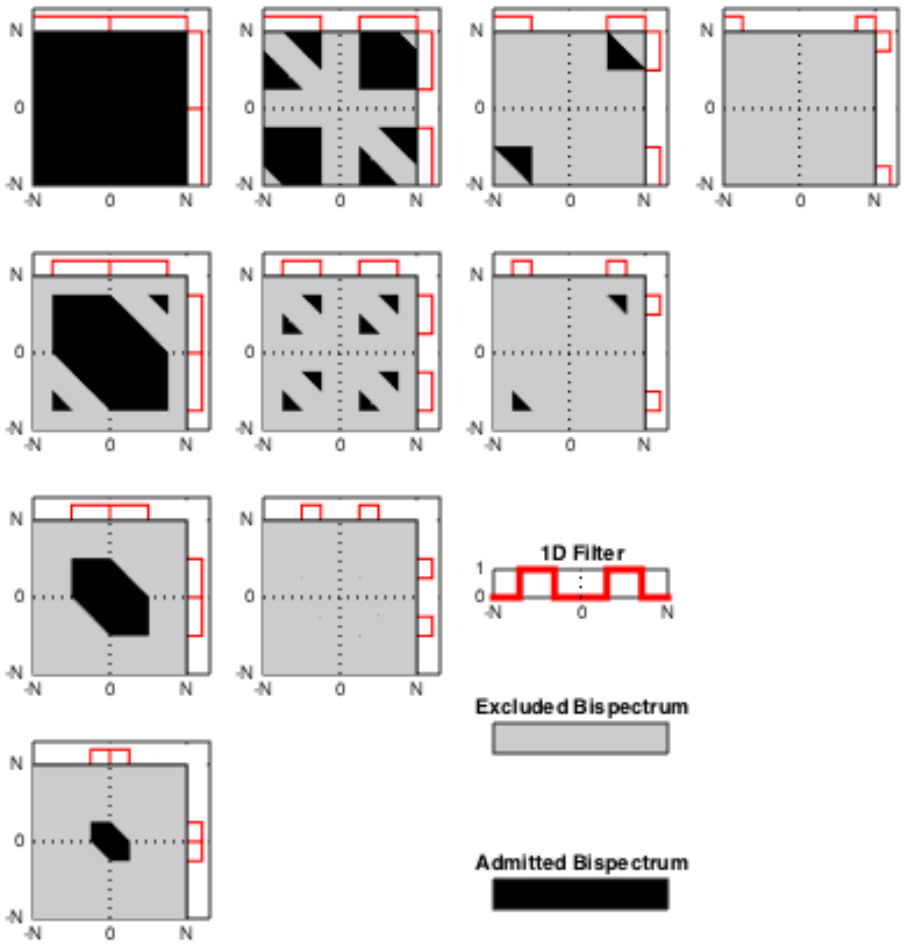

Figure 1.3: Effects of filtering on the bispectrum. Frequency axes are normalized relative to the Nyquist frequency (N), and the solid color depicts the part of the bispectrum respectively that is allowed to pass under a band-pass filter. Rows from top to bottom indicate greater filtering of high frequencies, and columns from left to right are greater filtering of low frequencies. Note that both negative and positive frequencies contribute to the bispectrum.

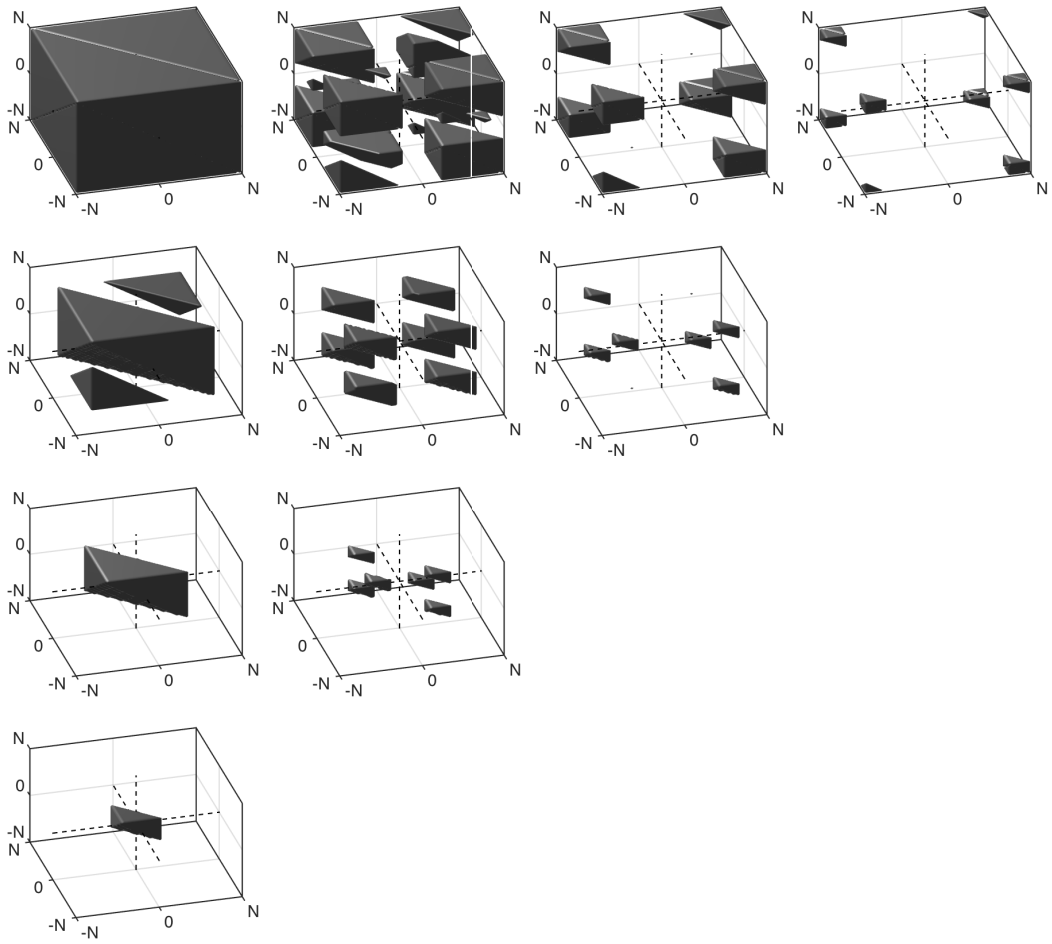

Figure 1.4: Same as Fig. 1.3, but depicting the effects of a top-hat band-pass filter on the trispectrum.

skewness and kurtosis, and because it provides a good fit to the radiosonde record. The results remain unchanged,however, if the simple models are forced with beta, gamma or chi-squared distributions. Similarly, the results hold for the case of Correlated Additive and Multiplicative noise, where no distribution is explicitly prescribed. When the time series is reduced to 10^5 points, the match becomes more noisy, with the implication that the finite radiosonde time series we analyze can be expected to have variable skewness and kurtosis structure even if the underlying statistical distributions are identical.

The analytical results describe the main features of the skewness observed in the filtered radiosonde data (Fig. 1.5a). Filtering low frequencies more rapidly decreases skewness than filtering high frequencies because, as depicted in Fig. 1.3, the sum of low-frequency pairs excludes more area of the bispectrum not already excluded by filtering a single member of the pair. The analytical results also explain a lobe of high skewness that extend out towards frequencies in the vicinity of half the Nyquist frequency, or in the case of the daily sampled radiosonde time series, 1/4 days^{-1}. This lobe is again a consequence of asymmetries in how filtering certain frequencies also excludes energy at interaction frequencies (see Eq. 1.1).

In contrast with skewness, correspondence between the kurtosis obtained from our analytical results and the radiosonde observations is poor. The Fourier representation shows contours of kurtosis that align with contours of constant filter bandwidth (Fig 1.5), whereas the data show maximum kurtosis when filtering everything but frequencies corresponding to synoptic timescales. Negative values of excess kurtosis observed when filtering all but a small number of frequencies (the diagonals in Fig. 1.2) reflects the fact that the excess kurtosis of a sine wave is -1.5. For the case of a normal process, wherein the Fourier coefficients are independently distributed, it can be shown that the expected sample kurtosis converges to the process excess Kurtosis of zero as

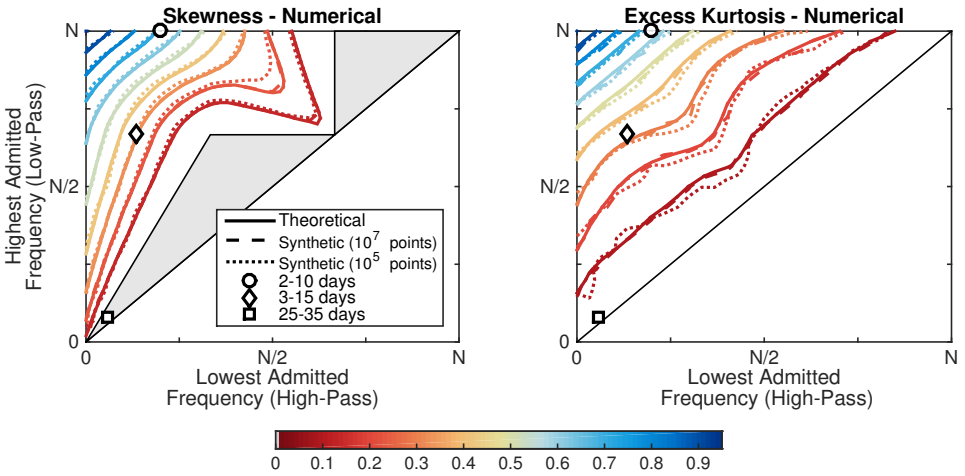

Figure 1.5: Skewness and excess kurtosis under band-pass filtering, relative to unfiltered values. Axes denote pass band limits, relative to the Nyquist frequency. Gray shaded area denotes values of skewness of exactly zero. **Left**: Numerical computation of a filtered isotropic bispectrum, and synthetic estimations based on independent non-normal noise. **Right**: Same as middle, but for excess kurtosis. Details of the numerical estimations are given in the appendix.

$K \propto -3/2n$, where n is the number of consecutive frequencies retained in the Fourier spectrum (see Appendix). Analytical results do not display this sinusoidal limit on account of continuous frequency resolution. The remainder of the mismatch between the analytical and observed kurtosis, however, points to substantive differences between filtering independent realizations from a Pearson distribution and actual temperature variability.

1.4 Numerical Models

An obvious deficiency in the foregoing analytical results is the presumption of independent realizations, whereas spectral estimates of radiosonde temperature have increasing energy toward lower frequencies (Figs. 1.1,1.2), indicative of memory between subsequent values. We represent correlations in atmospheric data using an autoregressive

process of order one, meant to represent the generally more energetic variability found at lower frequencies, either because of preferential damping of high frequency variations or the presence of memory in the record, e.g. due to the long timescales associated with ocean thermal inertia (Hasselmann, 1976),

$$x(t+1) = \rho x(t) + \eta(t), \tag{1.3}$$

where η represents independent realizations from a Pearson distribution. Specifying $\rho = 0$ gives the same skewness and kurtosis structure discussed previously (Fig. 1.5), at least up to variations associated with finite samples. Setting $\rho = 0.9$, however, gives results in agreement with observations (Fig. 1.6), especially with respect to maximum kurtosis at synoptic periods.

We also consider a a discretized version of the model presented by Sardeshmukh & Sura (2009) that involves correlated additive and multiplicative (CAM) noise. This is equivalent to an AR(1) process forced with non-normal noise, wherein the non-normality arises from state-dependency of perturbations. Such state-dependency has been modeled as arising from stochastic damping of Rossby waves (Sura et al., 2005) or stochastic advection of potential vorticity anomalies (Sura & Perron, 2010). The discretized model can be written as:

$$x(t+1) = \rho x(t) + b\eta_1(t) + (Ex(t) + g)\eta_2(t) - \frac{1}{2}Eg, \tag{1.4}$$

with η_1 and η_2 representing independent realizations of a standard normal distribution. The first term represents the autocorrelated nature of the process, the second and third term respectively represent the normal and non-normal innovations, and the last term ensures the process is stationary. Non-normality is introduced by a state-dependent and

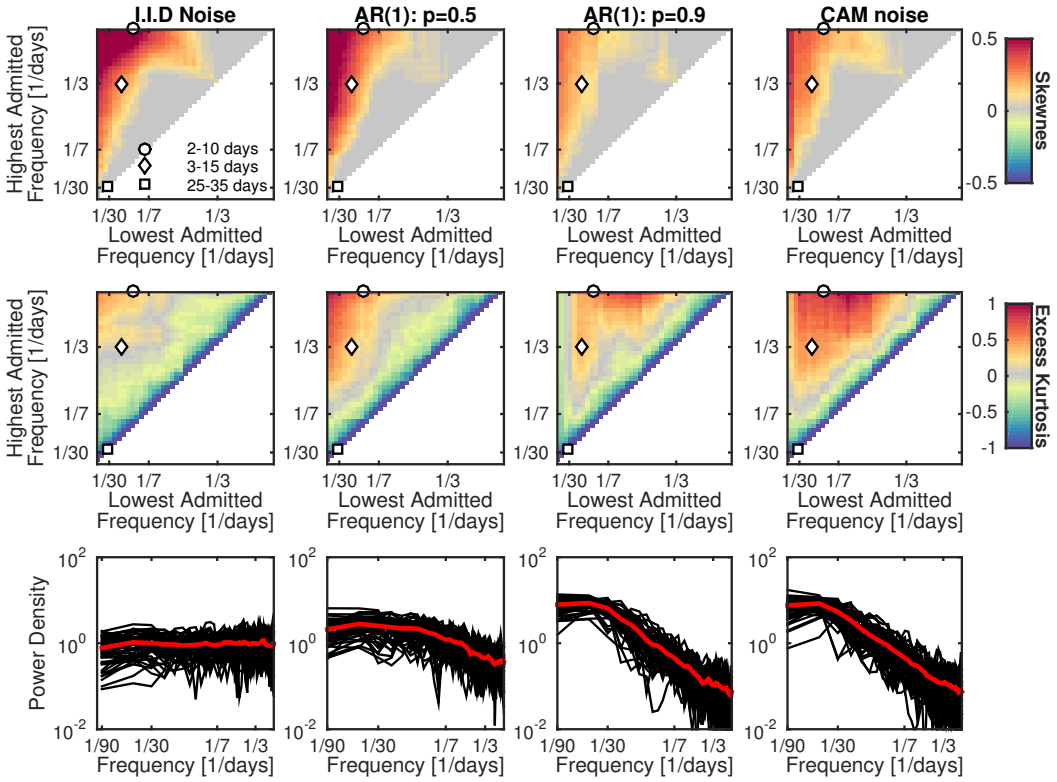

Figure 1.6: Same as Fig. 1.2, but for synthetic data. From left to right, columns present results models that are i.i.d., autoregressive order one with parameters of $\rho = 0.5$ and $\rho = 0.9$ (Eq. 1.3), and correlated additive and cumulative noise (Eq. 1.4). The i.i.d. and autoregressive models are driving by non-normal realization from a Pearson distribution. For each model, 50 independent ensemble members of length 90 are generated, with each member representing a season. Plotted results are the average after analyzing each season, except for the bottom panels which also shows individual realizations in black.

asymmetric amplification of the normal forcing, η_2. Specifying $\rho = 0.9, b = 0.1, E = 0.2$, and $g = 0.1$ gives output with a similar structure of skewness and kurtosis to the radiosonde data (Fig. 1.6).

Both numerical models reproduce the lobe of increased kurtosis present in radiosonde data. Heuristically, this lobe seems to be a consequence of the fact that high-pass filtering autocorrelated data to time scales shorter than the decorrelation time introduces a strong variance heteroskedasticity.

1.5 Further Discussion and Conclusions

Both the additive and multiplicative noise models fit the observations through inheriting non-normality from synoptic scale processes. This result can be reconciled with Schneider et al. (2015) finding normal variability at synoptic time scales but non-normal variability at longer time scales through considering the effects of band-pass filtering. Filtering to $1/3-1/15$ days^{-1} synoptic frequencies generally leads to insignificant deviations from normality in the radiosonde data (Fig. 1.2), in agreement with analytical and numerical predictions (Figs. 4,5). The findings of non-normal variability for a 1/25-1/35 day^{-1} pass-band results from the discrete frequency basis only retaining a small number of non-zero frequency terms, and, thus, tending towards a distribution determined by the sinusoidal Fourier basis. For completeness, we have also verified that applying the aforementioned filters to the five ERA-Interim 850hPa grid boxes examined by Schneider et al. (2015) gives equivalent results.

Rennert & Wallace (2009) also examine the presence of non-normality as a function of retained frequency. Using data filtered to retain frequencies below (30 day)$^{-1}$, between (6-30 day)$^{-1}$, and above (6 day)$^{-1}$ they attribute skewness to the coupling between the

low and intermediate frequency bands. Non-normality is then postulated to arise from processes with corresponding time scales, e.g., global teleconnections and Rossby-waves. Our findings, however, indicate that non-normality introduced at the highest resolved time-scales in the presence of memory leads to the appearance of greatest skewness when low frequencies are retained (Fig. 1.6). These results suggest that the generating mechanism for non-normal variations does not necessarily act at the frequencies at which greatest non-normality is observed.

Instead, our findings are consistent with processes introducing non-normality on time scales faster than the local decorrelation time scale. Schneider et al. (2015) suggest that mid-latitude temperature distributions are controlled by anomalies advected along a uniform gradient. Such a model may also account for non-normality introduced at synoptic time scales as additive noise, for example, through asymmetries in advection length scale or deviations from a constant gradient introduced by the finite size of the planet or the presence of jets (Luxford & Woollings, 2012). It remains unclear, however, whether the highest resolved frequency of $1/2$ day^{-1} is near that of the process giving rise to the non-normal distribution. Non-normality could also be introduced by atmospheric turbulence at much higher frequencies (Chu et al., 1996), and it will be useful to examine higher resolution temperature records in future work.

The frequency structure associated with skewness and kurtosis are, of course, not the only metrics by which to characterize non-normality associated with a time series. The relation between skewness and kurtosis in samples of both atmospheric and oceanic variability (Sura & Sardeshmukh, 2008; Sardeshmukh & Sura, 2009) has been shown to have a tighter clustering along a parabola than would be expected from purely mathematical considerations (Pearson, 1916; Rohatgi & Székely, 1989). Correlated additive and multiplicative noise can reproduce this parabolic clustering, although a number of

other processes may also give rise to similar relations (Sattin et al., 2009).

Tests for better distinguishing between different classes of non-normal generating processes have seen little progress (Sura & Hannachi, 2015). Autocorrelation in the data seems to be a primary obstacle. One promising technique in this regard is appropriately pre-whitening. Further possibilities include directly testing for multiplicative relationships, examining the extent to which spectral or higher order spectral features suggest the presence of aliasing, and developing approaches for handling non-stationarity associated with diurnal, seasonal, or longer timescale changes in distributional properties.

Appendix 2.1: Influence of discrete frequency space

Filters can cause the resulting signal to tend toward that of a sinusoid when the passband only admits a small number of frequencies. We derive a scaling relationship to quantify what is meant by a small number. First, consider the distribution of a sinusoid with amplitude a,

$$\frac{1}{\pi a\sqrt{1-(\frac{y}{a})^2}}. \tag{1.5}$$

The odd moments vanish on account of symmetry but excess Kurtosis is $-3/2$. The characteristic function of Eq. 1.5 equals $J_0\left(as\right)$, i.e. the zero-th Bessel function of the first kind scaled by amplitude a, and the characteristic function of the distribution of a sum of n such components is the product of their individual characteristic functions, $\prod_{j=1}^{n} J_0\left(a_j s\right)$.

Assuming all a_j equal unity and uniformly distributed phases, the k-th moment

becomes,

$$\mu_k = \mathrm{E}[Y^k] = (-i)^k \frac{d^k}{ds^k} \left(\prod_{j=1}^{n} J_0(s) \right) \Bigg|_{s=0} . \tag{1.6}$$

The fourth moment is solved for by setting $k = 4$, expanding using the chain rule while making use of identities relating derivatives of Bessel functions of the first kind of different orders, and evaluating at $s = 0$. Variance is similarly computed by setting $k = 2$. Excess kurtosis is then,

$$K = \frac{\mu_4}{\sigma^4} - 3 = -\frac{3}{2n}, \tag{1.7}$$

which is within 5% of a normal distribution once ten frequencies are included. There will be phase dependence between different frequencies for non-normal distributions, and synthetic tests confirm that convergence is slower under these conditions.

Appendix 2.2: Analytical and numerical filtering results

The bispectrum, B_F, and trispectrum, T_F, of the filtered time series are

$$B_F(f_1, f_2) = H_2(f_1, f_2) \cdot B(f_1, f_2), \tag{1.8a}$$

$$T_F(f_1, f_2, f_3) = H_3(f_1, f_2, f_3) \cdot T(f_1, f_2), \tag{1.8b}$$

where H_2 and H_3 are defined in terms of the filter $\hat{h}(f)$ as,

$$H_2(f_1, f_2) = \hat{h}\,(f_1) \cdot \hat{h}\,(f_2) \cdot \hat{h}^*\,(f_1 + f_2)\,, \tag{1.9a}$$

$$H_3(f_1, f_2, f_3) = \hat{h}\,(f_1) \cdot \hat{h}\,(f_2) \cdot \hat{h}\,(f_3) \cdot \hat{h}^*\,(f_1 + f_2 + f_3)\,. \tag{1.9b}$$

Since $\hat{x}(f)$ is periodic with period $2f_N$, the top-hat filter $\hat{h}(f)$ is defined accordingly:

$$\hat{h}\,(f) = \begin{cases} 1 & , f_L < (f + f_N \bmod 2f_N) - f_N < f_H \\ 0 & , \text{otherwise,} \end{cases} \tag{1.10}$$

with f_L, f_H the low- and high-frequency cutoffs.

Figure 1.5 shows estimates of skewness and kurtosis obtained by numerical integration. For each $[f_L, f_H]$ pair spanning a set of 50×50 possible combinations, theoretical estimates of skewness and kurtosis of the filtered data are computed by integrating according to equation 1.2. B_F and T_F are computed via (1.8-1.10) on a discrete grid of 1000×1000 and $100 \times 100 \times 100$ frequency bins respectively, and assuming $B(f_1, f_2) = 1$ and $T(f_1, f_2, f_3) = 1$. These values are compared with estimates obtained by filtering synthetic data to the same set of $[f_L, f_H]$ pairs.

For the bispectrum, we can write down analytical values for the relative skewness of the filtered data. Equation 1.10 defines seven different regions in the $[f_L, f_H] \in \{[0, f_N] \times [0, f_N] | f_L < f_H\}$ space of possible pass-bands, each with a different expression for the total area of the admitted bispectrum. The regions are depicted in Fig.1.7 as are analytical estimates of the filtered skewness. The lines a-e delimiting the different

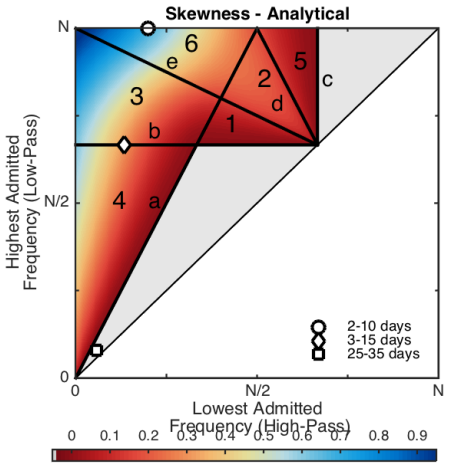

Figure 1.7: Skewness under band-pass filtering, relative to unfiltered values. Axes denote pass band limits, relative to the Nyquist frequency. Gray shaded area denotes values of skewness of exactly zero. Lines a-e delimiting different regions are given in Eqns. 1.11. The value of skewness in regions 1-6 are given in Eqns. 1.12

regions are

$$\begin{aligned} a : f_H &= 2f_L \\ b : f_H &= 2/3, f_L \leq 2/3 \\ c : f_L &= 2/3, f_H \geq 2/3 \\ d : f_H &= 2 - f_L \\ e : f_H &= 1 - f_L/2, \end{aligned} \tag{1.11}$$

while relative skewness in the six non-zero regions is

$$
\begin{aligned}
1: &\quad \frac{(2-3f_H)^2}{4(f_H-f_L)^{3/2}} \\
2: &\quad \frac{-2+3f_H-3f_H^2/4+3f_L-3f_Hf_L-3f_L^2/4}{(f_H-f_L)^{3/2}} \\
3: &\quad \frac{1-3f_H+3f_H^2-3f_Hf_L+3f_L^2}{(f_H-f_L)^{3/2}} \\
4: &\quad \frac{3(f_H-2f_L)^2}{4(f_H-f_L)^{3/2}} \\
5: &\quad \frac{(2-3f_L)^2}{4(f_H-f_L)^{3/2}} \\
6: &\quad \frac{-2+f_H(3-6f_L)+3f_L+9f_L^2/4}{(f_H-f_L)^{3/2}},
\end{aligned}
\tag{1.12}
$$

Appendix 2.3: Test for non-normality

The KS test for non-normality employed here accounts for autocorrelation of the data, as well as the reduced number of degrees of freedom relative to the sample size in filtered data. The test is similar to the Lilliefors variation on the KS test (Lilliefors, 1967). The test statistic is computed as the maximum deviation between the sample cumulative distribution function (CDF) and the CDF of a normal with identical mean and variance. A Monte-Carlo method is employed to compute the null distribution, by computing the maximum deviation between the sample CDF of phase randomized versions of the original data and the CDF of a normal distribution. The phase randomization ensures that the null model is normally distributed, while maintaining the same sample autocorrelation function and the same number of degrees of freedom as the original data. The test is consistent (5% type I errors at the nominal p=0.05 level) for the cases of autocorrelated data, subsampling for DJF, as well samples that have been normalized to unit variance and zero mean, thus avoiding the biases the standard KS test suffers from. A detailed analysis of the performance of this test is presented in

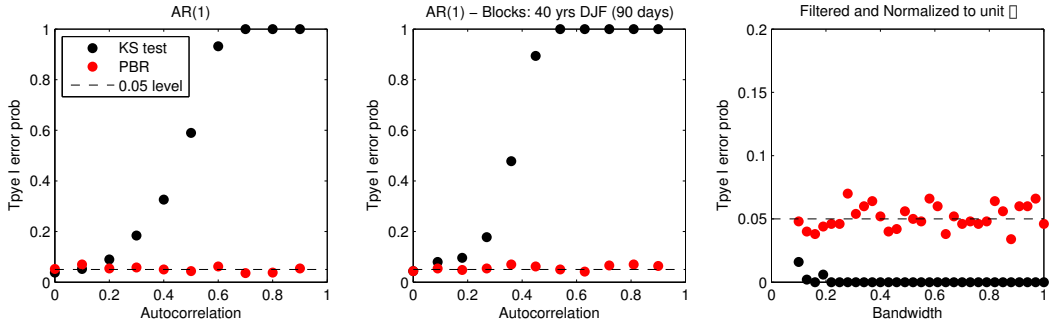

Figure 1.8: Probability of type-I errors (false positives) for the standard KS test and the KS test using phase randomization (KS_{PR}). At the nominal 95% critical value, a consistent test should obtain 0.05 false positives. **Left**: Normally distributed AR(1) process of length N=3500; **Middle**: Normally distributed AR(1) process of length 90*360=14,400, simulating 40 years of data. The data is then sub-sampled to 40 blocks of length 90 simulating an analysis of DJF temperature only. **Right**: I.I.D. Normally distributed data, filtered and normalized to unit variance. Probability of type-I error has been assessed using 500 iterations.

Figs. 1.8 and 1.9.

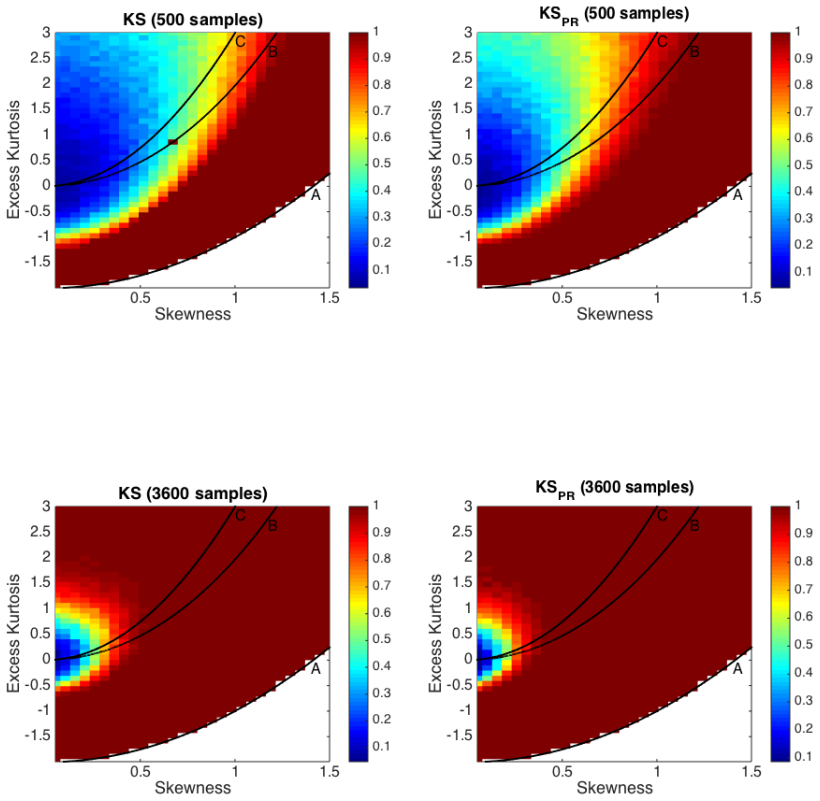

Figure 1.9: Estimates of the Statistical power of the test, defined as one minus the probability of false negatives, for a standard KS test **(left)** and the KS test using phase randomization (KS_{PR}) **(right)**. The data consists of 500 i.i.d. samples drawn from a Pearson'' distribution with zero mean, unit variance and varying skewness and kurtosis. The different regions depict the particular type of distributions. Curve **A** depicts the fundamental inquality of skewness and kurtosis. The region between **A** and **B** consists of Beta distributions; Curve B denotes the family of gamma distributions with unit variance (including chi-squared); Region between **B C** contains Fisher's F-distribution; Curve **C** contains inverse Chi-squared; The Region left of Curve **C** contains the Cauchy and Student's t-distribution.

2

Reconciling Historical and Model-Based Estimates of Climate Sensitivity

Abstract

The latest IPCC Assessment Report widened the Equilibrium Climate Sensitivity (ECS) range from 2–4.5 °C to 1.5–4.5°C in order to account for the lack of consensus between estimates based on models and historical observations. The historical ECS estimates range from 1.5–3°C and are derived assuming a linear radiative response to warming. A Bayesian methodology applied to 24 models, however, documents curvature arising from an evolving contribution of inter-annual through centennial modes of radiative response. Centennial modes display stronger amplifying feedbacks and ultimately contribute 28-68% (90% c.i.) of equilibrium warming, yet they comprise only 1–7% of current warming. Accounting for these unresolved centennial contributions brings historical records into agreement with model-derived ECS estimates.

2.1 Main Text

Estimates of Equilibrium Climate Sensitivity (ECS) from GCMs (Flato et al., 2013) and paleoclimate records (Rohling et al., 2012) are generally consistent with a range of 2–4.5°C, but a number of studies based on historical instrumental records (Gregory et al., 2002; Roe & Armour, 2011; Otto et al., 2013; Masters, 2014; Lewis & Curry, 2014) yield a lower range of 1.5–3°C (Forster, 2016). These systematically lower observational estimates have been interpreted as demonstrating that GCMs are overly sensitive to CO_2 forcing, and that the ultimate amount of warming that the Earth would experience at a given concentration of greenhouse gases is less than previously thought (Otto et al., 2013; Lewis & Curry, 2014).

A major challenge in inferring ECS from instrumental records is that the current

climate system is not in energetic equilibrium. On average, Earth's surface currently takes up between 0.1–0.9 W/m^2 more heat than it loses (Von Schuckmann et al., 2016), where this rate of heating is denoted as H. In order to extrapolate to the temperature at which radiative equilibrium would be re-established, a proportionality between changes in outgoing radiation and temperature is generally assumed (Gregory et al., 2002; Roe & Armour, 2011; Otto et al., 2013; Masters, 2014; Lewis & Curry, 2014; Forster, 2016), $\lambda = (F-H)/T$. F represents anomalies in downward radiative forcing and the difference gives the upward radiative response, $R = F - H$. If λ is assumed constant over time, ECS can be inferred by zeroing out H and assigning F equal to the forcing associated with a doubling of atmospheric CO_2, giving an estimated ECS of $F_{2\times}\lambda^{-1}$.

The validity of such an extrapolation is questionable, however, because of well-documented time dependence of λ in climate simulations (Murphy, 1995; Senior & Mitchell, 2000; Winton et al., 2010; Held et al., 2010; Bitz et al., 2012; Armour et al., 2013; Andrews et al., 2015; Gregory & Andrews, 2016). This time dependence is a prime suspect for the systematic differences between GCMs and historical estimates of ECS (Armour, 2016). Efforts at resolving the discrepancy, however, have been stymied by the lack of a methodology to quantitatively account for changes in λ (Forster, 2016). We derive a generalized fit of the change in λ in 24 GCM simulations from phase 5 of the Climate Model Intercomparison Project (Taylor et al., 2012), and use the results to estimate the bias inherent in historical energy budget constraints on climate sensitivity.

A number of simple linear models have been postulated to describe variability in λ. One set includes a fast-responding upper ocean that is coupled to a more slowly-responding deep ocean (Geoffroy et al., 2013), sometimes including an efficacy term that modifies how much surface warming is associated with oceanic heat fluxes (Winton et al., 2010; Rose et al., 2014). Another set of simple models represents the presence

of distinct regions having their own response time scale and feedbacks (Armour et al., 2013; Andrews et al., 2015). Both of these classes of models represent changing patterns of warming over decades and centuries, where delayed warming of the Southern Ocean (Armour et al., 2013; Andrews et al., 2015) and Eastern Equatorial Pacific (Held et al., 2010) are especially important because of the upwelling of deeper waters in these regions. Interannual time-dependence in the radiative feedbacks in response to warming has also been highlighted as leading to changes in λ, particularly in response to changes in cloud feedbacks in the Tropical Pacific (Andrews et al., 2015).

What model formulations are most useful for representing the heterogenous evolution of the forced response of GCMs remains a subject of ongoing research, but it is possible to generically represent how global mean temperature evolves in existing simple models (Winton et al., 2010; Rose et al., 2014; Geoffroy et al., 2013; Armour et al., 2013; Andrews et al., 2015), or any other linear system, using an eigenmode decomposition (Caldeira & Myhrvold, 2013),

$$T_n(t) = \frac{\alpha_n T_{2\times}}{\tau_n F_{2\times}} \exp(-t/\tau_n) \star F(t), \tag{2.1}$$

where the $\star$ represents the convolution operator. Each eigenmode of the temperature response, T_n, characterizes a rate and magnitude at which global warming is excited by radiative forcing. The rate of response is governed by the timescale, τ_n, and the ultimate magnitude by a fractional contribution, α_n, of the equilibrium climate sensitivity, $T_{2\times}/F_{2\times}$. The amount of warming realized at time t is also a function of the structure of the forcing, F, where eigenmodes with small τ are more strongly excited by high-frequency variations than eigenmodes with larger τ.

The radiative response, R, is represented following the standard approach (Gregory

et al., 2002) of assuming proportionality with T, but where we assign a distinct radiative feedback for each eigenmode of the temperature response, $R_n = \lambda_n T_n$.

We decompose simulations of the temperature and radiative response to increased CO_2 concentrations across 24 GCMs using a full Bayesian inference methodology. Our Bayesian approach allows for fitting simulated global average temperature, $T(t)$, and top-of-atmosphere energy flux, $H(t)$, in a manner that accounts for autocorrelated noise in T and H, uncertainties in extrapolating T and H to full equilibrium, and uncertainties in forcing, $F_{2\times}$. Further, the approach accounts for the fact that coefficients governing the distribution of allowable autocorrelations and uncertainties—the so-called hyperparameters—themselves need to be inferred from the observations. Three eigenmodes are used on the basis of fewer leading to systematic discrepancies between our fit and simulations during initial model adjustment (Fig. 2.8), and use of more eigenmodes failing to improve the fit as judged by a Bayesian Information Criterion. Bayesian fits generally parse the 140-300 year long simulations into distinct annual, decadal, and multi-centennial modes, though results differ slightly across GCMs (Table 2.2).

Fig. 2.1 illustrates the evolution of Earth's energy imbalance with global warming in a representative GCM. As H approaches zero there is typically an inflection such that $\lambda(t)$—or, minus the slope of H with respect to T—decreases in magnitude, and greater warming is required to achieve radiative equilibrium. The ECS across the ensemble has a median value of 3.5°C, with a 5–95% credible interval of 2.2–6.1°C. That the distribution of ECS is skewed (Roe & Baker, 2007) can be seen in Fig. 1 as the geometric result of uncertainties in $\lambda(t)$ having an asymmetric influence on the temperature at which top-of-atmosphere energy flux goes to zero. Individual GCM ECS estimates have maximum likelihood values similar to those reported by Assessment Report 5 of the Intergovernmental Panel on Climate Change (IPCC) (Flato et al., 2013) (Table 2.2).

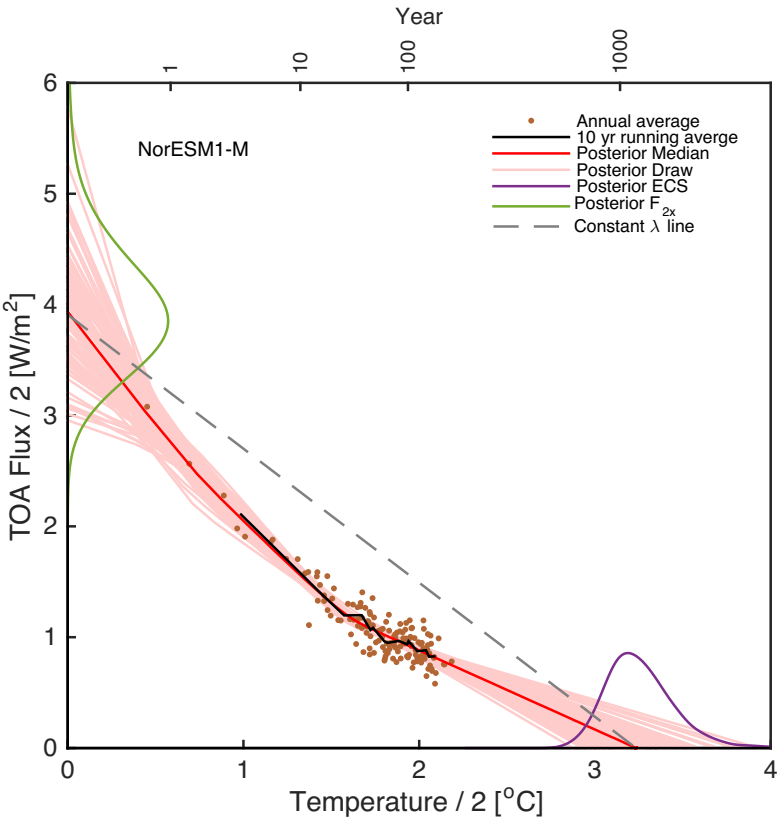

Figure 2.1: Evolution of top-of-atmosphere energy flux as a function of global temperature. Annual average (brown dots) and ten-year running average values (black line) of top-of-atmosphere energy flux (TOA Flux or H) and global-average temperature (T) are shown for a representative CMIP5 GCM (NorESM1-M) along with draws from the Bayesian fits (pinks lines) and their median (red line). The distributions of radiative forcing (green line along y-axis) and equilibrium climate sensitivity (purple line along the x-axis) are estimated as part of the Bayesian fit. Curvature in the fit can be compared against a constant equilibrium feedback parameter connecting median radiative forcing to median warming (gray dashed line), where convexity indicates a decrease in the magnitude of $\lambda(t)$ with time. Note that both T and H are halved in order to facilitate comparison of these results using quadrupled CO_2 forcing with the standard definition of equilibrium climate sensitivity using doubled CO_2. Simulation year is indicated by ticks at top of graph. NorESM1-M was chosen based on having the minimum total deviation from the ensemble median in ECS, ICS and $F_{2\times}$. (See Figs. 2.4-2.7 for the other 23 GCMs.)

The 95th percentile on ECS using is increased significantly, however, from 4.5 to 6.1 °C due to the inflection in $\lambda(t)$ not having been fully accounted for in previous estimates (Andrews et al., 2012; Forster et al., 2013).

As simulations warm, the net feedback operating at any instant represents a weighted average across each mode in the Bayesian fit,

$$\lambda(t) = \frac{\sum_{n=1}^{N} \lambda_n T_n(t)}{\sum_{n=1}^{N} T_n(t)}. \tag{2.2}$$

Estimates of sensitivity based on the historical record are typically obtained by dividing $F_{2\times}$ by $\lambda(t)$. We term this estimate the instantaneous climate sensitivity, $\mathrm{ICS}(t) = F_{2\times}/\lambda(t)$. ICS is sometimes referred to as effective ECS (Forster, 2016) when discussing instrumental estimates, but is distinct from other definitions of effective ECS used for GCMs (Bitz et al., 2012). Moreover, our point is that ICS may be an ineffective estimate of ECS.

Eq. 2.2 represents how different modes of warming cause variations in ICS over time. The slowest mode has a characteristic time scale of 350 years (c.i., 180–960 years) and contributes a median of 44% (c.i., 28–66%) of the total equilibrium warming across the Bayesian fits. Historical forcing, however, only weakly excites this slow mode, such that it only contributes 3% (c.i., 1–7%) of the historical warming (Fig. 2, Table 1). This contribution is computed by applying a best estimate of effective historical forcing from 1750 to 2011 provided in IPCC Assessment Report 5 (AR5) (Myhre et al., 2013) to the Bayesian fit associated with each GCM. This small response reflects that the slowest mode still remembers volcanic cooling episodes prevalent between 1750-1900, and that it is slow to realize warming from more recent increases in radiative forcing. By comparison, had the 2011 forcing of 2.2 W/m^2 been continuously applied since

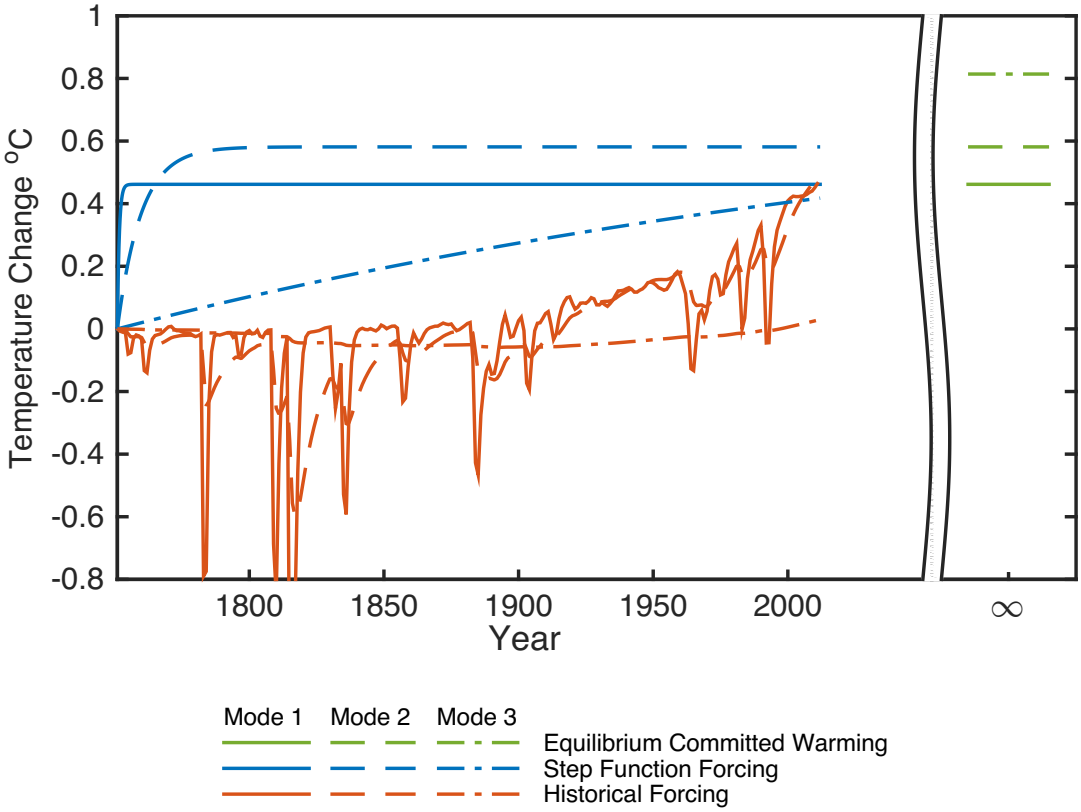

Figure 2.2: Evolution of fast and slow modes under historical forcing. Response of the CMIP5 GCMs to AR5 historical forcing from 1750–2011 (red) and to an abrupt increase in 1750 to the 2011 forcing value of 2.2 W/m^2 (blue). Responses are derived from the median Bayesian fit, which permits for parsing fast (modes 1 and 2) and slow (mode 3) contributions and estimating equilibrium conditions (green lines at right). Equilibrium responses are to the 2011 forcing of 2.2 W/m^2, also known as committed warming (Meehl et al., 2005; Wigley, 2005). Mode 3 accounts for 44% of equilibrium warming but only 3% of present warming. See Table 2.1 for the median and 5-95% range for relevant parameters.

eigenmode parameters	median	5%	95%
τ_1 [years]	0.8	0.2	2.6
λ_1 [W/m^2/°C]	1.6	0.4	4.0
Contribution to equilibrium warming	24%	10%	40%
Contribution to historical warming	47%	26%	75%
τ_2 [years]	9	4	37
λ_2 [W/m^2/°C]	1.4	0.8	2.8
Contribution to equilibrium warming	32%	15%	48%
Contribution to historical warming	49%	24%	72%
τ_3 [years]	350	180	960
λ_3 [W/m^2/°C]	0.8	0.3	1.6
Contribution to equilibrium warming	44%	28%	66%
Contribution to historical warming	3%	1%	7%

Table 2.1: Fast and slow modes contributions to equilibrium and historical warming. Posterior median and 5-95% credible interval values for the time scale and feedback factors of individual eigenmodes, and the fraction they contribute to equilibrium and historical warming.

1750, the slow mode would instead contribute 29% (c.i., 18–43%) of the warming. The structure of historical forcing is thus strongly biased toward sampling the faster modes of response with characteristic time scales of 0.8 years (c.i., 0.2–2.6 years) and 9 years (c.i., 4–37 years).

The faster modes of response have larger magnitudes of λ, with best estimates of 1.6 (c.i. 0.4–4.0) W/m^2/°C for the ultrafast and 1.4 (c.i. 0.8–2.8) W/m^2/°C for the fast mode, as compared to 0.8 (c.i. 0.3–1.6) W/m^2/°C for the slow mode. Larger magnitudes of λ indicate that less warming is required to achieve radiative equilibrium. Thus sampling essentially only the fast modes in historical estimates biases ICS values low

compared to ECS. In particular, sampling ICS in a manner consistent with observational estimates (Methods) leads to an ICS of 2.5°C (c.i. 1.6–5.6°C), or 1°C cooler than ECS (Fig. 3). These findings are also consistent with results from simulations using prescribed historical sea surface temperatures. Prescribed historical sea surface temperatures only weakly excite the slow mode of radiative response, analogous to runs wherein historical forcing is prescribed, and similarly lead to the inference of smaller equilibrium warming than is associated with CO_2 quadrupling experiments (Gregory & Andrews, 2016).

A recent review of observationally-based estimates of ICS shows a median value of 2°C and an 80% range of 1.6–3°C, where the range is obtained by trimming the highest and lowest estimates from across the ten available studies (Forster, 2016). This range of observationally-inferred ICS values is contained within and centered on the maximum likelihood value of the distribution of ICS values that we estimate from GCM simulations (Fig. 2.3b). Accounting for further issues involving forcing efficacy (Kummer & Dessler, 2014; Marvel et al., 2016) and temperature sampling methodology (Richardson et al., 2016) might narrow the distributions and bring them into even closer agreement.

Slow feedback contributions to warming are only weakly manifest in the current climate system because the rise in greenhouse gas concentrations that occurred primarily in the last 50 years has been active for a short period relative to the multi-centennial time scale of the dominant slow mode. This small manifestation makes it difficult, if not impossible, to accurately estimate ECS from historical observations, and highlights the importance of paleoclimate records for purposes of observationally constraining slow feedback processes (Rohling et al., 2012). Looking forward, these results also highlight that the warming current greenhouse gas concentrations have committed us to (Meehl et al., 2005; Wigley, 2005) will evolve according to a slow mode of response that is distinct from the present warming pattern.

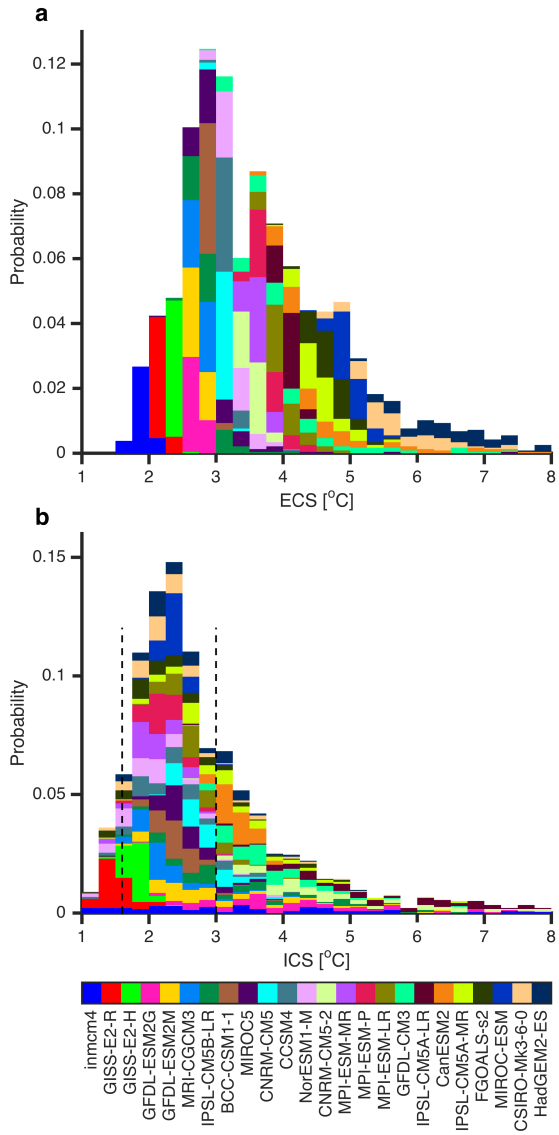

Figure 2.3: Equilibrium and instantaneous climate sensitivity distributions. a, Distribution of equilibrium climate sensitivity, ECS, from 5000 posterior draws of our Bayesian fit to each of 24 GCMs (indicated by colors). Aggregating across the posterior draws for all GCMs yields a median value of 3.4°C and a 5–95% credible interval of 2.2–6.1°C. **b,** Similar to (a) but for instantaneous climate sensitivity. ICS is obtained by applying AR5 historical forcing to our Bayesian fits, and has a median value of 2.5 and a 5–95% credible interval of 1.6–5.6°C. The range of ICS values estimated in historical studies (Forster, 2016) (vertical black lines) bracket the most likely GCM values, demonstrating consistency between observational and GCM results when they are appropriately compared (methods)

2.2 Methods

2.2.1 Eigenmode Decomposition

Eigenmode Decomposition is performed upon global and annual average temperature, T, and top-of atmosphere energy flux, H, using a full Bayesian statistical inference methodology. Simulations involving an instantaneous quadrupling of CO_2 are considered across 24 GCMs (Table 2.2). These simulations have the highest signal to noise ratio of the available CMIP5 simulations; do not suffer from issues of collinearity between forcing and response as in simulations with linear increase in forcing; are not subject to uncertainties in the structure of the applied forcing as in the historical GCM simulations; and the sustained forcing will most strongly excite the slowest modes of response. Additionally, values of ECS stated by the IPCC are based on these abrupt quadrupling simulations.

The constant forcing, denoted as $2F_{2\times}$, permits for simplification of Eq. 1(23)

$$T(t) = 2T_{2\times} \sum_{n=1}^{3} \alpha_n \left(1 - \exp\left(-t/\tau_n\right)\right), \tag{2.3}$$

where the evolution toward an equilibrium warming of $2T_{2\times}$ involves three eigenmodes having contributions of relative magnitudes, α_n, and characteristic time scales, τ_n.

The radiative response to increased global average surface temperature is represented as evolving along the same eigenmodes,

$$R(t) = 2F_{2\times} \sum_{n=1}^{3} \beta_n (1 - \exp(-t/\tau_n)), \tag{2.4}$$

toward a radiative equilibrium wherein R equals the radiative forcing, $2F_{2\times}$. Fractional

contributions of eigenmode n are given by β_n. It follows that the radiative feedback associated with mode n is,

$$\lambda_n = \frac{R_n}{T_n} = \frac{\beta_n F_{2\times}}{\alpha_n T_{2\times}}, \tag{2.5}$$

where $F_{2\times}/T_{2\times}$ is the net equilibrium feedback. Previously described simple linear models for changes in $\lambda(t)$(13,15,16,20,22) all assume three or fewer modes such that their temporal evolution can be fully described using Eqs. 2.3–2.5 (see Supplementary Text).

The radiative response, R, is not directly available from GCM simulations. However, the top-of-atmosphere energy flux, H, can be related to the radiative response as $H = F - R$, yielding,

$$H(t) = 2F_{2\times}\left[1 - \sum_{n=1}^{3} \beta_n \left(1 - \exp\left(-t/\tau_n\right)\right)\right], \tag{2.6}$$

for the abrupt CO_2 quadrupling experiment. Note that responses are corrected for GCM drift using the standard methodology wherein drift values for T and H are computed in pre-industrial control runs and subtracted from the quadrupling simulation.

2.2.2 Bayesian Inference

Bayesian Inference is performed by conditioning the parameters in Eqs. 2.3 and 2.6 on simulated time-series of T and H. Variability that is not linearly related to the forced response in T and H—which arises, for example, from internal modes of variability such as El Niño—is represented as a normally distributed order-one autoregressive processes with auto-correlation parameters, ρ_T and ρ_H, and variances, $\sigma_T^2/(1-\rho_T^2)$ and $\sigma_H^2/(1-\rho_H^2)$. This formulation is supported by the posterior residuals (Figs. 2.9 and 2.10)

having a structure consistent with an autoregressive process and exhibiting only minor deviations from a normal distribution when subjected to a Kolmogorov-Smirnov test that accounts for autocorrelation (Proistosescu et al., 2016).

Uniform priors are prescribed for parameters defined on bounded intervals. Autocorrelations ρ_T and ρ_H are defined on the interval 0 to 1, and α_n and β_n are defined on the surfaces defined by the sum of points, $\sum_{n=1}^{3} \alpha_n = 1$ and $\sum_{n=1}^{3} \beta_n = 1$. Unbounded parameters are given weakly informative exponential priors that help the sampling scheme converge more quickly by focussing sampling away from implausible regions. $T_{2\times}$ and $F_{2\times}$ have exponential priors with means of 10°C and 10 W/m^2, respectively; timescales τ_1, τ_2, and τ_3 have exponential priors with means of 10, 100, and 1000 years; and standard deviations σ_T and σ_H have exponential priors with means of 10°C and 10 W/m^2. The posterior maximum likelihood value for each parameters is consistent within 1% even if the means of all exponential priors are either increased or decreased by a factor of 10, whereas the upper 95% value associated with each parameter is consistent within 10%. This low sensitivity to drastic changes in priors demonstrates that the simulation results from GCMs determine the posterior estimates.

Full Bayesian statistical inference is performed using the statistical modeling platform STAN (Carpenter et al., 2016) In particular, a joint posterior distribution is obtained for $F_{2\times}$, $T_{2\times}$, τ_n, α_n, β_n, σ_T, σ_H, ρ_T, and ρ_H using a Hamiltonian version of Markov Chain Monte Carlo with a No-U-Turn Sampler (Hoffman & Gelman, 2014). Posterior distributions for radiative feedbacks, λ_n, are computed from the joint posterior using Eq. 2.5. Results shown in the main text are from 5000 draws across five chains of 1000 realizations. Each chain is sampled after an initial burn-in period of 500 draws. In order to provide a complete representation of our results, these draws from the joint posterior are provided in supplemental material.

2.2.3 Instantaneous Climate Sensitivity

ICS is inferred for the historical period by applying historical forcing to the estimated eigenmode response for each GCM. In particular, we apply the best estimate AR5 historical forcing as used in previous studies, to draws from the posterior distribution of the eigenmode decomposition for each GCM to obtain time series of T and H.

In the main text, $\lambda(t)$ is computed as $(F - H)/T$ where anomalies in forcing, heat uptake, and global temperature are computed as the departures in 2011 from preindustrial conditions, and ICS is obtained as $F_{2\times}/\lambda(t)$. Note that values of $F_{2\times}$ associated with each posterior draw can vary from the 3.7 W/m^2 assumed in the AR5 estimate of historical forcing. In order to propagate this uncertainty, F is multiplied by $F_{2\times}/3.7$ for each draw before obtaining values of H and T. We obtain a median ICS value across all draws and GCMs of 2.5°C (c.i. 1.5–5.4 °C).

Inferences of 'effective ECS' from the observational record, which we refer to as ICS, have been obtained using different intervals, but these alternate choices do not much influence our estimates. For example, if $\lambda(t) = (\Delta F - \Delta H)/\Delta T$ is computed between the intervals 1859–1883 and 1995–2011, as per Lewis & Curry (2014) we again obtain an ICS of 2.5°C (c.i. 1.4–5.8°C). Similarly, using intervals of 1860–1879 and 2000–2009, as per Otto et al. (2013), gives 2.6°C (c.i. 1.5–5.2°C), and using the difference between values in 1955 and 2011, as per Masters (2014), gives 2.6°C (c.i. 1.4–5.5°C).

2.3 Supplementary Text

We provide solutions based on linear response theory for two simple models used in the literature and cited in the main text to describe time dependency of the net radiative feedback.

2.3.1 Linear Model Response

Let $\mathbf{y}$ be the state vector the climate system. If we assume that $\mathbf{y}$ describes small anomalies around a state of equilibrium, the evolution of the climate system in response to an external forcing $\mathbf{F}$ is described by

$$\frac{d}{dt}\mathbf{y} = \mathbf{J}\mathbf{y} + \mathbf{F}(t), \tag{2.7}$$

where $\mathbf{J}$ is the Jacobian of the system. We apply classic linear system methods and diagonalize the Jacobian as

$$\mathbf{J} = \mathbf{P}^{-1}\mathbf{D}\mathbf{P}, \tag{2.8}$$

where $\mathbf{P}$ is the modal matrix containing the eigenvectors of the Jacobian, and $\mathbf{D}$ is the diagonal spectral matrix containing the eigenvalues. The exponential decay time scales of the eigenmodes are the negative inverses of the eigenvalues, such that $\tau_n = -1/\mathbf{D}_{nn}$. The system has the well known solution (Greenberg, 1988):

$$\mathbf{y} = \mathbf{y}\,(0) + \int_0^t \mathbf{P}^{-1} e^{\mathbf{D}\tau}\mathbf{P}\mathbf{F}\,(t-\tau)\,d\tau. \tag{2.9}$$

Global mean temperature is a weighted sum of members of the state vector corresponding to surface temperature,

$$T = \mathbf{w}_\mathbf{T}^\mathbf{t}\mathbf{y}. \tag{2.10}$$

Further, we assume that the system starts from a state of equilibrium, $\mathbf{y}(0) = 0$, and that there is a single forcing time series, $F(t)$, imposed on the system that is projected

onto the different dimensions of the system according to

$$\mathbf{F} = \mathbf{w_F} F(t). \tag{2.11}$$

We can now write the global temperature response, T, as:

$$T = \int_0^t \mathbf{w_T} \mathbf{P}^{-1} e^{\mathbf{D}\tau} \mathbf{P} \mathbf{w_F} F\left(t-\tau\right) d\tau. \tag{2.12}$$

$$T = \sum_n \sum_k \sum_l \mathbf{w_T}_l \mathbf{P^{-1}}_{ln} \mathbf{P}_{nk} \mathbf{w_F}_k \exp\left(-t/\tau_n\right) \star F(t). \tag{2.13}$$

If we divide and multiply by τ_k, and by the equilibrium sensitivity parameter,

$$\frac{T_{eq}}{F_{eq}} = -\mathbf{w_T} \mathbf{J}^{-1} \mathbf{w_F} = -\mathbf{w_T} \mathbf{P}^{-1} \mathbf{D}^{-1} \mathbf{P} \mathbf{w_F} \tag{2.14}$$

we recover Eqn. 1 in the main manuscript

$$T = \frac{T_{eq} \alpha_n}{F_{eq} \tau_n} \sum_n \exp\left(-t/\tau_n\right) \star F(t), \tag{2.15}$$

where

$$\alpha_n = \frac{\sum_m \sum_k \mathbf{w_T}_m \mathbf{P^{-1}}_{mn} \tau_n \mathbf{P}_{nk} \mathbf{w_F}_k}{\mathbf{w_T} \mathbf{P^{-1}} \mathbf{D^{-1}} \mathbf{P} \mathbf{w_F}}. \tag{2.16}$$

with $\sum_n \alpha_n = 1$.

The net global radiative response, R, will likewise be a linear function of the state vector,

$$R_g = \mathbf{w_R} \mathbf{y}, \tag{2.17}$$

which has to satisfy the condition that in equilibrium $R_{eq} = F_{eq}$, and thus

$$\mathbf{w_R P^{-1} D^{-1} P w_F} = 1. \tag{2.18}$$

We can now write

$$R = \sum_m \sum_k \sum_n \mathbf{w_R}_m \mathbf{P^{-1}}_{mn} \mathbf{P}_{nk} \mathbf{w_F}_k \exp\left(-t/\tau_n\right) \star F(t), \tag{2.19}$$

$$R = \sum_n \beta_n \frac{\exp\left(-t/\tau_n\right)}{\tau_n} \star F(t), \tag{2.20}$$

with

$$\beta_n = \sum_m \sum_k \mathbf{w_R}_m \mathbf{P^{-1}}_{mn} \tau_n \mathbf{P}_{nk} \mathbf{w_F}_k \tag{2.21}$$

The radiative feedbacks associated with each eigenmode can be computed (as per Eqn. 5 in the main text) as $\lambda_n = \beta_n / \alpha_n F_{eq} / T_{eq}$

$$\lambda_n = \frac{\sum_m \sum_k \mathbf{w_R}_m \mathbf{P^{-1}}_{mn} \tau_n \mathbf{P}_{nk} \mathbf{w_F}_k}{\sum_m \sum_k \mathbf{w_T}_m \mathbf{P^{-1}}_{mn} \tau_n \mathbf{P}_{nk} \mathbf{w_F}_k} \tag{2.22}$$

2.3.2 Regional feedbacks model:

In order to explain the time dependency of the net feedback, a three box model has been proposed (Armour et al., 2013). The model has three independent regions, each evolving with their own characteristic time scale and radiative feedback,

$$C_j \frac{dT_j}{dt} = -\lambda_j T_j + F, \tag{2.23}$$

$$T = \frac{1}{3} \sum_{j=1}^{n} T_j. \tag{2.24}$$

The matrices associated with this model are

$$\mathbf{w_F} = \begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix}, \quad \mathbf{w_T} = \begin{bmatrix} 1/C_1 \\ 1/C_2 \\ 1/C_3 \end{bmatrix}, \quad \mathbf{w_R} = \begin{bmatrix} \lambda_1/3 \\ \lambda_2/3 \\ \lambda_3/3 \end{bmatrix}, \tag{2.25}$$

$$\mathbf{J} = \begin{bmatrix} -\frac{\lambda}{C_1} & 0 & 0 \\ 0 & -\frac{\lambda_2}{C_2} & 0 \\ 0 & 0 & -\frac{\lambda_3}{C_3} \end{bmatrix}. \tag{2.26}$$

Since the Jacobian matrix is already diagonal, each region represents an eigenmode, such that the solution is trivial:

$$\tau_j = C_j/\lambda_j, \tag{2.27}$$

$$\alpha_j = \frac{1}{3}\frac{\lambda_j^{-1}}{\sum_{j=1}^{n}\lambda_j^{-1}}, \tag{2.28}$$

and the radiative feedbacks associated with each eigenmode are equal to the regional feedbacks. Energy fluxes between regions can be modeled by adding symmetrical off-diagonal terms in the Jacobian matrix, however in that case the radiative feedbacks associated with each eigenmode would no longer be equal to the regional feedbacks.

2.3.3 Two-layer ocean model:

Another standard conceptualization of the evolution of the system is one where a fast-responding upper ocean is coupled to a slow-responding deep ocean,

$$C_s \frac{dT_s}{dt} = -\lambda_{eq} T_s - \varepsilon\gamma(T_s - T_d) + F, \tag{2.29}$$

$$C_d \frac{dT_d}{dt} = \gamma(T_s - T_d). \tag{2.30}$$

$C_{s,d}$ are the heat capacities of the surface and deep ocean, and $T_{s,d}$ are the temperatures. Thermal coupling between the surface and deep is represented by γ, the equilibrium feedback parameter is λ_{eq}, and radiative forcing is F. Also included is a a term for the efficacy of ocean heat uptake, ε. This term can be understood in the context of an effective forcing whereby the same global radiative forcing can illicit different global temperature responses (Hansen et al., 2005), but where it is the transfer of heat between surface and deeper layers that forces surface temperature (Winton et al., 2010; Held et al., 2010). The radiative response of this two-layer ocean model can be written as

$$R = \lambda_{eq} T - (1 - \varepsilon)(T - T_d), \tag{2.31}$$

and the matrices associated with this model are

$$\mathbf{w_F} = \begin{bmatrix} 1/C_s \\ 0 \end{bmatrix} \quad \mathbf{w_T} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \mathbf{w_R} = \begin{bmatrix} \lambda_{eq} + (\varepsilon - 1) \\ 1 - \varepsilon \end{bmatrix}, \tag{2.32}$$

$$\mathbf{J} = \begin{bmatrix} \frac{-\lambda_{eq} + \varepsilon\gamma}{C_s} & \frac{\varepsilon\gamma}{C_s} \\ \frac{\gamma}{C_d} & \frac{-\gamma}{C_d} \end{bmatrix}. \tag{2.33}$$

They can be used to solve for:

$$\tau_{1,2} = \frac{2C_dC_s}{(\varepsilon\gamma + \lambda_{eq})\,C_d + \gamma C_s \pm \Delta} \tag{2.34}$$

$$\alpha_{1,2} = \frac{\Delta \pm (\varepsilon\gamma - \lambda_{eq})\,C_d + \gamma C_s}{2\Delta} \tag{2.35}$$

$$\lambda_{1,2} = \lambda_{eq}\frac{1+\varepsilon}{2} + (\varepsilon - 1)\,\frac{\gamma C_s + \varepsilon\gamma C_d \pm \Delta}{2\varepsilon C_d} \tag{2.36}$$

where

$$\Delta = \sqrt{((\varepsilon\gamma + \lambda_{eq})C_d + \gamma C_s)^{\,2} - 4\gamma\lambda_{eq}C_dC_s} \tag{2.37}$$

As expected, for the case of unit efficacy the feedbacks along the two eigenmodes are the same. Another formalism proposed to account for time-variability in feedback strengths is that of a virtual forcing (Rugenstein et al., 2016), but because this representation is mathematically equivalent to the two-layer ocean model described here, the same form of eigenmode solution holds.

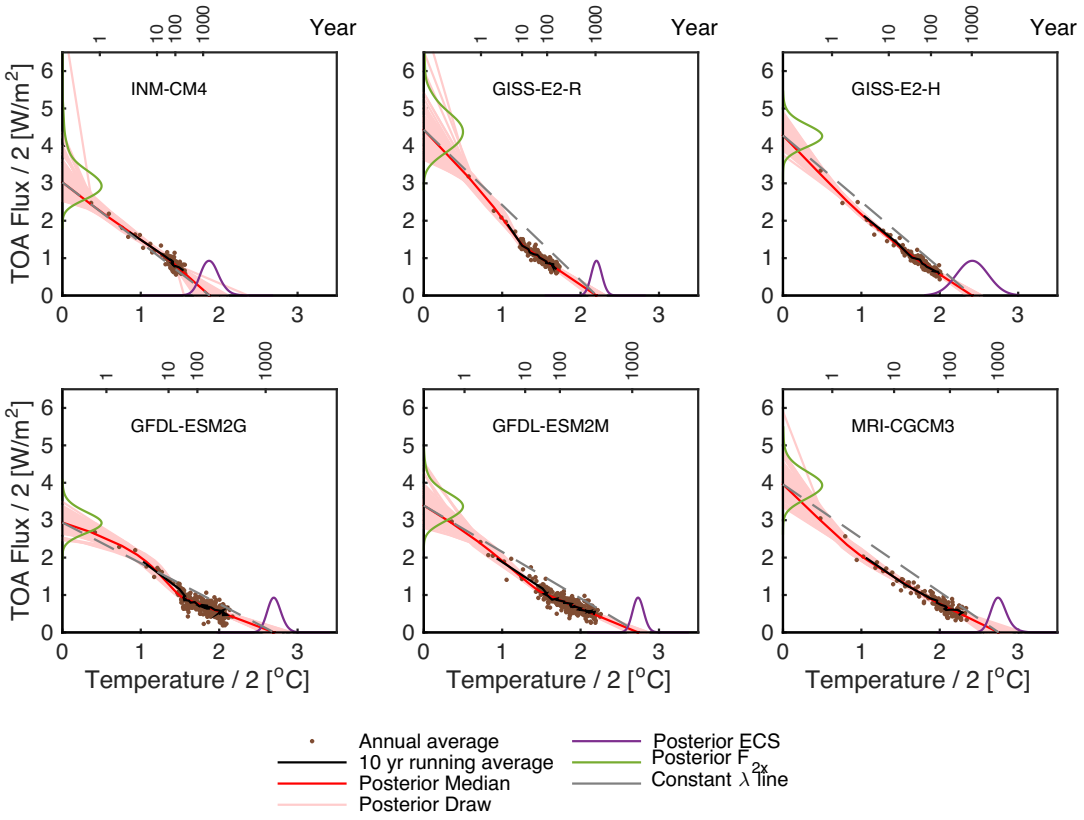

Figure 2.4: Fig. 2.1 - continued

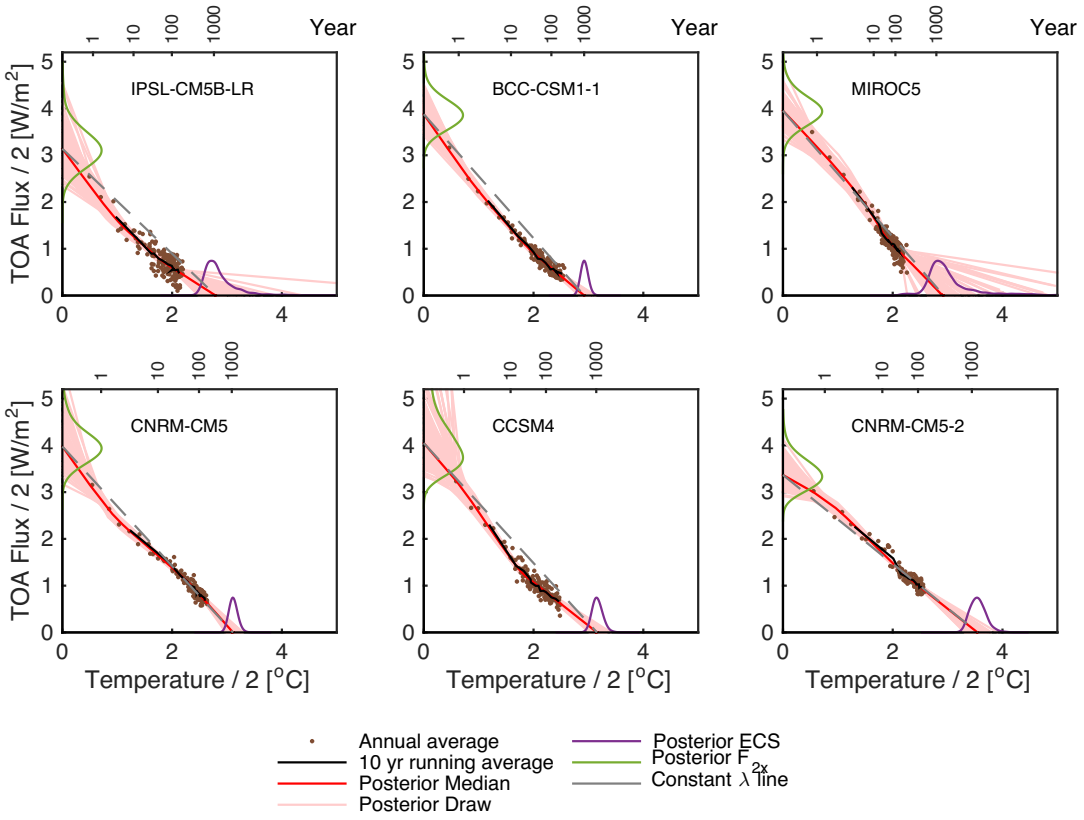

Figure 2.5: Fig. 2.1 - continued

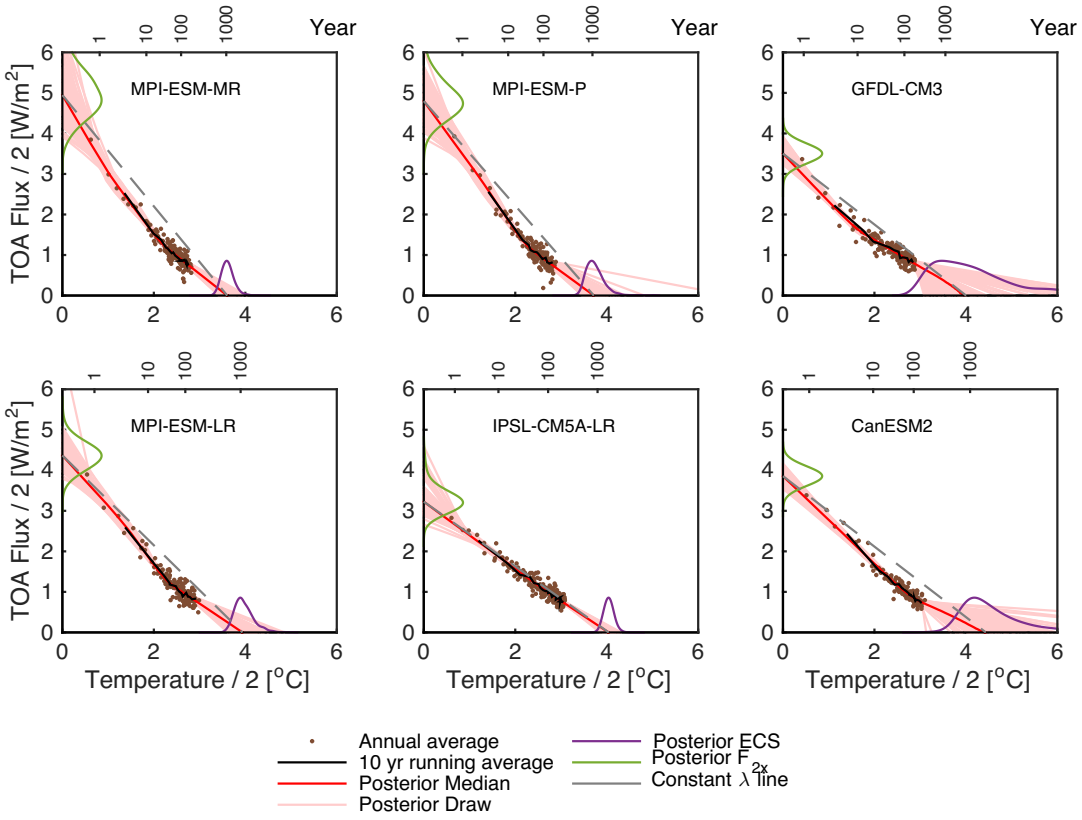

Figure 2.6: Fig. 2.1 - continued

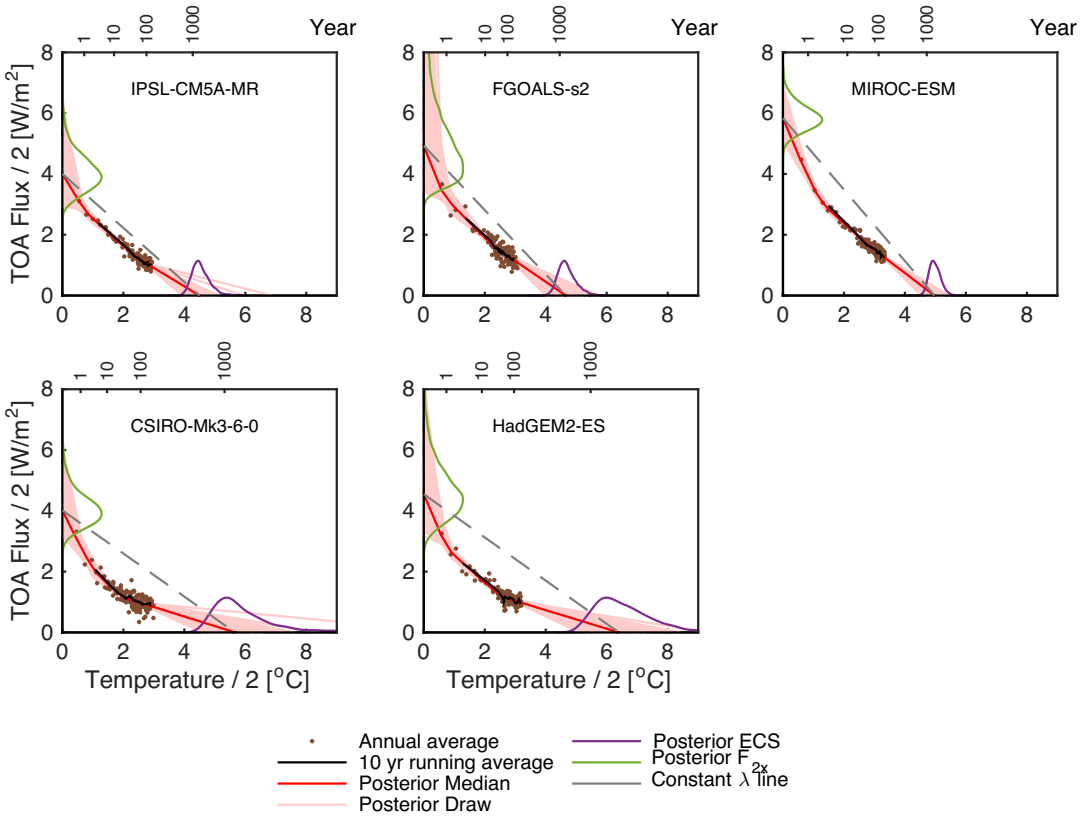

Figure 2.7: Fig. 2.1 - continued

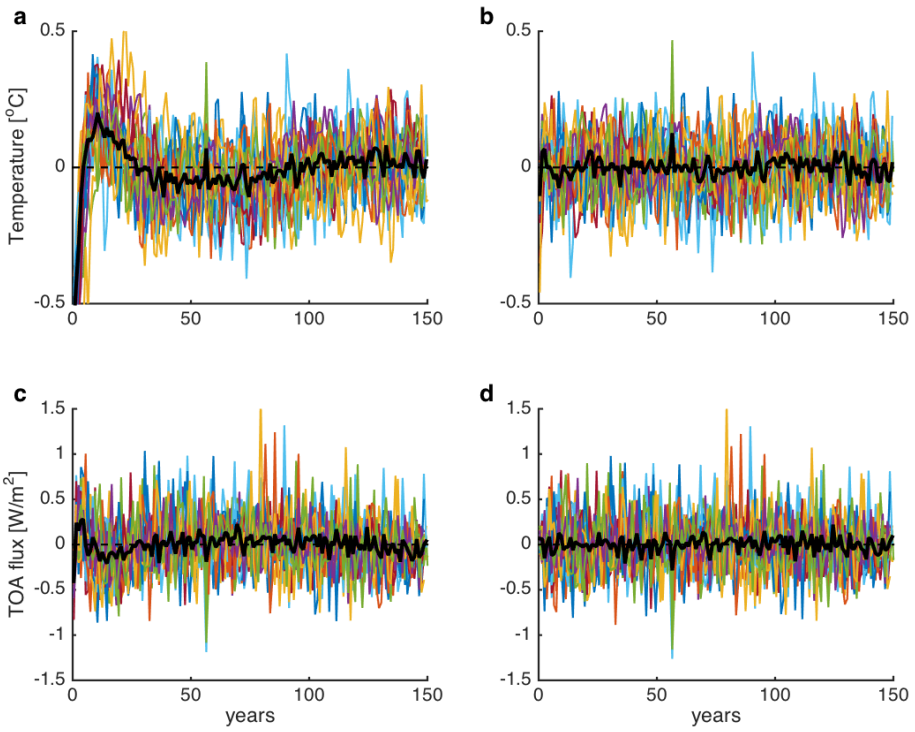

Figure 2.8: Residual temperature and top-of-atmosphere energy flux between GCM simulations and Bayesian fits. Residuals are shown for two-eigenmode (**a,c**) and three-eigenmode (**b,d**) fits for each GCM (colored lines). The median residual across models (thick black line) shows systematic residuals for the two-eigenmode fits during initial decades after abrupt quadrupling of CO_2 concentrations of model adjustment that are not present in the three-eigenmode fit. All eigenmode fits use posterior maximum likelihood values.

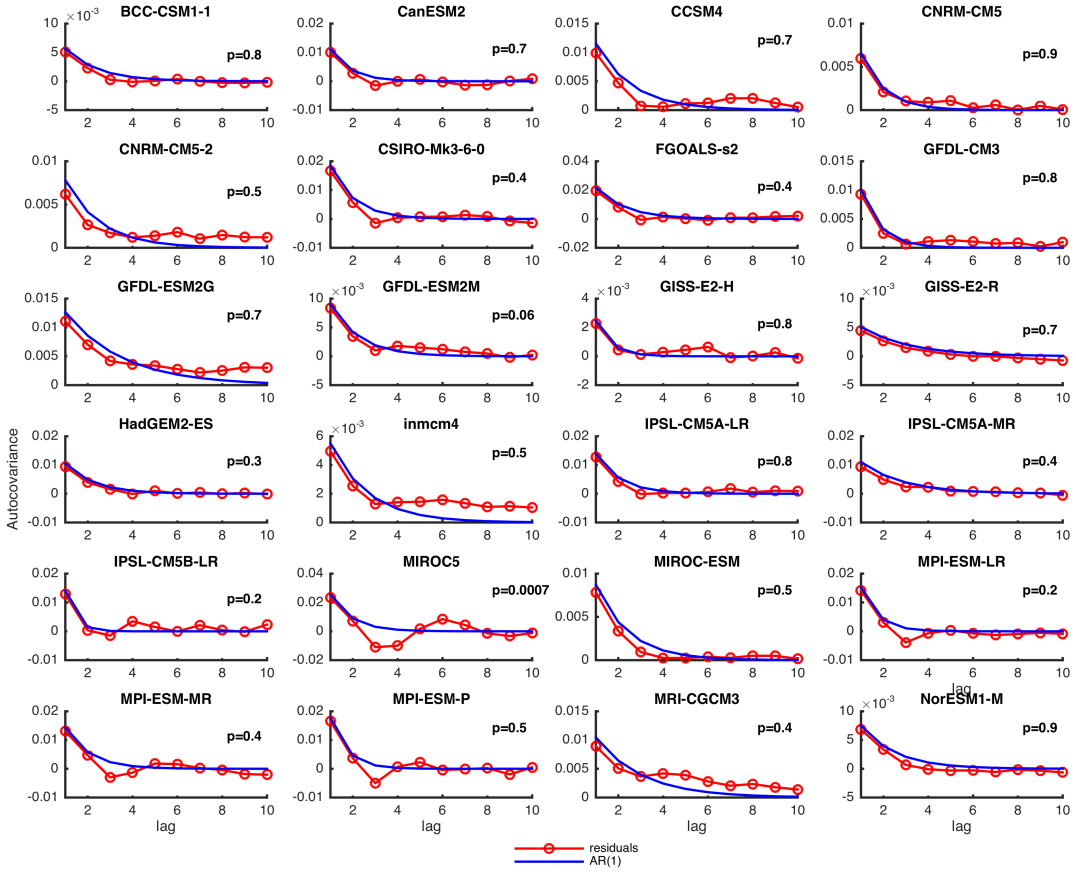

Figure 2.9: Structure of temperature residuals. Auto-covariance of the residuals between the GCM simulations and the Bayesian fit for each model (red), and that inferred for residual noise by the Bayesian fit (blue, from values of σ_T and ρ_T). Maximum likelihood parameters associated with each Bayesian fit are used in these calculations. The p-value of a Kolmogorov-Smirnov normality test for the residuals that accounts for autocorrelation (Proistosescu et al., 2016) is also displayed for each GCMI. Only MIROC 5 shows significant deviations from normality ($p < 0.05$) in the temperature residuals.

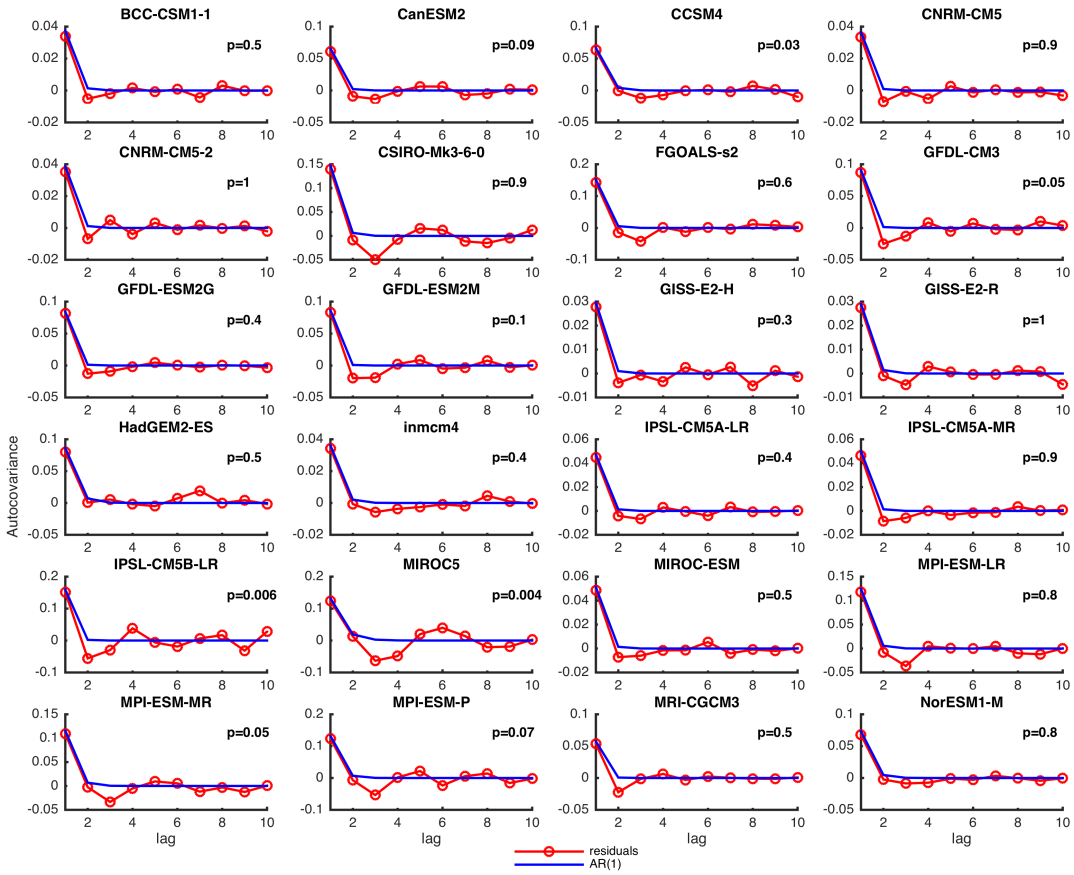

Figure 2.10: Structure of energy flux residuals. Same as Fig. 2.9, but for the top-of-atmosphere energy flux, H. Statistically significant deviations from a normal distribution are found for MIROC 5, CCSM4 and IPSL-CM5B-LR. Notwithstanding the significance of the p-value for MIROC 5, four rejections out of 48 trials across the results for temperature and energy flux is consistent with that expected when the null hypothesis holds and a test is performed at the 5% significance level.

model	τ_1	τ_2	τ_3	α_1	α_2	α_3	λ_1	λ_2	λ_3	σ_T	σ_H	ρ_T	ρ_H	$F_{2\times}$	ECS	ICS
		years					$^\circ$C/(W/m^2)			$^\circ$C	W/m^2			W/m^2	$^\circ$C	$^\circ$C
inmcm4	0.4	5	535	0.19	0.53	0.28	1.7	1.5	1.9	0.06	0.19	0.55	0.06	3.1	1.9	2.3
GISS-E2-R	0.4	3	229	0.30	0.25	0.44	2.1	3.0	1.4	0.06	0.17	0.63	0.05	4.4	2.2	2.5
GISS-E2-H	0.7	8	191	0.33	0.31	0.37	2.2	1.8	1.4	0.05	0.17	0.23	0.03	4.3	2.4	2.5
GFDL-ESM2G	1.0	6	407	0.31	0.23	0.46	0.5	2.5	0.8	0.08	0.29	0.68	0.01	2.9	2.7	2.4
GFDL-ESM2M	0.7	6	348	0.20	0.31	0.49	1.2	1.9	0.9	0.08	0.29	0.46	0.01	3.4	2.7	2.4
MRI-CGCM3	0.7	11	209	0.28	0.35	0.36	2.1	1.3	1.0	0.08	0.24	0.62	0.01	4.0	2.7	3.2
IPSL-CM5B-LR	1.1	15	278	0.31	0.27	0.42	1.6	1.1	0.7	0.12	0.40	0.10	0.01	3.1	2.8	2.4
MIROC5	1.5	9	442	0.40	0.20	0.38	1.2	2.0	1.2	0.15	0.36	0.35	0.14	4.0	2.9	2.3
BCC-CSM1-1	0.8	8	157	0.26	0.35	0.39	1.7	1.4	1.0	0.06	0.19	0.50	0.04	3.9	2.9	2.6
CNRM-CM5	0.6	9	271	0.25	0.46	0.29	1.7	1.0	1.3	0.07	0.19	0.38	0.03	4.0	3.1	2.6
CCSM4	0.3	5	212	0.19	0.34	0.47	1.1	1.8	0.9	0.09	0.26	0.54	0.06	4.0	3.2	2.6
NorESM1-M	0.5	7	342	0.18	0.28	0.54	2.1	1.6	0.7	0.07	0.27	0.53	0.06	3.9	3.2	2.5
CNRM-CM5-2	0.5	7	380	0.21	0.39	0.40	0.6	1.2	0.9	0.07	0.20	0.53	0.03	3.4	3.5	2.7
MPI-ESM-MR	0.5	7	234	0.23	0.34	0.43	2.0	1.6	0.9	0.11	0.34	0.40	0.06	5.0	3.6	2.5
MPI-ESM-P	0.8	9	255	0.30	0.26	0.44	1.5	1.8	0.8	0.13	0.36	0.26	0.05	4.8	3.7	2.3
MPI-ESM-LR	1.2	9	267	0.28	0.25	0.46	1.1	1.7	0.8	0.12	0.35	0.26	0.05	4.3	3.9	2.6
IPSL-CM5A-LR	1.0	20	528	0.26	0.35	0.39	0.8	0.8	0.8	0.11	0.22	0.40	0.03	3.2	4.0	2.4
GFDL-CM3	2.7	54	1010	0.35	0.30	0.36	1.2	0.7	0.6	0.10	0.30	0.31	0.02	3.5	4.1	2.4
CanESM2	2.5	30	597	0.38	0.19	0.43	1.1	1.2	0.5	0.10	0.26	0.33	0.03	3.9	4.4	2.6
IPSL-CM5A-MR	0.4	11	434	0.15	0.37	0.49	1.8	0.9	0.6	0.08	0.22	0.60	0.03	4.0	4.5	2.6
FGOALS-s2	0.2	5	403	0.10	0.37	0.53	3.2	1.0	0.7	0.13	0.39	0.47	0.04	5.0	4.7	2.6
MIROC-ESM	0.7	10	453	0.19	0.35	0.46	2.7	0.9	0.8	0.08	0.23	0.51	0.03	5.8	5.0	2.5
CSIRO-Mk3-6-0	0.9	11	417	0.14	0.15	0.70	2.0	1.2	0.3	0.12	0.38	0.39	0.04	4.0	5.7	2.4
HadGEM2-ES	0.3	9	688	0.10	0.28	0.62	2.8	0.9	0.3	0.09	0.29	0.46	0.08	4.6	6.4	2.7
Ensemble Median	0.7	9	354	0.24	0.32	0.44	1.6	1.4	0.8	0.09	0.27	0.45	0.03	3.9	3.4	2.5
Ensemble 5th percentile	0.2	4	179	0.10	0.15	0.28	0.4	0.8	0.3	0.05	0.17	0.15	0.00	2.9	2.2	1.6
Ensemble 95th percentile	2.6	37	956	0.40	0.48	0.66	4.0	2.8	1.6	0.14	0.40	0.68	0.16	5.9	6.1	5.6

Table 2.2: Median values for parameters in the Bayesian fit to each GCM. Ensemble values at bottom are computed assuming each GCM is equally likely.

3

On the uncertainty of peak warming toward higher cumulative emissions

Abstract

The temperature response to future emissions scenarios is examined with a focus on the uncertainties associated with the more slowly evolving components of the climate system. A Bayesian formalism is used to linearly decompose various response timescales of 24 CMIP5 simulations to CO_2 forcing. Using a similar decomposition of the response of 16 models of the carbon cycle to carbon emissions, permits for examining the probability of various values of peak warming as a function of emissions pathways. Whereas the median peak warming increases by about one degree Celsius per 1000 GtC, the 95th percentile of warming increases at twice the rate. This increasing uncertainty reflects the fact that the transient climate response and carbon cycle control the spread of possible temperatures for cumulative emissions below 1500 GtC, but that uncertainties in the slow physical modes of response are dominant for larger cumulative emissions.

3.1 Introduction

Thermal inertia associated with the large heat capacity of the oceans causes the earth's surface to be out of balance with incoming radiation. Thus, even if atmospheric concentrations of greenhouse gases were kept constant, the earth will still experience significant temperature increases, slowly equilibrating over centennial time scales (Hansen et al., 1985; Baker & Roe, 2009). The long term equilibrium warming assuming fixed atmospheric composition is sometimes termed unrealized or committed warming (Meehl et al., 2005; Wigley, 2005).

Reducing emissions of greenhouse gases to zero will not, however, lead to constant concentrations, as the ocean and land carbon reservoires will slowly uptake a significant

proportion of the emitted greenhouse gasses. This interaction between the time evolution of the forcing agents and the physical response leads to scenarios wherein zeroing out emissions leads to a peak warming that is realized in the transient regime, on the order of decades after cessation of emissions (Matthews & Caldeira, 2008). Indeed, future global warming scenarios are often framed in terms of specific peak warming targets (Stocker, 2013).

Since CO_2 and heat are mixed into the ocean at similar rates, the thermal inertia of the oceans is counteracted to a large degree by the inertia of the carbon cycle (Solomon et al., 2009; Matthews & Weaver, 2010; Matthews & Solomon, 2013). This approximate balancing of the physical and thermal inertias leads to decreased sensitivity of the peak response to the trajectory of the forcing, such that the peak transient response is often discussed in terms of the response to cumulative emissions (Matthews et al., 2012). The relationship between cumulative emissions and peak warming, however, entails substantial uncertainties (Allen et al., 2009)

Past studies of the uncertainty associated with peak warming have used conceptual models (Armour & Roe, 2011; Stocker, 2013; Good et al., 2011; Castruccio et al., 2014) or a limited number of either earth models of intermediate complexity (EMICs), (Allen et al., 2009; Solomon et al., 2009) or full general circulation models (GCMs) (Matthews & Zickfeld, 2012; Arnell et al., 2013). Conceptual models and EMICs have been useful for exploring the qualitative behavior of the system, but may not adequately represent the possible range of response.

GCMs are a more comprehensive representation of our understanding of the climate system, though computational requirements constrain the number and length of scenarios that can be simulated. In Chapter 2, we employed a Bayesian framework to decompose the response of the CMIP5 ensemble to CO_2 forcing, and which we employ

here in order to further examine the uncertainty of the temperature response to a range of emissions scenarios.

3.2 Emission Profiles and Response Functions

We use a basis set of idealized profiles to simulate the wide range of possible emission scenarios. In order to keep the problem tractable, we follow Allen et al. (2009) and parameterize these profiles as a function of two parameters: the year in which decarbonization of the energy system begins in earnest, and the time scale of the transition to zero emissions. Historical concentrations of radiative agents are prescribed following an estimate of historical forcing up until 2011 IPCC (2013, Annex II), after which emissions are assumed to increase at a rate of 1.7% per year (Allen et al., 2009). To represent the transition to zero emissions, the trend toward increasing emissions is multiplied by a half-gaussian applied at the year at which the transition is defined and with a timescale that is set as the standard deviation of the gaussian. The sigmoidal structure of the half-gaussian provides for a smooth transition. Three representative emission profiles are depicted in Fig. 3.1a. These simple emissions scenarios do not provide an exhaustive description of the possible evolution of the concentration of forcing agents, though the linear response function framework that we will derive could be utilized to ascertain the distribution of the response to any given forcing scenario.

In order to reduce the complexity of the basis set we treat all future anthropogenic emissions in terms of their carbon equivalent. Non-CO_2 forcing agents are fixed at their 2011 concentrations, and a value of 3.7 W/m^2 per doubling of CO_2 is assumed. Both aerosols and non CO_2 greenhouse gasses are major contributors to forcing, though they largley cancel one another out. Out of a total best estimate of 2011 forcing of 2.2 W/m^2,

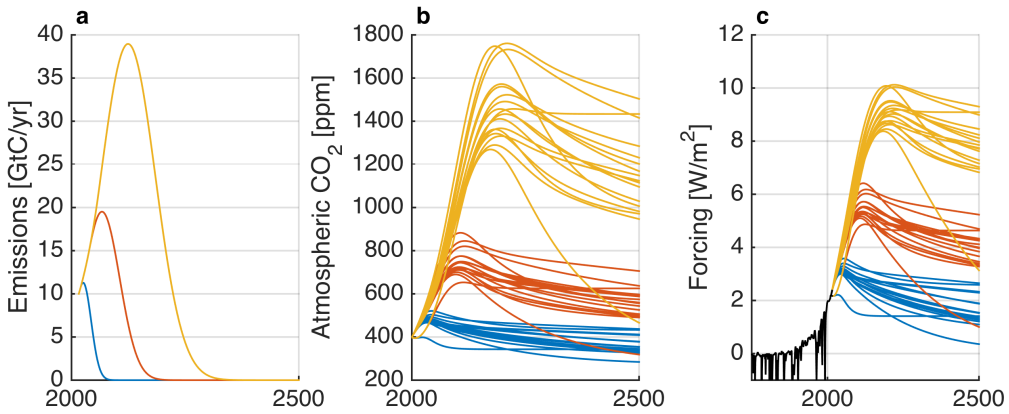

Figure 3.1: Illustrations of forcing profiles used. **a**: Benchmark emission profiles with decarbonization start dates of 2020(blue), 2050(red) and 2080(yellow), and decarbonization time scales of 25, 55, and 85 years respectively. **b**: CO_2 concentrations for each of the three benchmark emission profiles. Each line represeRnts a distinct emission profile convolved with the linear response function of a distinct carbon cycle model. **c:** Radiative forcing associated with concentrations from **b**. Qualitatively, the three emission profiles are similar to RCP 2.6, RCP 4.5, and RCP 6.

1.8 W/m^2 is attributable to CO_2 forcing IPCC (2013). The short life time of aerosols, however, would be expected to lead to increased rates of warming if emissions are ceased abruptly (Armour & Roe, 2011; Matthews & Zickfeld, 2012).

The uptake of CO_2 by the oceans and biopshere is a complex but first-order control on atmospheric concentrations. We follow Joos et al. (2013) and Ricke & Caldeira (2014) in using a simple linear response function to describe the evolution of atmospheric CO_2 concentrations for a given emission pulse. In particular, Joos et al. (2013) provides exponential decay curves for CO_2 concentrations following carbon emissions into an atmosphere with present day concentrations. The exponential curves are fit to 16 earth system models of intermediate and full complexity, and are described by three decay times, representing multi-annual, multi-decadal, and multi-millennial time scales. The resulting CO_2 concentrations and radiative forcing associated with particular emission

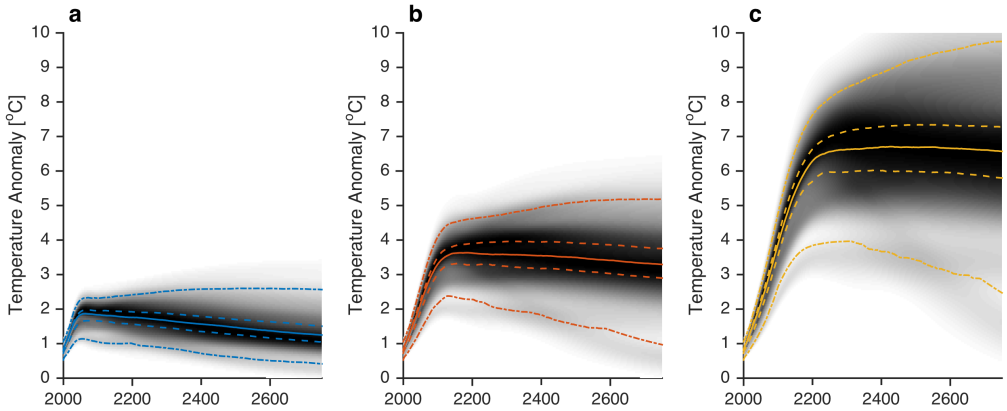

Figure 3.2: Distribution of temperature responses to the three representative emission profiles illustrated in Fig. 3.1. Median(solid lines), 30th and 60th percentiles (dashed lines), and 5th and 95th percentiles (dashed-dot lines) are depicted.

profiles are illustrated in Fig. 3.1b,c.

The global mean temperature response of the climate system to a given forcing time series is modeled using the three time scale linear response function derived in Chapter 2. For each combination of a decarbonization start time and duration, the resulting emission profile is convolved with each of the 16 different carbon cycle response functions to obtain forcing profiles, which are in turn convolved with 240 draws from the posterior distribution of the Bayesian fit to the CMIP5 models consisting of 10 draws for each model. Thus, the resulting distribution of temperature responses for a given emissions profile contains 3840 samples.

A major limitation of using a linear response function is that it assumes the response is independent of the size of the emission pulse, whereas the equilibrium airborne fraction of CO_2 increases with cumulative emissions, leading to superlinear responses for large emissions (Zickfeld & Herrington, 2015). Another major limitation particular to our analysis, along with earlier such analyses Solomon et al. (2009); Ricke & Caldeira

(2014), is that the carbon uptake is assumed independent of the climate response. Our expectation is that greater warming would lead to less ocean carbon uptake, thus creating the potential for even larger peak warming large cumulative emissions scenarios. In this respect, the present analysis can be considered a lower bound on the uncertainty in peak warming.

3.3 Temperature Response Probability

The distribution of possible temperature responses to three representative emission profiles is illustrated in Fig. 3.2. On average each of the three carbon cycle time scales is smaller than the associated physical response time scale. The relatively slower response of the carbon cycle leads to peakedness in the temperature response. For a given CO_2 emission pulse, temperature will increase on the time scale of the physical system up until it is in quasi-equilibrium with the radiative forcing. Afterwards, temperature decreases along with radiative forcing on the time scales of the carbon cycle. Note that in about $\sim 5\%$ of joint response functions have the carbon cycle responding more rapidily than temperature, in which case the temperature peaks before emissions reach zeros.

The time at which peak temperature is achieved relative to peak emissions depends on which time scales are excited. For an RCP2.6-like scenario (Fig. 3.2a) peak temperature is achieved on the order of one decade after peak emissions, in agreement with earlier results (Ricke & Caldeira, 2014). This decadal response reflect a combination of timescale associated with fast temperatures responses coming into equilibrium before decaying along with CO_2 concentrations. For a larger cumulative emissions (Fig. 3.2c), however, temperatures do not start falling until ~ 200 years after peak emissions because the slow mode of temperature response plays a more important role.

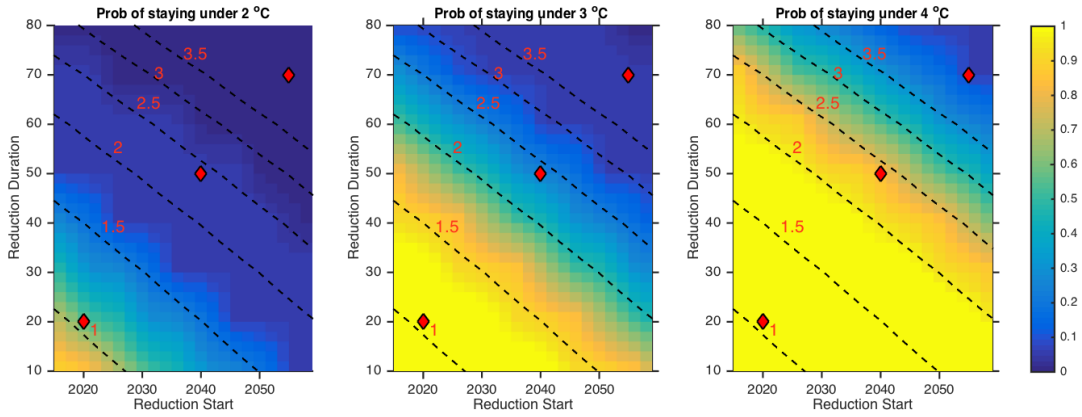

Figure 3.3: Probability of peak temperature reached by 2800 staying bellow 2°C (**a**), 3°C (**b**) and 4°C (**c**) as a function of the emission reduction start year and duration. Diamond markers denote the three representative emission profiles illustrated in Figs. 3.1 and 3.2. Dashed black lines represent lines of constant cumulative emissions, with valued denoted in red in units of 10^{12} tonnes C.

For each emission profile we compute the probability of staying below a given temperature threshold as the fraction of ensemble samples that do not cross the threshold by the end of the integration in 2800 (Fig. 3.3). As expected, peak warming is highly insensitive to emission pathways for a given value of the cumulative emissions, at least across the basis set considered here.

3.4 Contributions to Uncertainties in Peak Warming

Fig. 3.4 shows how the contribution to the 95% of peak temperatures changes as a function of cumulative emissions.

Based on expert judgement, the IPCC quotes 1000 and 1210 GtC of cumulative emissions as leading, respectively, to a 66% or 50% probability of peak temperatures remaining below 2 degrees Celsius of warming (Stocker et al., 2013). Our quantita-

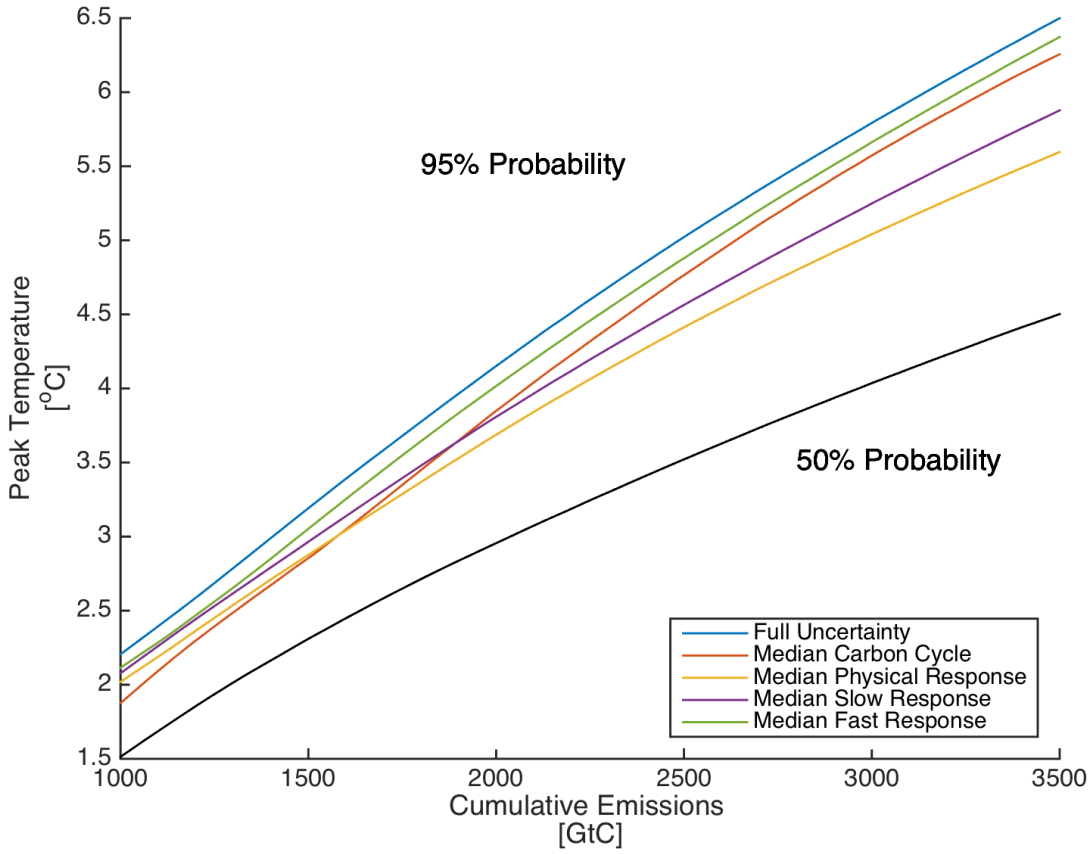

Figure 3.4: Cumulative emissions and peak temperatures at 50% and 95% levels for probability of non-exceedance. Each point represents a different emission profiles. The median response is depicted in black. Colored lines indicate the temperature threshold with a 95% non-exceedance probability if the spread associated with individual components is collapsed unto the respective median value.

tive assessment is similar with these probabilities according with 1100 or 1180 GtC of cumulative emissions, but where uncertainties increase more quickly with cumulative emissions than judged by the IPCC.

We assess the relative contributions of uncertainties in producing peak warming by isolating three different sources associated with the carbon cycle, fast warming responses (eigenmodes 1 and 2), and the slow warming response (eigenmode 3). Through dividing the distribution of temperature responses along these three dimensions and successively collapsing each dimension to its median value, we evaluate the various contributions to uncertainty. Fig. 3.5 shows the evolution of the 5-95% uncertainty range for the three benchmark emission profiles. The relative contribution of uncertainties exhibits strong dependence on the cumulative emissions for the profiles considered. Similarly to the timing of peak temperature, this will be controlled by the processes excited, with slow response modes and the carbon cycle dominating the uncertainty on longer time scales.

The relative contribution of uncertainties exhibits strong dependence on the cumulative emissions for the profiles considered. For cumulative emissions of $\sim$1500 GtC or less, the uncertainty in the response of the carbon cycle is dominant, whereas for greater cumulative emissions, the slow temperature response modes contribute the larger share of uncertainty.

3.5 Conclusions

By jointly considering uncertainties in both the carbon cycle response to carbon emissions and the physical response to a given radiative forcing, we have derived a probabilistic distribution of the temperature response to a variety of future scenarios. The framework presented provides the most comprehensive assessment to date of the emis-

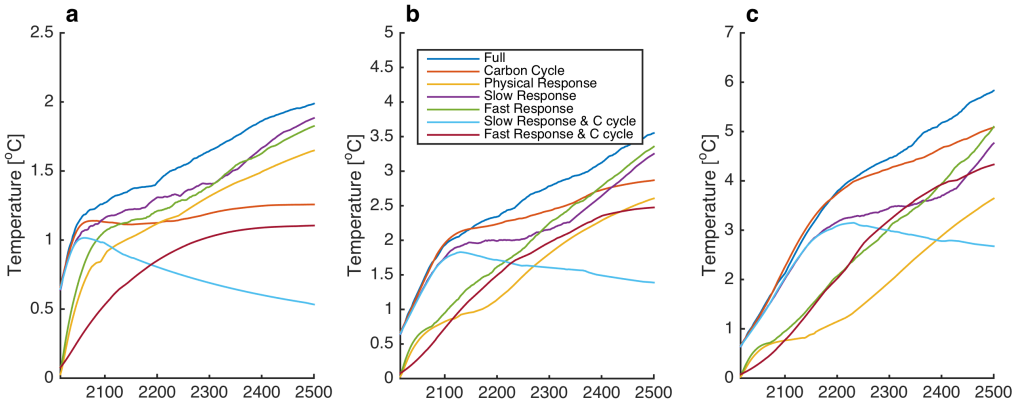

Figure 3.5: Evolution of the 5-95% uncertainty range in temperature projections for the three benchmark scenarios considered in Figs. reffig:profiles. The blue lines indicates the full uncertainty range associated with the temperature projections (Fig. 3.2). Colored lines indicate how the uncertainty ranges change if the distribution of individual components is collapsed unto their median values.

sion budget associated with a given peak temperature threshold and a given probability of exceedance.

It has been suggested that waiting prior to taking measures to abate climate change is reasonable because we will then know more about how the system operates. Indeed, efforts at reducing the uncertainty in the response to cumulative emissions should be directed towards reducing uncertainty in the transient climate response for low temperature thresholds such as 2°C. However, as we necessarily also contemplate the implications of higher thresholds, additional uncertainties associated with more-slowly evolving components of the climate system come to the fore. Insomuch as waiting to abate carbon emission entails the likelihood of greater cumulative emissions, our learning about the transient response will be negated by greater ignorance of peak warming associated with more slowly-evolving responses.

There are a number of shortcoming and potential improvements to the present anal-

ysis. Fundamentally, the model ensemble does not represent the full range of possible uncertainty due to covarying biases (Tebaldi & Knutti, 2007) and missing processes. Lacking knowledge about the error and bias covariance structure, or the relationship between diagnostic and prognostic skill, Knutti et al. (2010) recommends refraining from attempts to further restrain the uncertainty, such as through model weighting approaches. Another significant limitation is the assumption of no covariance between the physical response and the carbon cycle response. The Bayesian posterior distribution of the climate response used here addresses the covariance between the magnitude of the temperature response and its time scale structure (Ricke & Caldeira, 2014). The formalism is not extended to the carbon cycle, however, as the models used to derive the carbon response function (Joos et al., 2013) are different from the CMIP5 models used to derive the physical response. Future work should focus on applying the Bayesian formalism derived in Chapter 2 to ensembles of comprehensive earth system models.

4

Detecting Orbital Variability in Oligo-Miocene Climate Records

Abstract

We address the problem of detecting quasi-periodic variability at orbital frequencies within pre-Pleistocene climate records using depth-derived and orbitally tuned chronologies. Many studies describing orbital variability in pre-Pleistocene sediment hosted isotope records employ climatic records that are set on orbitally tuned chronologies, without accounting for the bias in spectral power estimates introduced by orbital tuning. In this study we develop a method to quantify the effects of tuning upon spectral estimates and, in particular, to more properly determine the statistical significance of spectral peaks associated with orbital frequencies. We apply this method to two marine sediment δ^{18}O records spanning the Oligo-Miocene, from ODP cores 1090 and 1218. We find that using linear age-depth relationships reveals statistically significant spectral peaks matching eccentricity in core 1090, and obliquity and precession in core 1218, where the last appears most significant. Tuning the chronologies to the orbital solutions of Laskar et al. (2004) increases the statistical significance of the precession peak, whereas the obliquity and eccentricity peaks become indistinguishable from those expected from tuning noise. This result can be understood in that tuning records with high signal to noise ratios tends to lead to more significant spectral peaks, whereas a linear age-depth relationship is better suited for detecting peaks when signal to noise ratios are low. We also demonstrate this concept using synthetic records.

4.1 Introduction

The presence of a response to orbital variations, in so much as it can be objectively and confidently identified in paleoclimate records, provides valuable insight into the climate

system. As a response to a known forcing, orbital variability in paleoclimate records is useful in understanding the sensitivity of the climate system to changes in radiative forcing. Additionally, since we are able to compute astronomical configurations and insolation curves over the entire Paleogene, correlation of sediment-hosted signals and orbital curves provides for the possibility of constructing accurate orbital chronologies (Hinnov, 2004; Pälike et al., 2006a; Westerhold et al., 2008), insomuch as the orbital solutions are themselves well constrained (Laskar et al., 2004).

Variability at orbital frequencies, also called Milankovitch frequencies, was first demonstrated in late-Pleistocene marine sediment cores in the obliquity and precession bands (Hays et al., 1976), and, later, in the obliquity band for the early Pleistocene (Ruddiman et al., 1986). The quality and quantity of Pleistocene Proxy records has permitted for cycle by cycle identification of the orbital imprint in climate records, and orbitally tuned chronologies have been developed by matching these signals against astronomical solutions calculated back in time (Lisiecki & Raymo, 2005). These astrochronologies have been shown to be in close agreement with ages obtained from depth-derived age models based on stacked records (Lisiecki & Raymo, 2007; Huybers, 2007), as well as radiometric measurements (Shackleton et al., 1990).

Pre-Pleistocene records also have been shown to contain significant orbital band variability, absent any orbital tuning. Records obtained during Ocean Drilling Program (ODP) Leg 154, on the Ceara Rise (Curry et al., 1995) have been shown to have significant obliquity and eccentricity band variability (Weedon et al., 1997; Zachos et al., 1997). Furthermore Sexton et al. (2011) demonstrated significant variability at precession, obliquity and eccentricity frequencies in an Eocene aged magnetic susceptibility record from ODP site 1258.

However, many records have been evaluated for orbital behavior only after some

amount of tuning is performed. Subsequent work on the Ceara Rise records, for example, used the age model of Shackleton et al. (1999), who calibrated magnetic susceptibility first to an orbital curve consisting mostly of obliquity with a small precession component, and then to the 405 kyr eccentricity. Based on this age model, the spectra for various oxygen and carbon isotope records from sites 926 and 929 were recomputed, (Paul et al., 2000; Zachos et al., 2001; Pälike et al., 2006a), and peaks at orbital frequencies were found to be significant when compared against a red noise background (Mann & Lees, 1996). Although the possibility of excess spectral power resulting from tuning was noted (Paul et al., 2000), the influence of orbital tuning upon such spectral estimates has not been formally quantified.

More recent studies have focused on records from ODP site 1090 on the Agulhas ridge in the south Atlantic (Billups et al., 2002, 2004), and from ODP site 1218 in the equatorial Pacific (Wade & Pälike, 2004; Pälike et al., 2006b). Both site 1090 (Channell et al., 2003) and site 1218 (Lanci et al., 2005) have a complete magnetostratigraphic record, containing all major magnetic reversals within the interval spanned by the record. Age models were obtained by assigning the magnetic reversals ages based on the polarity time scale and then refined by successive tuning of physical properties and stable isotopes, to an orbital curve consisting of eccentricity, obliquity and precession. Significance levels in the form of confidence intervals were computed for the spectral power estimates of tuned records (Pälike et al., 2006b), but again not accounting for the fact that the records have been tuned.

It is useful to distinguish between the problem of constraining the chronology of a record using orbital information and the problem of objectively identifying orbitally driven components in the variability. Here we focus only on the latter, using two approaches to this problem, that we expect to be broadly applicable to other orbital

detection problems in geological records. The first approach is in the spirit of Hays et al. (1976): a purely depth-derived age model is used and the spectrum is evaluated for significant concentrations of energy in the orbital bands. The problem one quickly runs up against is that age uncertainty can easily disperse concentrations of spectral energy present at orbital frequency (Perron & Huybers, 2009). There is an open question whether, for such uncertain records, variability may be more readily detected using limited orbital assumptions when constructing an age model. In our second approach, we explore the use of orbital tuning in detecting orbital variability, through using a Monte Carlo method to account for the bias that tuning introduces in spectral estimates.

4.2 Age Models

We analyze two marine sediment δ^{18}O records from ODP site 1090 (Billups et al., 2002) and ODP site 1218 (Wade & Pälike, 2004) as case studies for testing for Milankovitch signals in pre-Pleistocene climate records. We selected these records for several reasons. First, we required δ^{18}O records with high enough sampling resolution such that variability at orbital frequencies can be spectrally resolved, if present. For the ∼40 kyr obliquity and ∼20 kyr precession periods sampling intervals smaller than ∼20 kyr and ∼10 kyr, respectively, are required by the Nyquist theorem. Given distortions of high frequency terms down to about half the Nyquist frequency due to interpolation (Rhines & Huybers, 2011), even higher sampling resolutions are preferred. Site 1090 spans the late Oligocene to early Miocene (25-16.5 Ma) and is sampled on average every 5.8 kyr. The high resolution subset of site 1218, described in Wade & Pälike (2004), spans 30.0 to 26.5 Ma and has an average sampling rate of 4.3 kyr. Thus, both records have sufficiently high resolutions and are long enough to encompass several cycles of even the longer

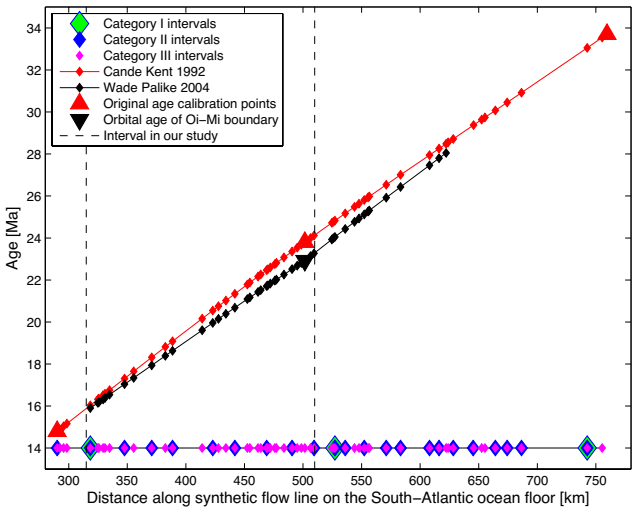

Figure 4.1: Geomagnetic Polarity Time Scale (GPTS). The horizontal line displays the synthetic flow line in the South Atlantic. The average distances between reversals was calculated using a combination of finite rotation poles and stacked magnetic profiles, under the assumption that sea-floor spreading rates varied slowly in time and were normally distributed in space (Cande & Kent, 1992). The large triangles are the radiometric age control points. For comparison, the original GPTS of Cande & Kent (1992) is compared with the orbitally tuned version of Wade & Pälike (2004).

Milankovitch cycles, such as the 405 kyr eccentricity. Second we require cores from which magnetostratigraphic records are available, such that a purely depth-derived age model independent of orbital assumptions can be built based on correlation of magnetic polarity reversals to the geomagnetic polarity time scale. Both sites yielded complete magnetostratigraphic polarity records (Channell et al., 2003; Lanci et al., 2005), and the starting point for our age models is the Cenozoic geomagnetic polarity time scale.

Within the last two decades, the geomagnetic polarity time scale has seen several adjustments and revisions. However, in all its present incarnations it is based on the work of Cande & Kent (1992, 1995) who obtained a composite geomagnetic polarity sequence of normal and reversed magnetic polarity intervals across the South Atlantic

ocean floor and computed the mean relative distances between reversals along an average direction of seafloor motion. A time scale was then obtained using nine absolute age calibration points, assuming quasi-uniform spreading rates in between each pair of calibration points, and performing a cubic spline interpolation against distance on the flow line to obtain a distance-to-age relationship for the magnetic reversals (Fig. 4.1).

In the original study, the calibration points were radiometric ages correlated with the magnetic anomaly sequence through the magnetostratigraphy of marine sediments overlying ocean crust (Cande & Kent, 1992). However, more recent studies (Billups et al., 2004; Wade & Pälike, 2004; Pälike et al., 2006a) use a calibration point at the Oligo-Miocene boundary that was astronomically derived (Shackleton et al., 1999). This age was chosen to maximize the similarity between the Oligo-Miocene climate signal and orbital parameters, making any studies using the astronomically calibrated geomagnetic polarity timescales somewhat biased towards showing an orbital relationship.

Starting from the ages of magnetic reversals in the untuned polarity time scale, we build purely depth-derived age models for the δ^{18}O records. Age calibration points are assigned to the magnetic reversals, with ages linearly interpolated with mean composite depth between these calibration points to obtain a depth-to-age curve for the entire length of the core. We asses the depth-to-age relationship for two versions of the age model, each based on a slightly different geomagnetic polarity time scale (Cande & Kent, 1992; Huestis & Acton, 1997). Both versions are independent of orbital assumptions and differ only in their estimation of the average position of magnetic reversals on the ocean floor.

Quantifying the uncertainties in the age model poses several difficulties. The calibration point at the Oligo-Miocene boundary has a chronogram estimated age of 23.8Ma $\pm$1 Myr (Harland, 1990). Uncertainties in the lengths of the intervals between reversals

on the sea floor profile are also significant (Cande & Kent, 1992). The larger intervals are estimated from fairly small sample sizes, ranging from 5 to 9 profiles, and with relative errors ranging between 4% to 17% at two standard deviations, assuming normality. The smaller sub-intervals often come from one or two measurements, with no estimated uncertainties. Thus, the errors in the timing of each magnetic reversal are a compound of errors in both the age calibration and the interval lengths, and they will accumulate for intervals in between two radiometric age calibration points. These magnetic reversals are in turn used as age calibration points for depth-derived age models, so the errors propagate through several layers of interpolation. Estimating all these compounded errors is beyond the scope of this paper, but we will explore the sensitivity of the results to the selection of age models.

4.3 A Statistical Test for Orbital Signals

4.3.1 Purely Depth-Derived Age Models

Using the age models based on untuned polarity time scales, we perform a statistical test for the presence of quasi-periodic variability at orbital frequencies within Oligo-Miocene δ^{18}O records. We start by using the Cande & Kent (1992) version, since it is the standard used in the literature. Our test is based on searching for statistically significant concentrations of spectral energy at the orbital frequencies. The records are re-interpolated on equally sampled grids, at the average time step, and normalized to zero mean and unit variance. Spectral power is computed using a periodogram estimator.

The statistical significance of spectral peaks is judged against the null hypothesis that the data simply represent random stochastic variability. This approach is similar

to that of Hays et al. (1976), who tested the null hypothesis that the data represent a random signal having a general distribution of variance similar to the one observed in their low-resolution spectrum. We term the entire background variability represented by the broad-band continuum of spectral power as noise in order to distinguish it from potentially orbital signals. In other contexts of course this other variability may be considered as a signal as it includes unforced random climate variability due to internal processes, and possible nonlinear responses to orbital variability. The background variability is modeled here as an autoregressive process of order 1, or AR(1), which results in a spectrum that starts out as red noise at high frequency and tapers to white noise at low frequency, in a manner similar to the spectrum of the proxies. This type of spectrum is typical for a system with finite memory, where random disturbances are dissipated from the system over a finite time scale (Roe, 2009). In discrete time, an AR(1) process is described by,

$$X_t = \phi X_{t-\Delta t} + \epsilon_t, \tag{4.1}$$

where ϕ is the autocorrelation coefficient and ϵ represents a normally distributed random disturbance with mean zero. We fit the AR(1) process to the observed spectrum using the algorithm detailed in Schneider & Neumaier (2001) and Neumaier & Schneider (2001). For ODP 1090 the estimated autocorrelation coefficient is $\phi = 0.73$ and the variance of the ϵ disturbances is 0.47 at a time step of 5.8 kyr. For ODP 1218 the estimated autocorrelation coefficient is $\phi = 0.87$ and the variance of the ϵ disturbances is 0.41 at a time step of 4.3 kyr.

A Monte Carlo approach is then employed to numerically estimate confidence intervals for our hypothesis test. We generate 10,000 realizations of the AR(1) process used

to model the background noise, and from these realizations we estimate the distribution of spectral energy at each frequency in the spectrum of the noise. The 95% confidence interval is set so that, at each frequency, 95% of the realizations of the AR(1) process have energy bellow this level. Thus, if a peak in the spectrum of the δ^{18}O record rises above the confidence interval, then the null hypothesis that the peak is attributable to background noise can be rejected with 95% confidence. We choose this Monte Carlo approach because we are interested in having individual realizations of the null hypothesis that will later allow us to directly quantify the bias that tuning introduces into the background.

Narrowband concentrations of spectral energy are observed at 1/445 kyr^{-1} in the ODP 1090 data (Fig. 4.2) and at 1/23 kyr^{-1} and 1/41 kyr^{-1} in the ODP 1218 data (Fig. 4.3). The 1/445 peak is near, but not exactly at, the Oligo-Miocene eccentricity frequency of 1/405 kyr^{-1} (Laskar et al., 2004), and the slight frequency offset could be a consequence of the age-uncertainties. By the frequency shift theorem (Bracewell, 2000), errors in the estimated duration of the record are inversely proportional to shifts in the frequency, $f_{estimated} = f_{true} \cdot T_{true}/T_{estimated}$. The uncertainty in the total length of the record is at least as large as the 1 Myr (1σ) uncertainty in the age of the radiometric calibration point at the Oligo-Miocene transition that is used in the calibration of both age models (Fig. 4.1). Given a $\pm$1 Myr error in the total length of the record, we would expect a 1/405 kyr^{-1} eccentricity peak to surface anywhere in a frequency band between 1/456 kyr^{-1} and 1/363 kyr^{-1}. The multitude of precession peaks near $\sim$1/20 kyr^{-1} make it difficult to ascertain whether a frequency offset is present, and the obliquity peak is consistent with no stretching or squeezing in ODP 1218.

The 445 kyr peak (ODP 1090, Fig. 4.2a) and both the 41 kyr and 23 kyr peaks (ODP 1218, Fig. 4.3a) are above the 95% confidence interval, but a few issues immediately arise

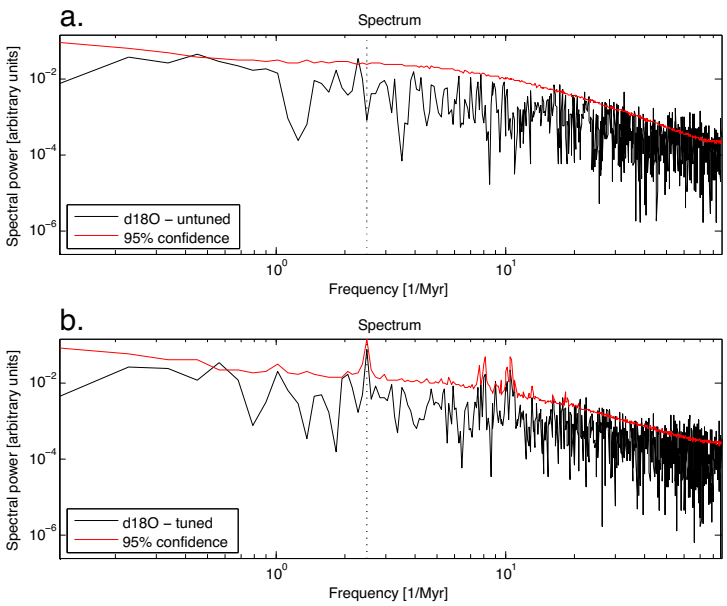

Figure 4.2: Periodogram estimator of spectral power for the δ^{18}O record at ODP site 1090, for both untuned and tuned data. The red lines represents the 95% confidence interval, computed using a Monte Carlo method and assuming an AR(1) background noise. (*a*) Depth Derived age model based on the Cande & Kent (1992) polarity time scale. Significant concentration of spectral energy is found near the $1/405$ kyr^{-1} eccentricity frequency, denoted by the dotted line. (*b*) Orbital age model, with confidence intervals adjusted for the bias introduced through tuning the background noise to eccentricity.

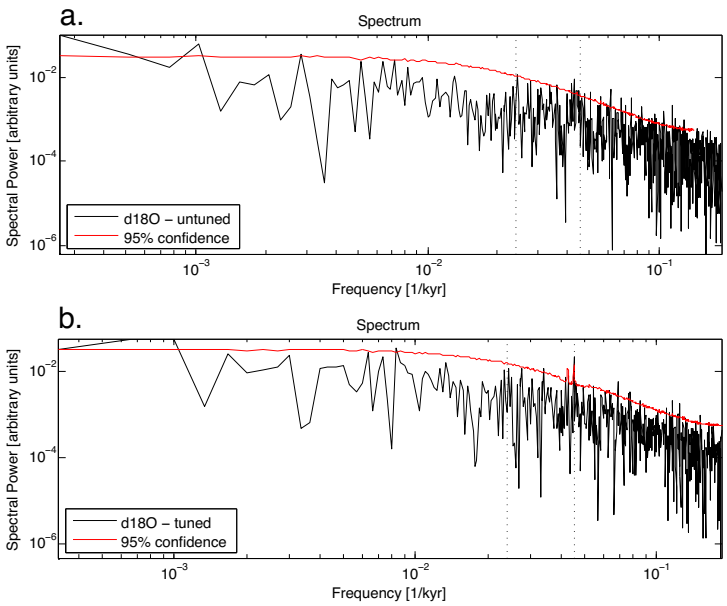

Figure 4.3: Similar to Fig. 4.2, but for ODP 1218. (*a*) Depth Derived age model based on the Cande & Kent (1992) polarity time scale. Significant concentrations of spectral energy is found near the $1/23$ kyr^{-1} precession and $1/41$ kyr^{-1} obliquity frequency, denoted by the dotted lines. (*b*) Orbital age model, with confidence intervals adjusted for the bias introduced through tuning the background noise to precession. Shows that the precession peak is still significant even after tuning.

in interpreting the implications of their significance. The first issue is that not all the peaks are exactly at the orbital frequency, which introduces a subtlety in the formulation of the null hypothesis. Confidence intervals are used to estimate the probability that a peak *at a particular frequency* is attributable to background noise. However, since we know that the peak might be shifted due to errors in the total interval length, we have searched for peaks in a rather broad frequency band, and there is a larger probability that background noise will produce at least one peak which is spuriously significant within this frequency interval as opposed to at a single frequency. For the 1/405 kyr^{-1} eccentricity peak, for example, we estimated that timing uncertainties could shift the peak to anywhere in the 1/456 kyr^{-1} and 1/363 kyr^{-1} interval. Thus, in our discrete spectra we have effectively extended our search to six different frequency bands. Reformulating the null hypothesis to account for the fact that we are searching for a peak amongst multiple frequencies, the significance levels of the 1/405 kyr^{-1} eccentricity peak and the 1/41 kyr^{-1} decrease to just under the 95% critical value, but the 1/23 kyr^{-1} precession remains significant at 95% confidence. Furthermore, the significance of the 445 kyr peak is sensitive to the choice of the age model. When we use an age model based on the slightly different polarity time scale of Huestis & Acton (1997), the eccentricity band has no significant concentration of power, at any candidate frequency. The fact that the results of the statistical test are sensitive to the specific formulation of the null hypothesis and the choice of age models warrants the exploration of a different approach.

4.3.2 Tuned Age Models

We now asses the efficacy of using orbital tuning to test for orbital variability, which is interesting to explore two reasons. First, it may be that tuning allows for better

identification of orbital behavior. Second, many other studies have tuned their records and then attempted to identify statistically significant peaks in the resulting spectral estimates, and we would like to better understand the effects of the tuning process.

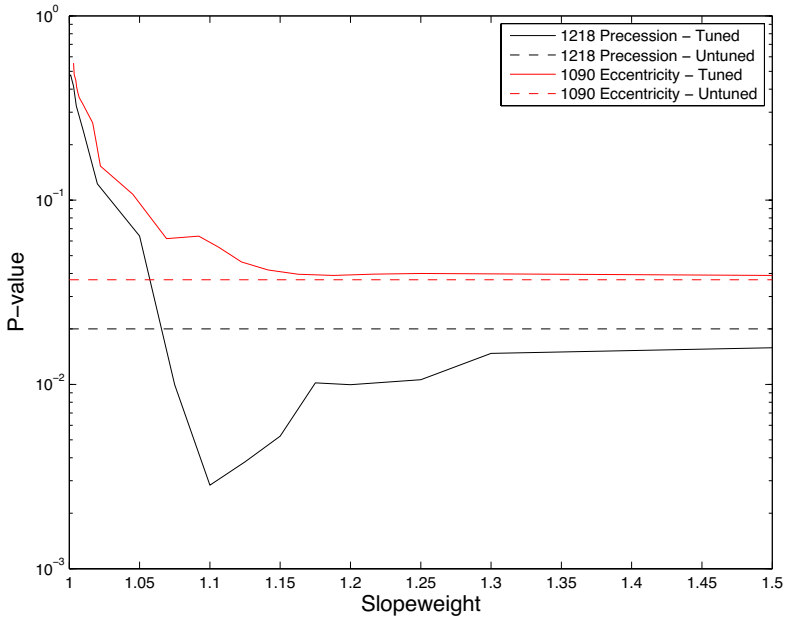

Figure 4.4: P-values for the spectral peaks at the 1/405 kyr^{-1} eccentricity frequency in ODP 1090 and 1/22 kyr^{-1} precession frequency in ODP 1218. The P-value of a spectral peak is equal to one minus its significance, and indicates the probability that a peak drawn from a sample of the null hypothesis distribution has spectral power equal or greater than the observed peak. The slope-weight of the tuning algorithm is a measure of the allowable age model adjustment. The higher the slope-weight, the smaller the adjustments that the tuning algorithm is allowed to make. As the slope-weight increases, the P-values will asymptote to the untuned case.

Similar to previous studies, we start by assigning the records a linear age to depth relationship based on the polarity time scale, and then use an objective tuning algorithm to match the δ^{18}O record to each orbital curve in Laskar et al. (2004). Tuning is accomplished using a dynamical time warping algorithm Lisiecki & Lisiecki (2002) that chooses a depth to age relation for the proxy records that minimizes the sum of

squared differences between the proxy and the orbital curve. The amount of age model adjustment the tuning algorithm is allowed to make can be controlled by introducing a paramater called the slope-weight, which penalizes deviations from the original age model (Pälike, 2002). The higher the slope-weight, the less tuning can deviate from the accumulation rate implied by the linear age-depth relationship. Not surprisingly, after tuning, the periodograms of the δ^{18}O records show sharp peaks at the orbital frequencies (Fig. 4.2b, Fig. 4.3b). However, this result does not necessarily mean the peak is statistically significant. A record comprised solely of stochastic variability will exhibit sharp peaks in the spectral domain if subjected to the same tuning algorithm. Most realizations of the AR(1) process used to model the background noise spectrum of the δ^{18}O record, when tuned, exhibit sharp peaks that appear significant when contrasted to the spectral estimate of the untuned noise. Thus, revised critical values and 95% confidence intervals need to be estimated that account for the effects of tuning.

The revised null hypothesis is that peaks in a spectral estimate are due to broadband stochastic variability that has been tuned erroneously. Revised confidence intervals are then estimated via a Monte Carlo approach by taking realizations of the AR(1) process used to model background noise and subjecting them to the same tuning algorithm as the δ^{18}O record (Fig. 4.2b, Fig. 4.3b).

The behavior and results of this statistical test depends on the amount of adjustment the tuning algorithm is allowed to impose. With more adjustment (i.e., a lower slope-weight parameter) the orbital peak in the record becomes stronger, but so too becomes the probability of overtuning, or spuriously injecting a large amount of spectral power from the background variability (Muller & MacDonald, 2002). An optimal configuration of the tuning algorithm would ideally minimize the amount of spectral power that can be injected into a narrow band from the stochastic background continuum and, at the

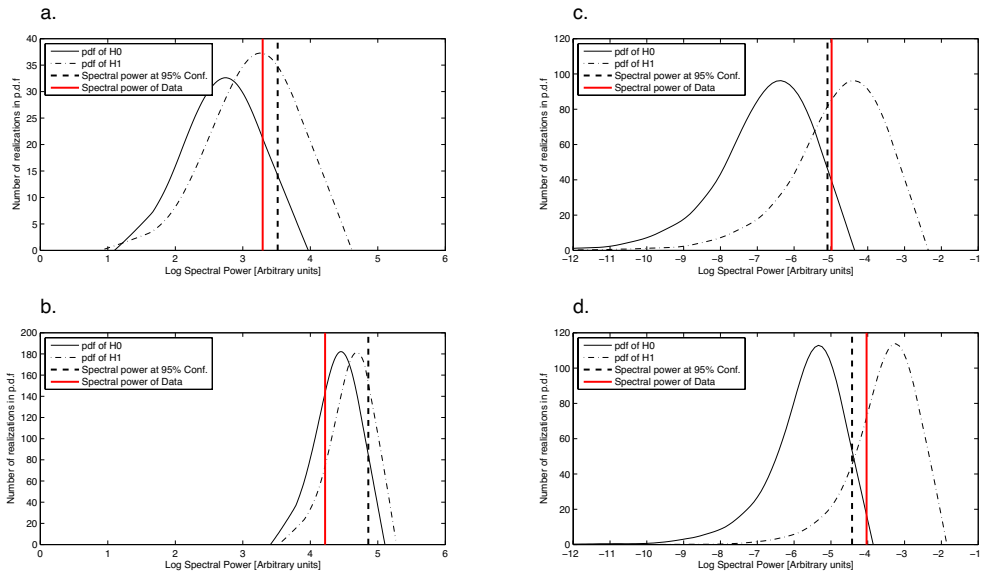

Figure 4.5: Hypothesis test for the presence of orbital variability: Spectral power distribution of H_0 (solid), H_1 (dashed) and δ^{18}O (red) . The parameters for the alternate hypothesis are a 0.5 signal-to-noise ratio and a 0.25 jitter. (*a*) ODP 1090, eccentricity, untuned H_0, H_1, δ^{18}O; (*b*) ODP 1090, eccentricity, tuned H_0, H_1, δ^{18}O; (*c*) ODP 1218, precession, untuned H_0, H_1, δ^{18}O; (*d*) ODP 1218, precession, tuned H_0, H_1, δ^{18}O;

same time, maximize the amount of spectral power recovered from a distorted periodic signal.

For each orbital curve we tune to, we sweep across the full range of possible slope-weights and examine the p-value of the orbital peaks. Large values for the slope weight result in minimal tuning, while smaller values of the slope weight result in greater tuning. The p-value of a spectral peak is a measure of how high the tuned orbital peaks rise above the adjusted confidence intervals. By definition, it will represent the probability that a realization of the null hypothesis, tuned to the same orbital curve, will have equal or more spectral power than the tuned peak in the δ^{18}O record. As the amount of age model adjustment is minimized, the p-value of the peaks asymptote to the untuned

values (Fig. 4.4). We perform the tuning exercise for each of the three main orbital parameters, for both the ODP 1090 and ODP 1218 records. For the orbital frequencies where there was no significant concentration of spectral power using the untuned age model, no significant spectral peaks appear after tuning. The 1/405 kyr^{-1} and the 1/41 kyr^{-1} peaks have a similar behavior, in that for large to moderate amounts of tuning the spectral peaks are not significant. Apparently, the tuning algorithm injects more spurious power from the background noise than is recovered from the distorted orbital signal, assuming an orbital signal is present.

A different behavior is observed for the 1/23 kyr^{-1} peak where, for a moderate amount of tuning, the significance of the peak increases dramatically. The robustness of the results for precession across testing configuration lead us to confidently conclude that significant precession band variability is present in the ODP 1218 $\delta^{18}O$ record. This result is, to our knowledge, the first unbiased statistical test for orbital variability using orbitally tuned records.

4.3.3 Statistical Power of the Test

The foregoing method permits for a test of whether orbital variability is present using orbital tuning. However, the results of tuning are uneven, with the strongest results becoming more significant and others less so. This pattern may reflect the nature of the true signal or the characteristics of the test. To explore these possibilities, we turn to analyzing synthetic records to determine under what circumstances a tuned or untuned approach to testing for orbital variability is most powerful. The power of the test is a statistical measure of the separation of the null and alternate hypotheses. Specifically, it is the probability of rejecting the null hypothesis, H_0, when the alternate hypothesis, H_1, is true, and should not be confused with p, which is the probability that H_0 is

wrong. As we will show, statistical power is also a useful metric for assessing if the significance of a spectral peak increases or decreases with tuning.

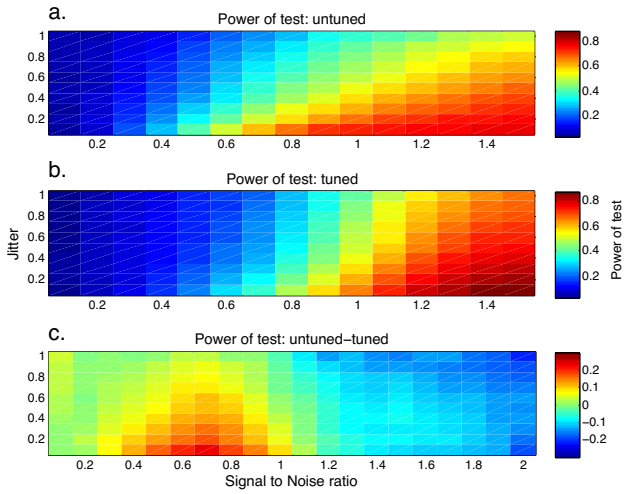

Figure 4.6: The power of the test, i.e one minus the probability of erroneously failing to reject the null.(*a*) for a test based on a purely depth-derived age model. (*b*) for a test based on an orbitally tuned age model. (*c*) The difference in statistical power between the two tests (untuned minus tuned). As expected, statistical power increases as the signal-to-noise ratio increases and as the amount of jitter decreases. The test based on a purely depth-derived age model performs better for signal-to-noise ratios bellow about one, while the test based on an orbitally tuned age model performs better for signal-to-noise ratios above one.

The same null hypothesis, H_0, and associated AR(1) process described in section 3.1

are used here. For the alternate hypothesis, H_1, the record is modeled as a sum of AR(1) noise and an eccentricity curve (Laskar et al., 2004), placed on a corrupted age model. The parameters of H_1 are the signal-to-noise ratio, defined as the ratio of the variance of the orbital curve to the total variance of the AR(1) noise, and the jitter, defined as the variance growth rate of the age model uncertainty (Huybers & Wunsch, 2004). From Monte Carlo ensembles, we can estimate the probability density functions of spectral power in the orbital bands for H_0 and H_1. The power of the test is then directly computed as one minus the percentage of ensemble members of the H_1 that fall under the 95% confidence intervals derived from H_0 (Fig. 4.5a). When tuning is performed on realizations of H_0 and H_1, additional power is injected into the orbital bands, so both the tuned ensembles of H_0 and H_1 will have more spectral power. If the relative gain for H_0 is larger, then the power of the test decreases, and vice-versa (Fig. 4.5b).

The relative gain in spectral power of H_1 will depend on the signal-to-noise ratio and the amount of age model corruption. We perform a parameter space exploration over a reasonable range of jitter for the sediment accumulation rate (Huybers & Wunsch, 2004), and the signal to noise ratios. Thus, for each of the two tests, statistical power is computed as a function of the signal-to-noise ratio and the amount of jitter (Fig. 4.6). We only examine what happens to an orbital signal with the 405 kyr eccentricity period, with an interval duration equal to the duration of the ODP 1090 record. For higher frequency signals, the power of the test will decrease more rapidly with jitter, as higher frequencies are more easily corrupted by age model uncertainties (Huybers & Wunsch, 2004; Perron & Huybers, 2009). As expected, there is a trade off between the signal-to-noise ratio and jitter: the power of the test increases as the signal-to-noise ratio increases or as the degree of jitter decreases.

The test for orbital forcing that uses a purely depth-derived age model has a higher statistical power if the signal to noise ratio is lower than about one, while the test based on an orbitally tuned age model has a higher statistical power for signal-to-noise ratios higher than about one, at least assuming a moderate degree of jitter. The suggestion is that orbital tuning can be used to sharpen periodic signals that are strong relative to the background variability, but tests based on orbitally tuned chronologies perform poorly in detecting signals in a predominantly noisy record, as evidenced by the increase in overlap between the distributions of the null and alternate hypothesis (Fig. 4.5). The behavior of the synthetic records is consistent with the results obtained for ODP 1090 and ODP 1218 where the only peak that increased in significance with adjusted confidence intervals, from a p-value of 0.02 to a p-value of 0.003, was the precession peak, which already had the lowest p-value of all three orbital peaks.

4.4 Conclusions

We have examined two different spectral domain tests for the presence of quasi-periodic orbital forcing in paleoclimate records inherently dominated by a broad-band distribution of variance, using two δ^{18}O records from the Oligocene and Miocene as case studies. The first test is based on a purely depth-derived age model, devoid of orbital forcing assumptions and employing a linear age-to-depth relation between magnetostratigraphic calibration points, while the second test is based on records that are orbitally tuned to orbital curves. Without accounting for the spectral energy introduced in the orbital bands from tuning the background, the null hypothesis would be erroneously rejected at the 95% significance level. The need to reformulate the null hypothesis and adjust confidence intervals when using astrochronologies is a major caveat that no previous

study appears to have quantitatively accounted for. We have developed a Monte Carlo type methodology for estimating the statistical significance of spectral peaks in tuned records, and used it to perform an unbiased test for orbital variability in pre-Pleistocene climate records set on orbitally tuned chronologies. This methodology should be generally applicable to other orbitally tuned records.

Using untuned age models, significant spectral peaks were detected at eccentricity (ODP 1090) obliquity (ODP 1218) and precession (ODP 1218) frequencies. The use of orbitally tuned age models decreased the significance levels of the weaker eccentricity and obliquity peaks but increased the significance level of the stronger precession peak. Experiments on synthetic records are consistent with these results and suggest that tests using untuned age models are better at detecting periodic variability in noisier records, with a signal-to-noise ratio lower then about one, while orbital tuning may improve the test for distorted records that nonetheless have strong orbital components. Because the signal-to-noise ratio is difficult to estimate apriori, the use of both types of tests appears warranted.

5

Discussion and Future Challenges

We have analyzed and solved several problems pertaining to the temporal structure of climate variability and response. In Chapter 1 we analyzed the time scale structure of non-normality in atmospheric temperature time series. Using the relationship between higher order spectra and higher order moments of a distribution, we have shown how the later decay under filtering. This demonstration provides a quantification of a well known empirical result in signal processing, extending previous abstract proofs to finite observational time series. More importantly, we have reconciled observations of non-normality at low frequencies with mechanisms acting at the highest observed frequencies. Additionally, we have developed a non-normality test that accounts for auto correlation in the data.

In Chapter 2 we analyzed the time scale structure of the mean temperature response. We have showed that within the CMIP5 ensemble, the value of the net radiative feedbacks is larger along slower modes of the response function. These modes, however, are relatively weakly excited by the structure of historical forcing, and therefore estimates of climate sensitivity are biased towards the faster modes of response possessing weaker feedbacks. Accounting for this bias reconciles estimates of sensitivity drawn from historical records with those obtained from climate models.

In Chapter 3 we used the probabilistic distribution of the response function of temperature to radiative forcing derived in Chapter 2, and convolved it with a similar response function for atmospheric CO_2 concentrations in response to carbon emissions. This allowed us to compute the evolution of the global temperature response to any given future emission scenario. Using a representative basis set of emission profiles we computed the cumulative carbon emission budget required in order to achieve a given probability of staying below a set temperature threshold. Finally, we estimated the sensitivity of the uncertainty in the temperature response to reducing uncertainty in

the response of different components.

In Chapter 4 we developed a rigorous hypothesis test for the presence of orbital variability in records that have undergone orbital tuning. We generally find that orbital tuning increases the significance of periodic component that already have relatively large signal to noise ratio relative to the background continuum, whereas it can actually decreases the significance of weak orbital components by injecting large amounts of spurious spectral power from the background continuum.

A pressing problem highlighted by the uncertainty analysis presented in Chapter 3 is the need to better constrain the slow modes of temperature response to radiative forcing. As shown in Chapter 2, however, these modes are only very weakly expressed in global mean temperature responses. A more detailed analysis of the joint spatio-temporal patterns of response will be required as part of future efforts. Several problems arise, however, in analyzing local response patterns. Estimation of the equilibrium global response is predicated upon the condition that linear response in the top of the atmosphere imbalance reaches asymptotes to zero. This definition of equilibrium does not hold for the local response in the presence of changing atmospheric and oceanic heat transports. One possible solution is computing a full joint distribution of the local temperature and radiative responses under the constraint that the global mean radiative imbalance goes to zero. Preliminary attempts have proven computationally prohibitive.

An approximation for the local responses can be obtained under the assumption that local temperatures evolve along the same eigenmodes derived from the global response. For each abrupt quadrupling experiment in the CMIP5 a multiple linear regression is performed between local temperature and radiative imbalance, and decaying exponential functions with time scales set to the posterior mode of the bayesian inference for the global response function, thus obtaining the local response along each mode. This

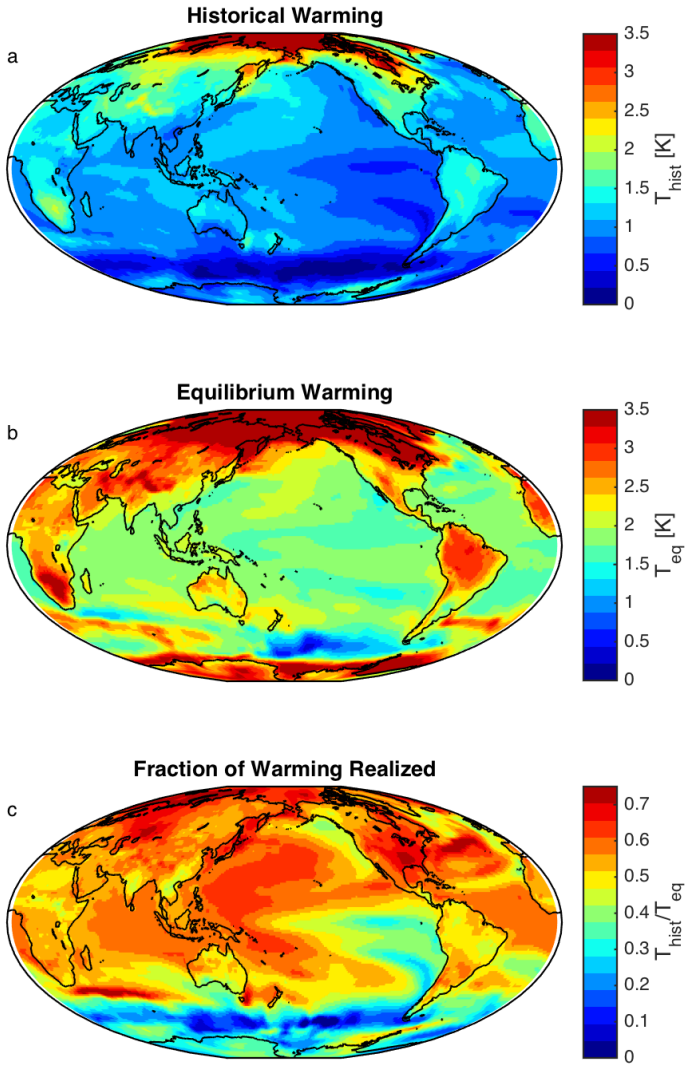

Figure 5.1: Ensemble mean spatial warming patterns obtained by locally projecting the three eigenmode decomposition of the CMIP5 global response through a multiple linear regression. **a**: Historical pattern of warming under AR5 historical forcing. **b**: Equilibrium warming pattern. **c**: Fraction of local equilibrium temperature realized under historical forcing.

approach allows for an estimation of an approximate ensemble mean response function, although it does not permit for a straightforward estimation of the associated uncertainty. As a check on the approximation, a global mean of local responses in top of the atmosphere radiative imbalance is computed, and is found to asymptote to values within $\sim$0.2 W/m^2 for each GCM. The local response functions are then convolved with the AR5 estimate of historical forcing up to 2011, to obtain an approximation for patterns of local temperature anomalies relative to pre-industrial times (Fig. 5.1a). This is contrasted with the equilibrium patterns of warming associated with a 2011 value of effective global radiative forcing of 2.2 W/m^2, to obtain the local relative fraction of equilibrium warming (Fig. 5.1b,c). Some of the observed features are to be expected. The interior of northern continents exhibit a large degree of relative equilibration with the forcing, whereas regions associated with strong oceanic upwelling, such as the souther ocean, the east tropical Pacific and the north Atlantic exhibit the lowest degree of equilibration.

The relatively weak historical temperature changes and large internal variability expected in radiatively important regions such as the southern ocean and the east Pacific will complicate any attempt to infer the local response from observational records. Palaeoclimate proxies may provide better constraints on the longer response time scales. In particular, the background continuum of variability may prove useful in constraining intermediate time scales situated between those resolved by observational records and the orbital bands associated with higher signal to noise ratio. Fig. 5.2 illustrates preliminary estimates of the background radiative forcing relative to previous estimates of the background temperature spectra for low latitude sea surface temperatures (Huybers & Curry, 2006).

Time domain linear response functions such as those used in Chapters 2 have an

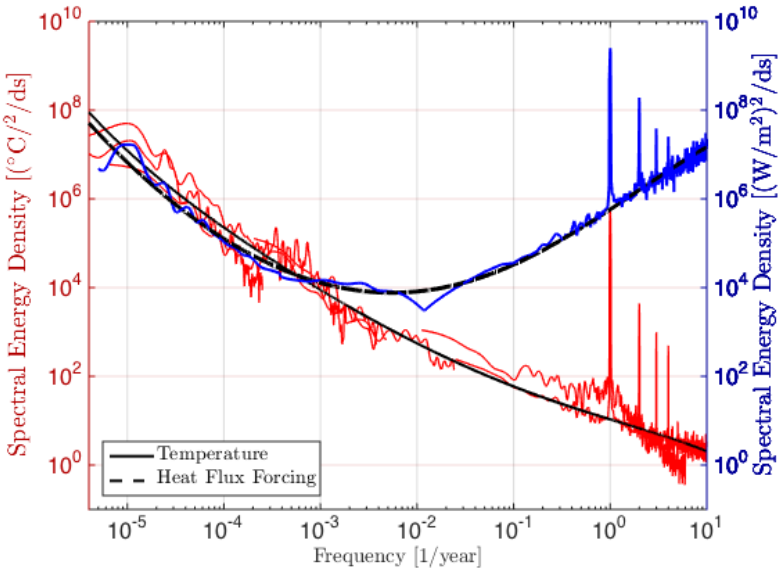

Figure 5.2: Patchwork estimates of the variance spectra for the heat flux forcing of the low-latitude surface ocean and the temperature response. Records used include HadCRU instrumental estimates, reanalysis data and paleoclimate proxies consisting of Mg/Ca, Sr/Ca, Alkenones,^{10}Be and ice core bubble greenhouse gas measurements.

analogy in the spectral domain in the form of transfer functions. The square root of the spectra illustrated in Fig. 5.2 can be used to constrain such a transfer function (MacMynowski et al., 2011). A preliminary estimate of the resulting gain function is consistent with the likely range of the IPCC estimate of equilibrium sensitivity on multi-centennial time scales, but continues to increase before tapering off on the time scales associated with glacial cycles. The gain functions derived from these records, however, represent local response functions. Complex relationships can exist between local and global spectra, particularly at sub-centennial time scales (Laepple & Huybers, 2014; Rypdal et al., 2015). Thus significant challenges still persist in drawing inferences on the global response.

The work presented in the body of the thesis draws inferences about the structure of climate variability and response primarily by analyzing individual time series. The spatial patterns, however, are expected to also change as a function of time scale. Better models for the joint spatio-temporal structure in both historical records and paleoclimate proxies are needed before the temperature response to anthropogenic emissions can be better constrained empirically. Past observations in both long term historical records and paleoclimate records are sparse. The primary challenge going forward appropriate framework will be required to obtain the probable distributions of global responses conditional on the sparse local observations. Additionally, a sufficiently good model for the temporal and spatial autocorrelation structure of local internal variability is required in order to separate the forced response.

References

Allen, M. R., Frame, D. J., Huntingford, C., Jones, C. D., Lowe, J. A., Meinshausen, M., & Meinshausen, N. (2009). Warming caused by cumulative carbon emissions towards the trillionth tonne. *Nature*, 458(7242), 1163–1166.

Andrews, T., Gregory, J. M., & Webb, M. J. (2015). The dependence of radiative forcing and feedback on evolving patterns of surface temperature change in climate models. *J. Climate*, 28, 1630–1648.

Andrews, T., Gregory, J. M., Webb, M. J., & Taylor, K. E. (2012). Forcing, feedbacks and climate sensitivity in CMIP5 coupled atmosphere-ocean climate models. *Geophys. Res. Lett.*, 39, 1–7.

Armour, K. C. (2016). Projection and prediction: Climate sensitivity on the rise. *Nature Climate Change*, 6, 896–897.

Armour, K. C., Bitz, C. M., & Roe, G. H. (2013). Time-Varying Climate Sensitivity from Regional Feedbacks. *J. Climate*, 26, 4518–4534.

Armour, K. C. & Roe, G. H. (2011). Climate commitment in an uncertain world. *Geophysical Research Letters*, 38(1).

Arnell, N. W., Lowe, J., Brown, S., Gosling, S., Gottschalk, P., Hinkel, J., Lloyd-Hughes, B., Nicholls, R., Osborn, T., Osborne, T. M., et al. (2013). A global assessment of the effects of climate policy on the impacts of climate change. *Nature Climate Change*, 3(5), 512–519.

Baker, M. B. & Roe, G. H. (2009). The shape of things to come: Why is climate change so predictable? *J. Climate*, 22, 4574–4589.

Billups, K., Channell, J. E. T., & Zachos, J. (2002). Late Oligocene to early Miocene geochronology and paleoceanography from the subantarctic South Atlantic. *Paleoceanography*, 17(1), 010000–1.

Billups, K., Pälike, H., Channell, J. E. T., Zachos, J. C., & Shackleton, N. J. (2004). Astronomic calibration of the late Oligocene through early Miocene geomagnetic polarity time scale. *Earth and Planetary Science Letters*, 224, 33–44.

Bitz, C. M., Shell, K. M., Gent, P. R., Bailey, D. a., Danabasoglu, G., Armour, K. C., Holland, M. M., & Kiehl, J. T. (2012). Climate sensitivity of the community climate system model, version 4. *Journal of Climate*, 25(9), 3053–3070.

Bracewell, R. (2000). The fourier transform & its applications 3rd Ed.

Brillinger, D. R. (1965). An introduction to polyspectra. *The Annals of mathematical statistics*, (pp. 1351–1374).

Büeler, B., Enge, A., & Fukuda, K. (2000). Exact volume computation for polytopes: a practical study. In *Polytopes—combinatorics and computation* (pp. 131–154).: Springer.

Caldeira, K. & Myhrvold, N. P. (2013). Projections of the pace of warming following an abrupt increase in atmospheric carbon dioxide concentration. *Environ. Res. Lett.*, 8, 034–039.

Cande, S. C. & Kent, D. V. (1992). A new geomagnetic polarity time scale for the late Cretaceous and Cenozoic. *Journal of Geophysical Research*, 97, 13917–+.

Cande, S. C. & Kent, D. V. (1995). Revised calibration of the geomagnetic polarity timescale for the Late Cretaceous and Cenozoic. *Journal of Geophysical Research*, 100, 6093–6095.

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M. A., Guo, J., Li, P., & Riddell, A. (2016). STAN: A probabilistic programming language. *J Stat Softw*, VV, 1–43.

Castruccio, S., McInerney, D. J., Stein, M. L., Liu Crouch, F., Jacob, R. L., & Moyer, E. J. (2014). Statistical emulation of climate model projections based on precomputed gcm runs*. *Journal of Climate*, 27(5), 1829–1844.

Cavanaugh, N. R. & Shen, S. S. (2014). Northern hemisphere climatology and trends of statistical moments documented from ghcn-daily surface air temperature station data from 1950 to 2010. *Journal of Climate*, 27(14), 5396–5410.

Cavanaugh, N. R. & Shen, S. S. (2015). The effects of gridding algorithms on the statistical moments and their trends of daily surface air temperature*. *Journal of Climate*, 28(23), 9188–9205.

Channell, J., Galeotti, S., Martin, E., Billups, K., Scher, H., & Stoner, J. (2003). Eocene to Miocene magnetostratigraphy, biostratigraphy, and chemostratigraphy at ODP Site 1090 (sub-Antarctic South Atlantic). *Geological Society of America Bulletin*, 115(5), 607.

Chu, C. R., Parlange, M. B., Katul, G. G., & Albertson, J. D. (1996). Probability density functions of turbulent velocity and temperature in the atmospheric surface layer. *Water resources research*, 32(6), 1681–1688.

Curry, W. B., Shackleton, N. J., Richter, C., & et, a. (1995). Proc. ODP, Initial Reports, 154: College Station, TX (Ocean Drilling Program).

Director, H. & Bornn, L. (2015). Connecting point-level and gridded moments in the analysis of climate data*. *Journal of Climate*, 28(9), 3496–3510.

Donohoe, A. & Battisti, D. S. (2009). The amplitude asymmetry between synoptic cyclones and anticyclones: Implications for filtering methods in feature tracking. *Monthly Weather Review*, 137(11), 3874–3887.

Durre, I., Vose, R. S., & Wuertz, D. B. (2006). Overview of the integrated global radiosonde archive. *Journal of Climate*, 19(1), 53–68.

Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S., Collins, W., Cox, P., Driouech, F., Emori, S., Eyring, V., Forest, C., Gleckler, P., Guilyardi, E., Jakob, C., Kattsov, V., Reason, C., & Rummukainen, M. (2013). *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change: Evaluation of Climate Models*, chapter 9, (pp. 741– 866). Cambridge University Press: Cambridge, United Kingdom and New York, NY, USA.

Forster, P. M. (2016). Inference of climate sensitivity from analysis of earth's energy budget. *Annual Review of Earth and Planetary Sciences*, 44(1), 85–106.

Forster, P. M., Andrews, T., Good, P., Gregory, J. M., Jackson, L. S., & Zelinka, M. (2013). Evaluating adjusted forcing and model spread for historical and future scenarios in the CMIP5 generation of climate models. *J. Geophys. Res.: Atmos.*, 118, 1139–1150.

Frisch, U. & Sornette, D. (1997). Extreme deviations and applications. *Journal de Physique I*, 7(9), 1155–1171.

Garth, L. M. & Bresler, Y. (1997). The degradation of higher order spectral detection using narrowband processing. *Signal Processing, IEEE Transactions on*, 45(7), 1770–1784.

Geoffroy, O., Saint-martin, D., Olivié, D. J. L., Voldoire, a., Bellon, G., & Tytéca, S. (2013). Transient climate response in a two-layer energy-balance model. Part I: Analytical solution and parameter calibration using CMIP5 AOGCM experiments. *J. Climate*, 26, 1841–1857.

Good, P., Gregory, J. M., & Lowe, J. A. (2011). A step-response simple climate model to reconstruct and interpret aogcm projections. *Geophysical Research Letters*, 38(1).

Greenberg, M. D. (1988). *Advanced engineering mathematics*. Prentice-Hall.

Gregory, J. & Andrews, T. (2016). Variation in climate sensitivity and feedback parameters during the historical period. *Geophysical Research Letters*, 43(8), 3911–3920.

Gregory, J. M., Stouffer, R. J., Raper, S. C. B., Stott, P. A., & Rayner, N. A. (2002). An observationally based estimate of the climate sensitivity. *Journal of Climate*, 15(22), 3117–3121.

Hannachi, A. (2010). On the origin of planetary-scale extratropical winter circulation regimes. *Journal of the Atmospheric Sciences*, 67(5), 1382–1401.

Hansen, J., Russell, G., Lacis, A., Fung, I., Rind, D., & Stone, P. (1985). Climate response times: dependence on climate sensitivity and ocean mixing. *Science*, 229, 857–859.

Hansen, J., Sato, M., Ruedy, R., Nazarenko, L., Lacis, a., Schmidt, G. a., Russell, G., Aleinov, I., Bauer, M., Bauer, S., Bell, N., Cairns, B., Canuto, V., Chandler, M., Cheng, Y., Del Genio, a., Faluvegi, G., Fleming, E., Friend, a., Hall, T., Jackman, C., Kelley, M., Kiang, N., Koch, D., Lean, J., Lerner, J., Lo, K., Menon, S., Miller, R., Minnis, P., Novakov, T., Oinas, V., Perlwitz, J., Perlwitz, J., Rind, D., Romanou, a., Shindell, D., Stone, P., Sun, S., Tausnev, N., Thresher, D., Wielicki, B., Wong, T., Yao, M., & Zhang, S. (2005). Efficacy of climate forcings. *Journal of Geophysical Research D: Atmospheres*, 110(18), 1–45.

Hansen, J., Sato, M., Russell, G., & Kharecha, P. (2013). Climate sensitivity, sea level and atmospheric carbon dioxide. *Phil. Trans. R. Soc. A*, 371(2001), 20120294.

Harland, W. (1990). *A geologic time scale 1989*. Cambridge Univ Press.

Hasselmann, K. (1976). Stochastic climate models Part I. Theory. *Tellus*, 28(6), 473–485.

Hays, J. D., Imbrie, J., & Shackleton, N. J. (1976). Variations in the Earth's Orbit: Pacemaker of the Ice Ages. *Science*, 194, 1121–1132.

Held, I. M., Winton, M., Takahashi, K., Delworth, T., Zeng, F., & Vallis, G. K. (2010). Probing the fast and slow components of global warming by returning abruptly to preindustrial forcing. *J. Climate*, 23, 2418–2427.

Hinnov, L. A. (2004). Earth's orbital parameters and cycle stratigraphy. In J. Ogg, A. Smith, & F. Gradstein (Eds.), *A Geologic Time Scale 2004,* (pp. 55–62).

Hoffman, M. D. & Gelman, A. (2014). The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 15(1), 1593–1623.

Huestis, S. P. & Acton, G. D. (1997). On the construction of geomagnetic timescales from non-prejudicial treatment of magnetic anomaly data from multiple ridges. *Geophysical Journal International*, 129, 176–182.

Huybers, P. (2007). Glacial variability over the last two million years: an extended depth-derived agemodel, continuous obliquity pacing, and the pleistocene progression. *Quaternary Science Reviews*, 26(1-2), 37–55.

Huybers, P. & Curry, W. (2006). Links between annual, Milankovitch and continuum temperature variability. *Nature*, 441, 329–332.

Huybers, P., McKinnon, K. A., Rhines, A., & Tingley, M. (2014). U.s. daily temperatures: The meaning of extremes in the context of nonnormality. *Journal of Climate*, 27(19), 7368–7384.

Huybers, P. & Wunsch, C. (2004). A depth-derived Pleistocene age model: Uncertainty estimates, sedimentation variability, and nonlinear climate change. *Paleoceanography*, 19, 1028–+.

IPCC (2013). *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press.

Joos, F., Roth, R., Fuglestvedt, J., Peters, G., Enting, I., Bloh, W. v., Brovkin, V., Burke, E., Eby, M., Edwards, N., et al. (2013). Carbon dioxide and climate impulse response functions for the computation of greenhouse gas metrics: a multi-model analysis. *Atmospheric Chemistry and Physics*, 13(5), 2793–2825.

Knutti, R., Furrer, R., Tebaldi, C., Cermak, J., & Meehl, G. A. (2010). Challenges in combining projections from multiple climate models. *Journal of Climate*, 23(10), 2739–2758.

Kotulski, Z. & Sobczyk, K. (1981). Linear systems and normality. *Journal of Statistical Physics*, 24(2), 359–373.

Kummer, J. R. & Dessler, A. E. (2014). The impact of forcing efficacy on the equilibrium climate sensitivity. *Geophysical Research Letters*, 41(10), 3565–3568.

Laepple, T. & Huybers, P. (2014). Global and regional variability in marine surface temperatures. *Geophysical Research Letters*, 41(7), 2528–2534.

Lanci, L., Parés, J., Channell, J., & Kent, D. (2005). Oligocene magnetostratigraphy from equatorial Pacific sediments (ODP Sites 1218 and 1219, Leg 199). *Earth and Planetary Science Letters*, 237(3-4), 617–634.

Laskar, J., Robutel, P., Joutel, F., Gastineau, M., Correia, A., & Levrard, B. (2004). A long-term numerical solution for the insolation quantities of the Earth. *Astronomy and Astrophysics*, 428(1), 261–285.

Lewis, N. & Curry, J. A. (2014). The implications for climate sensitivity of AR5 forcing and heat uptake estimates. *Climate Dynamics*, 45(3-4), 1009–1023.

Lilliefors, H. W. (1967). On the kolmogorov-smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association*, 62(318), 399–402.

Lisiecki, L. & Lisiecki, P. (2002). Application of dynamic programming to the correlation of paleoclimate records. *Paleoceanography*, 17(4), 1049.

Lisiecki, L. & Raymo, M. (2007). Plio-pleistocene climate evolution: trends and transitions in glacial cycle dynamics. *Quaternary Science Reviews*, 26(1-2), 56–69.

Lisiecki, L. E. & Raymo, M. E. (2005). A Pliocene-Pleistocene stack of 57 globally distributed benthic δ^{18}O records. *Paleoceanography*, 20(26), A261003+.

Luxford, F. & Woollings, T. (2012). A simple kinematic source of skewness in atmospheric flow fields. *Journal of the Atmospheric Sciences*, 69(2), 578–590.

MacMynowski, D. G., Shin, H.-J., & Caldeira, K. (2011). The frequency response of temperature and precipitation in a climate model. *Geophys. Res. Lett.*, 38, 1–5.

Mallows, C. (1967). Linear processes are nearly gaussian. *Journal of Applied Probability*, (pp. 313–329).

Mann, M. & Lees, J. (1996). Robust estimation of background noise and signal detection in climatic time series. *Climatic Change*, 33(3), 409–445.

Marvel, K., Schmidt, G. A., Miller, R. L., & Nazarenko, L. S. (2016). Implications for climate sensitivity from the response to individual forcings. *Nature Climate Change*, 6(4), 386–389.

Masters, T. (2014). Observational estimate of climate sensitivity from changes in the rate of ocean heat uptake and comparison to cmip5 models. *Climate Dynamics*, 42(7-8), 2173–2181.

Matthews, H. D. & Caldeira, K. (2008). Stabilizing climate requires near-zero emissions. *Geophysical research letters*, 35(4).

Matthews, H. D. & Solomon, S. (2013). Irreversible does not mean unavoidable. *Science*, 340(6131), 438–439.

Matthews, H. D., Solomon, S., & Pierrehumbert, R. (2012). Cumulative carbon as a policy framework for achieving climate stabilization. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 370(1974), 4365–4379.

Matthews, H. D. & Weaver, A. J. (2010). Committed climate warming. *Nature Geoscience*, 3(3), 142–143.

Matthews, H. D. & Zickfeld, K. (2012). Climate response to zeroed emissions of greenhouse gases and aerosols. *Nature Climate Change*, 2(5), 338–341.

Meehl, G. A., Washington, W. M., Collins, W. D., Arblaster, J. M., Hu, A., Buja, L. E., Strand, W. G., & Teng, H. (2005). How much more global warming and sea level rise? *Science*, 307(5716), 1769–1772.

Muller, R. & MacDonald, G. (2002). *Ice ages and astronomical causes: data, spectral analysis and mechanisms*. Springer Verlag.

Murphy, J. (1995). Transient response of the hadley centre coupled ocean-atmosphere model to increasing carbon dioxide. part iii: analysis of global-mean response using simple models. *Journal of Climate*, 8(3), 496–514.

Myhre, G., Shindell, D., Bréon, F.-M., Collins, W., Fuglestvedt, J., Huang, J., Koch, D., Lamarque, J.-F., Lee, D., Mendoza, B., Nakajima, T., Robock, A., Stephens, G., Takemura, T., & Zhang, H. (2013). *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change: Anthropogenic and Natural Radiative Forcing*, chapter 8, (pp. 659– 740). Cambridge University Press: Cambridge, United Kingdom and New York, NY, USA.

Neumaier, A. & Schneider, T. (2001). Estimation of parameters and eigenmodes of multivariate autoregressive models. *ACM Transactions on Mathematical Software (TOMS)*, 27(1), 57.

Otto, A., Otto, F. E., Boucher, O., Church, J., Hegerl, G., Forster, P. M., Gillett, N. P., Gregory, J., Johnson, G. C., Knutti, R., et al. (2013). Energy budget constraints on climate response. *Nature Geoscience*, 6(6), 415–416.

Pälike, H. (2002). *Extending the geological calibration of the geological time scale*. PhD thesis, University of Cambridge.

Pälike, H., Frazier, J., & Zachos, J. C. (2006a). Extended orbitally forced palaeoclimatic records from the equatorial Atlantic Ceara Rise. *Quaternary Science Reviews*, 25, 3138–3149.

Pälike, H., Norris, R. D., Herrle, J. O., Wilson, P. A., Coxall, H. K., Lear, C. H., Shackleton, N. J., Tripati, A. K., & Wade, B. S. (2006b). The Heartbeat of the Oligocene Climate System. *Science*, 314, 1894–.

Papoulis, A. (1972). Narrow-band systems and gaussianity. *Information Theory, IEEE Transactions on*, 18(1), 20–27.

Paul, H. A., Zachos, J. C., Flower, B. P., & Tripati, A. (2000). Orbitally induced climate and geochemical variability across the Oligocene/Miocene boundary. *Paleoceanography*, 15, 471–485.

Pearson, K. (1916). Mathematical contributions to the theory of evolution. xix. second supplement to a memoir on skew variation. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, (pp. 429–457).

Perron, J. & Huybers, P. (2009). Is there an orbital signal in the polar layered deposits on Mars? *Geology*, 37(2), 155.

Perron, M. & Sura, P. (2013). Climatology of non-gaussian atmospheric statistics. *Journal of Climate*, 26(3), 1063–1083.

Proistosescu, C., Rhines, A., & Huybers, P. (2016). Identification and interpretation of nonnormality in atmospheric time series. *Geophysical Research Letters*, 43, 5425–5434.

Rennert, K. J. & Wallace, J. M. (2009). Cross-frequency coupling, skewness, and blocking in the northern hemisphere winter circulation. *Journal of Climate*, 22(21), 5650–5666.

Rhines, A. & Huybers, P. (2011). Estimation of spectral power laws in time uncertain series of data with application to the Greenland Ice Sheet Project 2 δ18O record. *Journal of Geophysical Research*, 116(D1), D01103.

Richardson, M., Cowtan, K., Hawkins, E., & Stolpe, M. B. (2016). Reconciled climate response estimates from climate models and the energy budget of earth. *Nature Climate Change*, 6, 931–935.

Ricke, K. L. & Caldeira, K. (2014). Maximum warming occurs about one decade after a carbon dioxide emission. *Environmental Research Letters*, 9(12), 124002.

Roe, G. (2009). Feedbacks, timescales, and seeing red. *Annual Review of Earth and Planetary Sciences*, 37, 93–115.

Roe, G. H. & Armour, K. C. (2011). How sensitive is climate sensitivity? *Geophysical Research Letters*, 38(14), 1–5.

Roe, G. H. & Baker, M. B. (2007). Why is climate sensitivity so unpredictable? *Science*, 318(5850), 629–632.

Rohatgi, V. K. & Székely, G. J. (1989). Sharp inequalities between skewness and kurtosis. *Statistics & probability letters*, 8(4), 297–299.

Rohling, E., Sluijs, a., Dijkstra, H., Köhler, P., van de Wal, R., a.S. von der Heydt, Beerling, D., Berger, a., Bijl, P., Crucifix, M., DeConto, R., Drijfhout, S., Fedorov, a., Foster, G., Ganopolski, a., Hansen, J., Hönisch, B., Hooghiemstra, H., Huber, M., Huybers, P., Knutti, R., Lea, D., Lourens, L., Lunt, D., Masson-Demotte, V., Medina-Elizalde, M., Otto-Bliesner, B., Pagani, M., Pälike, H., Renssen, H., Royer, D., Siddall, M., Valdes, P., Zachos, J., & Zeebe, R. (2012). Making sense of palaeoclimate sensitivity. *Nature*, 491, 683–691.

Rose, B. E. J., Armour, K. C., Battisti, D. S., Feldl, N., & Koll, D. D. B. (2014). The dependence of transient climate sensitivity and radiative feedbacks on the spatial pattern of ocean heat uptake. *Geophysical Research Letters*, 41(3), 1071–1078.

Rosenblatt, M. (1961). Some comments on narrow band-pass filters. *Quart. Appl. Math*, 18, 387–393.

Rozanov, Y. A. (1961). On the applicability of the central limit theorem to stationary processes which have passed through a linear filter. *Theory of Probability & Its Applications*, 6(3), 321–322.

Ruddiman, W. F., Raymo, M., & McIntyre, A. (1986). Matuyama 41,000-year cycles: North Atlantic Ocean and northern hemisphere ice sheets. *Earth and Planetary Science Letters*, 80, 117–129.

Ruff, T. W. & Neelin, J. D. (2012). Long tails in regional surface temperature probability distributions with implications for extremes under global warming. *Geophysical Research Letters*, 39(4).

Rugenstein, M. A., Gregory, J. M., Schaller, N., Sedláček, J., & Knutti, R. (2016). Multi-annual ocean-atmosphere adjustments to radiative forcing. *Journal of Climate*, 29(2016), 5643–5659.

Rypdal, K., Rypdal, M., & Fredriksen, H.-B. (2015). Spatiotemporal long-range persistence in earth?s temperature field: Analysis of stochastic–diffusive energy balance models. *Journal of Climate*, 28(21), 8379–8395.

Sardeshmukh, P. D. & Sura, P. (2009). Reconciling non-gaussian climate statistics with linear dynamics. *Journal of Climate*, 22(5), 1193–1207.

Sattin, F., Agostini, M., Cavazzana, R., Serianni, G., Scarin, P., & Vianello, N. (2009). About the parabolic relation existing between the skewness and the kurtosis in time series of experimental data. *Physica Scripta*, 79(4), 045006.

Schneider, T., Bischoff, T., & Płotka, H. (2015). Physics of changes in synoptic mid-latitude temperature variability. *J. Climate*, 28(6), 2312–2331.

Schneider, T. & Neumaier, A. (2001). Algorithm 808: ARfit—A Matlab package for the estimation of parameters and eigenmodes of multivariate autoregressive models. *ACM Transactions on Mathematical Software (TOMS)*, 27(1), 58–65.

Senior, C. A. & Mitchell, J. F. B. (2000). The time dependence of climate sensitivity. *Geophys. Res. Lett.*, 27, 2685–2688.

Sexton, P., Norris, R., Wilson, P., Pälike, H., Westerhold, T., Röhl, U., Bolton, C., & Gibbs, S. (2011). Eocene global warming events driven by ventilation of oceanic dissolved organic carbon. *Nature*, 471(7338), 349–352.

Shackleton, N., Berger, A., & Peltier, W. (1990). An alternative astronomical calibration of the lower Pleistocene timescale based on ODP Site 677. *Trans. R. Soc. Edinburgh Earth Sci*, 81, 251–261.

Shackleton, N. J., Crowhurst, S. J., Weedon, G. P., & Laskar, J. (1999). Astronomical calibration of Oligocene-Miocene time. *Royal Society of London Philosophical Transactions Series A*, 357, 1907–+.

Smirnov, N. V. (1939). Estimate of deviation between empirical distribution functions in two independent samples. *Bulletin Moscow University*, 2(2), 3–16.

Solomon, S., Plattner, G.-K., Knutti, R., & Friedlingstein, P. (2009). Irreversible climate change due to carbon dioxide emissions. *Proceedings of the national academy of sciences*, (pp. pnas–0812721106).

Stocker, T., Qin, D., Plattner, G.-K., Alexander, L., Allen, S., Bindoff, N., Bréon, F.-M., Church, J., Cubasch, U., Emori, S., Forster, P., Friedlingstein, P., Gillett, N., Gregory, J., Hartmann, D., Jansen, E., Kirtman, B., Knutti, R., KrishnaKumar, K., Lemke, P., Marotzke, J., Masson-Delmotte, V., Meehl, G., Mokhov, I., Piao, S., Ramaswamy, V., Randall, D., Rhein, M., Rojas, M., Sabine, C., Shindell, D., Talley, L., Vaughan, D., & Xie, S.-P. (2013). *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change: Technical Summary*, chapter TS, (pp.

33–115). Cambridge University Press: Cambridge, United Kingdom and New York, NY, USA.

Stocker, T. F. (2013). The closing door of climate targets. *Science*, 339(6117), 280–282.

Sura, P. & Hannachi, A. (2015). Perspectives of non-gaussianity in atmospheric synoptic and low-frequency variability. *Journal of Climate*, 28(13), 5091–5114.

Sura, P., Newman, M., Penland, C., & Sardeshmukh, P. (2005). Multiplicative noise and non-gaussianity: A paradigm for atmospheric regimes? *Journal of the atmospheric sciences*, 62(5), 1391–1409.

Sura, P. & Perron, M. (2010). Extreme events and the general circulation: Observations and stochastic model dynamics. *Journal of the Atmospheric Sciences*, 67(9), 2785–2804.

Sura, P. & Sardeshmukh, P. D. (2008). A global view of non-gaussian sst variability. *Journal of Physical Oceanography*, 38(3), 639–647.

Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, 93(4), 485–498.

Tebaldi, C. & Knutti, R. (2007). The use of the multi-model ensemble in probabilistic climate projections. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 365(1857), 2053–2075.

Von Schuckmann, K., Palmer, M., Trenberth, K., Cazenave, A., Chambers, D., Champollion, N., Hansen, J., Josey, S., Loeb, N., Mathieu, P.-P., et al. (2016). An imperative to monitor earth's energy imbalance. *Nature Climate Change*, 6(2), 138–144.

Wade, B. S. & Pälike, H. (2004). Oligocene climate dynamics. *Paleoceanography*, 19(26), A264019+.

Weedon, G., Shackleton, N., & Pearson, P. (1997). 5. The Oligocene time scale and cyclostratigraphy on the Ceara Rise, western equatorial Atlantic. In *Proceedings of the Ocean Drilling Program. Scientific Results*, volume 154 (pp. 101–114).

Westerhold, T., Röhl, U., Raffi, I., Fornaciari, E., Monechi, S., Reale, V., Bowles, J., & Evans, H. (2008). Astronomical calibration of the paleocene time. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 257(4), 377–403.

Wigley, T. M. (2005). The climate change commitment. *Science*, 307(5716), 1766–1769.

Winton, M., Takahashi, K., & Held, I. M. (2010). Importance of ocean heat uptake efficacy to transient climate change. *J. Climate*, 23, 2333–2344.

Zachos, J. C., Flower, B. P., & Paul, H. (1997). Orbitally paced climate oscillations across the Oligocene/Miocene boundary. *Nature*, 388, 567–570.

Zachos, J. C., Shackleton, N. J., Revenaugh, J. S., Pälike, H., & Flower, B. P. (2001). Climate Response to Orbital Forcing Across the Oligocene-Miocene Boundary. *Science*, 292, 274–278.

Zickfeld, K. & Herrington, T. (2015). The time lag between a carbon dioxide emission and maximum warming increases with the size of the emission. *Environmental Research Letters*, 10(3), 031001.