Conditional Simulation of Future Climate Under Changing Temporal Covariance Structures

W.B. Leeds, M.L. Stein, and E.J. Moyer

August 15, 2013

Abstract

1 Introduction

A wide body of evidence suggests that our climate is currently changing (Hansen et al., 2012), with at least part of the change attributed to increasing greenhouse gas (GHG) emissions (Lashof and Ahuja, 1990; Berliner et al., 2000; Qin et al., 2007). As a result, important public policy debates involve discussions of both mitigation and adaptation. These conversations require some understanding of future climate under different potential emissions trajectories, and leading to increased use of general circulation models (GCMs) for discussions in the public sphere. GCMs provide a physically-based model of the earth’s atmosphere under a given emissions scenarios, but are biased relative to observations (see Figure 1). However, GCMs may provide useful information about changes we might expect in the coming years (e.g., the “derivatives”), even if they do not necessarily provide accurate forecasts in an absolute sense (e.g., they do not give a reliable estimate for average temperature in the year 2070, or presently, but they may give valuable information as to the way temperature is changing between these two periods). The temperature and precipitation output from a GCM model can be used in a variety of impacts models to give a glimpse at the future state of various environmental processes for a given emissions scenario.
Impacts models (e.g., agricultural, hydrological) play an important role in this public policy debate. For example, agricultural impacts models provide a glimpse of future food production under various GHG emissions scenarios. These impacts models require inputs of temperature and precipitation at fine spatial and temporal resolution (typically daily). While GCMs provide temperature and precipitation output at high temporal resolution for various GHG scenarios, often, impacts models require a higher spatial resolution than is provided by the GCM (which are typically run at a resolution greater than 100 kilometers). As a result, spatial downscaling is an important area of research in impacts modeling.

All downscaling methodologies can be placed into one of two broad classes. First, dynamic downscaling is a deterministic approach where a regional climate model (RCM) is nested within a GCM (i.e., the GCM provides lateral and possibly vertical boundary conditions). The RCM is processed over a small region and at finer spatial resolution (often on the order of 50 kilometers) than is computationally realistic over the entire globe. RCMs therefore better resolve geography (a clear shortcoming of coarse spatial resolution GCMs), but have their own limitations (Pielke and Wilby, 2012; Kerr, 2013). In addition, even though RCMs are run over a smaller region, the fine-scale spatial resolution requires considerable computational effort, beyond what was already needed to run the GCM. Statistical downscaling of climate models, on the other hand, provides a computationally cheap alternative to simulating future climate (Wilby et al., 1998; Zorita and Von Storch, 1999). These methods are typically based on the estimation of a transfer function between large-scale atmospheric processes and local, observed climate, and evaluating this transfer function for GCM output for future emissions scenarios. Because model-observation discrepancy is also accommodated in statistical downscaling, the overall procedure is often referred to as bias-correction/statistical downscaling.

The Bias Correction Spatial Disaggregation (BCSD) method is perhaps the most commonly used approach (Wood et al., 2002, 2004). Here, the model-observation discrepancy
transfer function is a quantile-mapping procedure, with the most extreme case being a full transformation of the empirical CDF. This transfer function is then evaluated for future GCM output. This approach is appealing primarily because it considers entire distributional corrections, not simply mean corrections, as found in other methods.

Despite the ability to account for entire distributional corrections, it is our contention that the BCSD approach has two substantial drawbacks. First, because the approach relies on correcting marginal distributions, the dependence structure in the process of interest is ignored. Also, an approach based on the correction of marginal distributions is not aggregation consistent. To clarify, consider the problem of generating monthly downscaled temperature given daily observations and model output. One approach would be to perform the quantile-mapping procedure on the daily observations and model output, then aggregating the result to generate downscaled temperature at a monthly resolution. A second approach would be to aggregate the observations and model output to a monthly resolution first and then to perform the quantile-mapping on the monthly time series. These two approaches need not produce equivalent results, which may be problematic given the need to produce downscaled climate variables at different temporal resolutions.

Second, as with most bias-correction approaches, the general principle behind BCSD is to modify future GCM-output based on a transfer function that represents model-observation discrepancy. Due to the limitations of climate models, we contend that it is advantageous to approach the problem from what has at times been referred to as a “change-factor method” (see, e.g., Diaz-Nieto and Wilby, 2005). Such approaches modify the observations based on suggested changes in GCM output. The typical change-factor approach is known as the Delta method in the bias-correction/statistical downscaling literature (for details see, e.g., Eisner et al., 2012, and references therein). This approach involves shifting the mean of the climatology of the observations based on suggested changes from the GCM seasonal cycle. Let \( \{Z_{0,t}; t = 0, \ldots, T - 1\} \) represent observations of daily
mean temperature at a single spatial location. Suppose we wish to generate simulations of this variable at a future time period. Let \( \{\hat{\mu}_{1,t+\tau}; t = 0, \ldots, T-1\} \) represent the seasonal cycle estimated from the output of a GCM under a given emissions scenario, and let \( \{\hat{\mu}_{0,t}; t = 0, \ldots, T-1\} \) be the corresponding estimated GCM seasonal cycle under a baseline run (e.g., pre-industrial forcings). Then, using the Delta method, we shift the observational record to produce the future simulations

\[
Z_{1,t+\tau} = Z_{0,t} + (\hat{\mu}_{1,t+\tau} - \hat{\mu}_{0,t}).
\]

While quite simple, this approach is conceptually appealing for several reasons. Most importantly, the approach involves modifying the data based only on changes in the model output, rather than modifying future model output based on a transfer function used to estimate model-observation discrepancies. As a result, when the model suggests no mean shifts under a given emissions scenario, the conditional simulations are equivalent to the observations. Thus, all statistical properties of the observations (e.g., nonstationarity, nonlinearity, tail behavior, etc.) are preserved for the conditional simulations.

While the Delta method approach does provide useful insight to how one might approach generating simulations of future climate, it is limiting in the sense that it only accounts for changes in the mean, not the (spatial, temporal, or spatio-temporal) dependence structure of the process. Recent research suggests that future climate may not only involve an increase in mean temperatures, but it might involve an increase (or decrease) in overall “variability” as well (Karl et al., 1995; Timmermann et al., 1999; Schär et al., 2004). Importantly, this variability change may involve changes in dependence across different spatial and temporal scales. This temperature variability may have an impact on agricultural yield projections, for example, even if if it occurred in the absence of changes in mean temperatures (Wheeler et al., 2000). A framework that modifies the observations in order to account for changing dependence structures could provide a valuable approach to simulating future climate.

Modifying changing dependence structures (spatial, temporal, or spatio-temporal) across
multiple spatial and temporal scales is challenging. To achieve this goal, it helps to think about modifying quantities that are independent (or close to independent). In this regard, spectral-based approaches provide a natural framework. We propose an approach for modifying the discrete Fourier transform (DFT) of observations based on an estimated ratio of spectral distributions of model output. This approach shares important qualities with the Delta method, namely that we are modifying the observations based on changes suggested by the model output and, when the model suggests no changes (in either first or second-order moment characteristics), the conditional simulations are equivalent to the observations.

The remainder of the paper is as follows. Section 2 outlines the methodology, explaining how to modify the observations using a ratio of spectral densities, and how to estimate the given ratio. It also explains a potential approach to account for temporal nonstationarity in the data, as brought about by differences in variability across seasons. Section 3 provides an application to generating conditional simulations of daily mean temperature using the Community Climate System Model version 3 (CCSM3) and NCEP Climate Forecast System Reanalysis (CFSR) used as observations. We end the paper with a discussion of the results and future work in Section 4.

2 Methods

2.1 Motivation

Let \( \{Z_{0,t}; t = 0, \pm 1, \pm 2, \ldots\} \) represent a time series of an observable process of interest. Further, suppose \( \{Z_{0,t}\} \) is a stationary Gaussian process with \( \mathbb{E}(Z_{0,t}) = 0 \) and autocovariance function \( \gamma_0(h) = \text{Cov}(Z_{0,t}, Z_{0,t-h}) = \mathbb{E}(Z_{0,t}Z_{0,t-h}) \). Let \( \{Z_{1,t}; t = 0, \pm 1, \pm 2, \ldots\} \) represent the future process that we wish to simulate. Suppose that \( \{Z_{1,t}\} \) is also a stationary Gaussian process with \( \mathbb{E}(Z_{1,t}) = 0 \), but with autocovariance function \( \gamma_1(h) = \text{Cov}(Z_{1,t}, Z_{1,t-h}) = \ldots \)
\[ \mathbb{E}(Z_{1,t}Z_{1,t-h}), \] where \( \gamma_1(h) \) is not necessarily the same as \( \gamma_0(h) \). Thus, \( \{Z_{1,t}\} \) has a possibly different temporal dependence structure than \( \{Z_{0,t}\} \).

We are interested in modifying \( \{Z_{0,t}\} \) in order to generate a random process that is equal in distribution to \( \{Z_{1,t}\} \). The correlated temporal dependence of \( \{Z_{0,t}\} \) makes this non-trivial. However, the orthogonal nature of spectral representation makes it a natural framework. For example, \( Z_{0,t} \) can be written
\[
Z_{0,t} = \int_{-0.5}^{0.5} \exp(2\pi i \omega t) d\hat{Z}_0(\omega)
\]
where \( \hat{Z}_0(\omega) \) is a complex-valued random measure with mean zero and for disjoint sets \([a, b] \cap [c, d] = \emptyset\), \( \mathbb{E}\left(\hat{Z}_0([a, b])\overline{\hat{Z}_0([c, d])}\right) = 0 \), where \( \overline{x} \) represents the complex conjugate of \( x \). That is, \( \hat{Z}_0(\omega) \) (also referred to as the spectral process associated with \( \{Z_{0,t}\} \)) is an orthogonal Gaussian measure which can be modified separately at each frequency to produce a process with a different dependence structure after transforming back to the time domain.

Consider the spectral distributions associated with \( \{Z_{0,t}\} \) and \( \{Z_{1,t}\} \). Specifically, the spectral distribution associated with \( \{Z_{0,t}\} \), \( G_0(\omega) \), can be derived from \( \hat{Z}_0(\omega) \), \( \mathbb{E}\left[|d\hat{Z}_0(\omega)|^2\right] = dG_0(\omega) \), where \( G_0(\omega) \) is a positive measure. Then, assuming absolute summability of the autocovariance, i.e., \( \sum_{h=-\infty}^{\infty} |\gamma_0(h)| < \infty \), the spectral distribution is absolutely continuous and we can derive the spectral density \( g_0(\omega)d\omega = dG_0(\omega) \). Similarly, \( g_0(\omega) \) can be derived from the discrete Fourier transform of the autocovariance function \( \gamma_0(h) \), \( g_0(\omega) = \sum_{h=-\infty}^{\infty} \exp(-2\pi i \omega h) \gamma_0(h) \) and provides the same information about the dependence structure as \( \gamma_0(h) \). Under the same assumption, we could also consider the second-order characteristics of \( \{Z_{1,t}\} \) based on its spectral density \( g_1(\omega) = \sum_{h=-\infty}^{\infty} \exp(-2\pi i \omega h) \gamma_1(h) \).

The spectral densities \( g_0(\omega) \) and \( g_1(\omega) \) provide information regarding the dependence structure of \( \{Z_{0,t}\} \) and \( \{Z_{1,t}\} \), respectively, for frequencies, \( \omega \in (-0.5, 0.5] \). Then, for
\[
\rho_g(\omega) := \frac{g_1(\omega)}{g_0(\omega)}
\]
(here, we use the symbol := to mean defined as) we can modify the spectral process associated with \( \{Z_0(t)\} \) to generate a process

\[
Z_{1,t} \overset{L}{=} \int_{-0.5}^{0.5} \exp(2\pi i \omega t) \sqrt{\rho_g(\omega)} d\hat{Z}_0(\omega),
\]

that is equal in law to a stationary, Gaussian process with \( \mathbb{E}(Z_{0,t}) = 0 \) and covariance \( \gamma_1(h) \) (here, \( \overset{L}{=} \) indicates equality in law or equality in distribution). If \( g_1(\omega) = g_0(\omega), \forall \omega \in (-0.5, 0.5] \), \( \{Z_1(t)\} \) is equal in law to \( \{Z_0(t)\} \). In other words, in the event that the climate model indicates no changes in the dependence structure across all frequencies, the temporal dependence structure of the conditional simulations will be equivalent to the observations.

In this way, we derive the future, unobservable process in terms of the present process, modified by the ratio of their spectral densities.

So, one question becomes how to best represent \( \rho_g(\omega) \) when \( g_0(\omega) \) and \( g_1(\omega) \) are unknown, and \( g_1(\omega) \) cannot even be estimated because the future process has not yet been observed. In some applications, mechanistic models provide a general description of the process dynamics based on first principles. Let \( f_1(\omega) \) represent the spectral density associated with the computer model. For GCMs, it may be unreasonable to expect that \( f_1(\omega) \) is a close approximation to \( g_1(\omega) \). A more plausible assumption may be that, given the model’s suggested dependence structure under a baseline period, represented by \( f_0(\omega) \), that the estimated change in dependence structure is a decent approximation for the real changes in the dependence structure. In other words, rather than make the assumption that the model captures the correct dependence structure for a given GHG scenario, we assume:

\[
\rho_g(\omega) = \rho_f(\omega), \forall \omega \in (-0.5, 0.5],
\]

where \( \rho_f(\omega) := f_1(\omega)/f_0(\omega) \). It is possible that this will be a decent approximation, provided that the emissions scenario used to force the GCM is close to the “real” GHG scenario. That is, similar to the situation for the mean function illustrated in Figure 1,
we assume that, rather than capturing the dependence structure in an absolute sense, the GCM can provide a better approximation of the changes in the dependence structure. The details for carrying out the conditional simulation are provided in the following section.

2.2 Spectral-Based Conditional Simulation

Let \{Z_{0,t}; t = 0, ..., T - 1\} represent the observations of the process of interest, observed at regular time points. For now, assume that the process is stationary with \(E(Z_{0,t}) = 0\). We discuss how we account for seasonality in Section 2.4 and Section 2.5.

While in the previous section \(T = \infty\), in reality our observations are observed discretely over a finite period and the spectral process associated with \{Z_{0,t}\} is unknown. So first, we must approximate the modification of the true spectral process by using the discrete Fourier transform of the observations

\[
\hat{Z}_{0,k} = T^{-1/2} \sum_{t=0}^{T-1} Z_{0,t} \exp(-2\pi i \omega_k t),
\]

for \(\omega_k = k/T\) and \(k = -T/2 + 1, \ldots, T/2\). Here, \(\hat{Z}_{0,k}\) are complex-valued quantities, with \(\hat{Z}_{0,k} = \overline{\hat{Z}_{0,-k}}\). We can similarly define the cosine transform \(\hat{Z}_{0,k}^c = T^{-1/2} \sum_{t=0}^{T-1} Z_{0,t} \cos(2\pi \omega_k t)\) and sine transform \(\hat{Z}_{0,k}^s = T^{-1/2} \sum_{t=0}^{T-1} Z_{0,t} \sin(2\pi \omega_k t)\). The cosine and sine transforms relate to the DFT through the equality \(\hat{Z}_{0,k} = \hat{Z}_{0,k}^c - i\hat{Z}_{0,k}^s\). Because \(\hat{Z}_{0,k}^c\) and \(\hat{Z}_{0,k}^s\) are linear combinations of Gaussian processes, \(\hat{Z}_{0,k}^c\) and \(\hat{Z}_{0,k}^s\) are also Gaussian and, under the condition that \(\beta = \sum_{h=-\infty}^{\infty} |h||\gamma_0(h)| < \infty\) the following asymptotic results hold in terms of the dependence structure (see Shumway and Stoffer, 2011, for details) for the cosine transform

\[
\text{Cov}(\hat{Z}_{0,j}^c, \hat{Z}_{0,k}^c) = \begin{cases} 
  g_0(\omega_j)/2 + \epsilon_T, & j = k \\
  \epsilon_T, & j \neq k
\end{cases}
\]

the sine transform

\[
\text{Cov}(\hat{Z}_{0,j}^s, \hat{Z}_{0,k}^s) = \begin{cases} 
  g_0(\omega_j)/2 + \epsilon_T, & j = k \\
  \epsilon_T, & j \neq k
\end{cases}
\]
and also \( \text{Cov}(\hat{Z}_{0,j}^e, \hat{Z}_{s,k}^s) = \epsilon_T, \forall j, k \). Here, \( \epsilon_T \) can be shown to be bounded such that \( |\epsilon_T| \leq \beta / T \). As a result, our methodology is modifying nearly independent quantities in order to produce conditional simulations with a different dependence structure than the observations.

Let \( \hat{\rho}_f(\omega_k) \) be an estimate of the ratio of the spectral densities at \( \omega_k \). Then, the conditional simulations (not accounting for changes in mean) under a given scenario can be represented as

\[
Z_{1,t} = T^{-1/2} \sum_{k=-T/2+1}^{T/2} \sqrt{\hat{\rho}_f(\omega_k)} \hat{Z}_{0,k} \exp(2\pi i \omega_k t),
\]

suggesting the following covariance structure for \( \{Z_{1,t}\} \) for a given estimate \( \hat{\rho}_f(\omega_k) \)

\[
\mathbb{E}(Z_{1,t+h}Z_{1,t}) = T^{-1} \sum_{k=-T/2+1}^{T/2} \hat{\rho}_f(\omega_k) g_0(\omega_k) \exp(2\pi i \omega_k h).
\]

As a result, when \( \hat{\rho}_f(\omega_k) = 1, \ k = -T/2 + 1, \ldots, T/2, \) then \( Z_{1,t} = Z_{0,t}, \ t = 0, \ldots, T - 1. \) In the following section, we provide the details of a penalized likelihood approach to estimate \( \rho_f(\omega_k) \).

2.3 Estimation of the ratio of spectral densities \( (\rho_f(\omega_k)) \)

We propose a penalized likelihood approach for estimation of \( \rho_f(\omega_k) \), similar to the approach given in Pawitan and O’Sullivan (1994) for the estimation of one spectral density. Let \( \{Y_{0,t}; \ t = 0, \ldots, T - 1\} \) and \( \{Y_{1,t}; \ t = 0, \ldots, T - 1\} \) represent output from a climate model. Note that when \( \{Y_{0,t}\} \) and \( \{Y_{1,t}\} \) follow stationary, Gaussian distributions, the periodograms \( I_{j,k} = |\hat{Y}_{j,k}|^2 \), for \( j = 0, 1 \), follow (asymptotically) independent exponential distributions such that \( \mathbb{E}(I_{j,k}/f_{j,k}) \rightarrow 1 \) as \( T \rightarrow \infty \) (note \( f_{j,k} = f_j(\omega_k) \)). Let us write the likelihood in terms of \( \theta_{j,k} = \log(f_{j,k}) \). We propose using the following approximate penalized likelihood,

\[
\mathcal{L}(\theta_0) + \mathcal{L}(\theta_1) + \delta \mathcal{J}(\theta_0, \theta_1),
\]
where $\mathcal{L}(\theta_0)$ and $\mathcal{L}(\theta_1)$ represent the Whittle negative log-likelihood approximation (Whittle, 1954) $\mathcal{L}(\theta) = \sum_{k=-T/2+1}^{T/2} \{\theta_{j,k} + I_{j,k} \exp(-\theta_{j,k})\}$, for $j = 0, 1$. So, $\mathcal{L}(\theta_0)$ and $\mathcal{L}(\theta_1)$ provide an objective function that determines the fit to the data and $\mathcal{J}(\theta_0, \theta_1)$ is a penalty function.

Although penalties could be placed on the individual spectral densities themselves, for our analysis we only need an estimate of the ratio, and so we place the penalty on the ratio of the spectral densities

$$\mathcal{J}(\theta_0, \theta_1) = \mathcal{J}(\theta_1 - \theta_0).$$

Further, because we expect the ratio of spectral densities to be smoother than the individual spectral densities themselves, we are able to obtain a low-variance estimate of the ratio while increasing bias less than we would by smoothing each spectral density individually. Our penalty function is then placed on the $\ell^{th}$ derivative of $\lambda(\omega) = \theta_1(\omega) - \theta_0(\omega)$:

$$\mathcal{J}(\lambda) = (2\pi)^{-2\ell} \int_{-0.5}^{0.5} \{\lambda^{(\ell)}(\omega)\}^2 d\omega.$$  

Using Parseval’s Identity, this can be written as $\mathcal{J}(\lambda) = \sum_{k=-\infty}^{\infty} k^{2\ell} |\Lambda_k^*|^2$, where $\Lambda_k^*$ is the $k^{th}$ Fourier coefficient of $\lambda(\omega)$, $\Lambda_k^* = \int_{-0.5}^{0.5} \lambda(\omega) \exp(-2\pi ik\omega) d\omega$. Then, the penalty function $\mathcal{J}(\lambda)$ is approximated as $\mathcal{J}(\lambda) \approx \sum_{k=-T/2+1}^{T/2} k^{2\ell} |\Lambda_k|^2$, where $\Lambda_k$ is the discrete Fourier coefficient of $\lambda$, $\Lambda_k = T^{-1/2} \sum_{j=-T/2+1}^{T/2} \lambda(\omega_j) \exp(-2\pi ik\omega_j)$.

Estimation is carried out using an iterative least squares approach (McCullagh and Nelder, 1989), where

$$\mathcal{L}(\theta_j) \approx \sum_{k=-T/2+1}^{T/2} (m_{j,k} - \theta_{j,k})^2,$$

with

$$m_{j,k} = \theta_{j,k}^0 + (I_{j,k} - \exp(\theta_{j,k}^0)) \left( \frac{d\theta_{j,k}}{d\theta_{j,k}} \right)$$

$$= \theta_{j,k}^0 + I_{j,k} \exp(-\theta_{j,k}^0) - 1,$$
where the derivative \( d\theta_{j,k} / df_{j,k} \) is evaluated at \( \theta_{j,k}^{(0)} \). We point out that there is no weighting, but one could also accommodate for multiple realizations of model output, in which case weighting is necessary. Suppose, that \( I_{j,k} \) is actually the average of \( n_j \) periodograms from multiple realizations of the same climate scenario. In that case, we could proceed exactly as before, but instead we have

\[
\mathcal{L}(\theta_j) \approx \sum_{k=-T/2+1}^{T/2} w_{j,k} (m_{j,k} - \theta_{j,k})^2,
\]

where

\[
w_{j,k}^{-1} = \left( \frac{d\theta_{j,k}}{df_{j,k}} \right)^2 \text{var}(I_k) = n_j^{-1},
\]

where, again, \( d\theta_{j,k} / df_{j,k} \) is evaluated at \( \theta_{j,k}^{(0)} \). In this way, we can easily accommodate the use of multiple realizations for a given computer model scenario. Also, this can handle the situation in which there is a different number of independent realizations for the baseline and scenario runs. We proceed with an estimation procedure based on one realization, but note that this can easily be extended to accommodate multiple realizations.

The approximation to the entire objective function can be written as

\[
\sum_{k=-T/2+1}^{T/2} \left[ (m_{1,k} - \theta_{0,k} - \lambda_k)^2 + (m_{0,k} - \theta_{0,k})^2 + \delta k^2 |\Lambda_k|^2 \right],
\]

where, for a given smoothing parameter \( \delta \), we can iterate back and forth between estimates of \( \theta_{0,k} \) and \( \lambda_k \) until convergence. Therefore, the ratio of spectral densities can be estimated using the algorithm provided in Appendix A.

Note that we do not discuss the choice of smoothing parameter \( \delta \), but in this case assume it is known. In a situation in which multiple realizations of a climate scenario exist, it may be desirable to choose \( \delta \) based on a cross-validation study. At a minimum, one should consider multiple choices for \( \delta \) and consider the sensitivity to the results.
2.4 Adjustment of the seasonal cycle

The previous section assumed that the process of interest was a stationary Gaussian process with mean zero. In general, this is not a reasonable assumption for processes, such as daily temperature, that involve a strong seasonal component. So, before estimating the spectral ratio, we remove the seasonal trend of the observations, as well as for the model output.

As mentioned in the introduction, the Delta method uses model output for changes in first-order characteristics (e.g., seasonal mean) estimated from the model output. This is typically done by taking the difference in the seasonal cycles of the base and scenario time slices for the GCM, and then adding this difference onto the observations, as shown in Equation (1). Where the time series consists of daily observations, $\hat{\mu}_{0,t}$ and $\hat{\mu}_{1,t}$ could represent monthly means. Another option would be to consider annual harmonics to estimate the seasonal cycle:

\[
\hat{\mu}_{z,t} = \hat{\mu}_z + \sum_{k=1}^{K} \hat{R}_{z,k}\cos(2\pi \omega_k t + \hat{\phi}_{z,k})
\]
\[
\hat{\mu}_{0,t} = \hat{\mu}_0 + \sum_{k=1}^{K} \hat{R}_{0,k}\cos(2\pi \omega_k t + \hat{\phi}_{0,k})
\]
\[
\hat{\mu}_{1,t} = \hat{\mu}_1 + \sum_{k=1}^{K} \hat{R}_{1,k}\cos(2\pi \omega_k t + \hat{\phi}_{1,k}),
\]

for $\omega_k = k/365.25$. Then, $\hat{\mu}_0$ and $\hat{\mu}_1$ are the overall means for the observations, base, and scenario periods, respectively and $\hat{R}_{z,k}$, $\hat{R}_{0,k}$ and $\hat{R}_{1,k}$ are the estimated amplitudes at $\omega_k$, for the observations, base, and scenario periods, respectively. Lastly, $\hat{\phi}_{z,k}$, $\hat{\phi}_{0,k}$ and $\hat{\phi}_{1,k}$ are the estimated phase shifts for $\omega_k$.

2.5 Accounting for seasonal nonstationarity

Thus far we have assumed that the deseasonalized observations and model output, $\tilde{Z}_{0,t} = Z_{0,t} - \hat{\mu}_z$, $\tilde{Y}_{0,t} = Y_{0,t} - \hat{\mu}_0$, and $\tilde{Y}_{1,t} = Y_{1,t} - \hat{\mu}_1$ are (temporally) stationary. However, this need not be the case, and in applications involving daily mean temperature, it likely
is not the case. Figure 2 shows the log-averaged periodograms by season for a given pixel for the base and scenario period (the pixel is located in midwestern United States, centered at 39° N and 90° W). Clearly, the spectral density function for the base period are different for the different seasons, with the winter months showing higher variability across all frequencies then the other seasons, and the summer months showing lower variability across all frequencies. Note that, in the case of the scenario period, variability across frequencies in the winter has decreased, and is now roughly the same as spring and fall, but the summer variability is still roughly the same, and is lower across most all frequencies. Thus, the assumption of temporal stationarity is likely not reasonable. However, the log-periodograms are nearly parallel, suggesting that it may be reasonable to treat the process as a uniformly modulated process (Priestley, 1988). Specifically, we consider \( \tilde{Z}_{0,t} = D_{z,t}z_{0,t}, \)
\( \tilde{Y}_{1,t} = D_{1,t}y_{1,t} \), and \( \tilde{Y}_{0,t} = D_{0,t}y_{0,t}, \) after deseasonalizing, where \( \{z_{t}\}, \{y_{1,t}\}, \) and \( \{y_{0,t}\} \) are stationary processes (corresponding to the observations, model output under scenario period, and model output under base period, respectively). Then, \( \{D_{z,t}\}, \{D_{1,t}\}, \) and \( \{D_{0,t}\} \) are modulation constants to be estimated. Thus, if we can find suitable values for the modulation constants, then we can perform the spectral density estimation on \( \{y_{1,t}\} \) and \( \{y_{0,t}\} \), in order to modify the DFT of \( \{z_{0,t}\} \), and then multiply the constants back at the end to account for the nonstationarity across seasons. Note that this implies that our DFT-transform method will not account for seasonal changes in the dependence structure, but rather will maintain the relative differences that were available in the data (e.g., if variability corresponding to 2-day periods changes differently in the spring than in the fall, we will not be able to accommodate that with this approach).

Considering the process as uniformly modulated now consists of choosing and estimating the modulation constants. One simple option would be to standardize the observations by a seasonal standard deviation. However, it is more reasonable to consider constants that would vary smoothly with time of year, rather than having a constant value across each
season. Here we illustrate how we compute the modulation constants for the deseasonalized observations, \( \{\tilde{Z}_{0,t}\} \). Let \( I_t(\omega_k) \) represent an empirical periodogram of \( \{\tilde{Z}_{0,t}, \ldots, \tilde{Z}_{0,t+44}\} \) for frequency \( \omega_k = k/45 \). Then,

\[
T_t(\omega_k) = \frac{1}{45} \sum_{m=t-44}^{t} I_t(\omega_k)
\]

is the average of those periodogram estimates, and provides an estimate of the spectral density at frequency \( \omega_k \) for time \( t \). Now, we also average the estimate across years. For notational convenience, let \( \overline{T}_{d,y}(\omega_k) \) represent \( T_t(\omega_k) \), where \( d \) refers to the day of the year \( (d = 1, \ldots, 365) \) and \( y \) refers to the year \( (y = 1, \ldots, n_y) \). Then,

\[
\overline{T}_d(\omega_k) = \sum_{y=1}^{n_y} \overline{T}_{d,y}(\omega_k),
\]

and finally we sum up the variability across all frequencies and take the square root to produce the modulation constant for day \( d \)

\[
D_d = \sqrt{\sum_{k} \overline{T}_d(\omega_k)}.
\]

Figure 2 also shows the log-averaged periodograms for the modified process. While they are not equal across seasons, clearly the log-periodograms are much closer than previously, suggesting that most of the seasonal non-stationarity can be accounted for in this manner.

### 2.6 Outline of approach

We have outlined the three primary steps of our approach: the spectral density estimation and DFT modification, the seasonal cycle removal, and the modulation. However, we have presented these out of order in order to emphasize our contribution, namely the modification of the DFT of the observations. In order to highlight the correct order for generating the conditional simulations, starting with \( \{Z_t\} \), \( \{Y_{0,t}\} \), and \( \{Y_{1,t}\} \), the observations, base period time series, and scenario period time series, respectively, take the following steps:
1. Following the instructions in Section 2.4, deseasonalize \( \{Z_t\} \), \( \{Y_{0,t}\} \), and \( \{Y_{1,t}\} \) to get \( \{\tilde{Z}_t\} \), \( \{\tilde{Y}_{0,t}\} \), and \( \{\tilde{Y}_{1,t}\} \).

2. Following the instruction in Section 2.5, demodulate \( \{\tilde{Z}_t\} \), \( \{\tilde{Y}_{0,t}\} \), and \( \{\tilde{Y}_{1,t}\} \) to get \( \{ Z^*_t \} \), \( \{ Y^*_{1,t} \} \), and \( \{ Y^*_{0,t} \} \).

3. Estimate the spectral densities of \( \{ Y^*_{1,t} \} \) and \( \{ Y^*_{0,t} \} \), following the steps given in Appendix A. Then, use the estimated spectral densities to modify the discrete Fourier transform of \( \{ Z^*_{0,t} \} \), producing \( \{ Z^*_{1,t} \} \).

4. Multiply back the demodulation constants, \( \{ D_{z,t} \} \), \( \{ D_{0,t} \} \), and \( \{ D_{1,t} \} \), to produce

\[
\tilde{Z}_{1,t} = D_{z,t}(D_{1,t}/D_{0,t})Z^*_{1,t}.
\]

5. Add back seasonal cycle to produce \( Z_{1,t} = \hat{\mu}_{z,t} + (\hat{\mu}_{1,t} - \hat{\mu}_{0,t}) + \tilde{Z}_{1,t} \).

3 Application

In this section, we illustrate our methodology with an application related to simulating future surface air temperature under possibly changing dependence structures. The GCM output used is part of an ensemble of simulations completed by the Center for Robust Decision Making on Climate and Energy Policy (RDCEP). Specifically, we used the Community Climate System Model, version 3 (CCSM3), a coupled model that has atmosphere, ocean, land surface and sea-ice components (see Collins et al., 2006; Yeager et al., 2006, for details). The model was run at the “T31” spatial resolution (48 × 96 grid cells with a spatial resolution of approximately 3.75° × 3.75°). For the baseline run, all GHG concentrations were held constant at pre-industrial values. We use the end of our pre-industrial GHG run, which was run out over 2000 model years. For our scenario run, all GHG concentrations were held constant at pre-industrial values except CO₂, which followed a historical trajectory until 2010 and then increased to 700 parts per million (ppm) in 2100, at which
point it stabilized. In this case, we use as our scenario period, 8000-8031. The fact that
the model has been run for so long after stabilization of CO$_2$ suggests that we should be
close to equilibrium, particularly over land. We point out that CCSM3 uses a 365, rather
than a 365.25, day year. As a result, the Fourier frequencies will be slightly different for
the CCSM3 output than for the reanalysis data. We account for this by a simple linear
interpolation of the log spectral density ratio to the frequencies corresponding to the data.

For our observations, we used the NCEP Community Forecast System Reanalysis
(CFSR) data set of daily mean surface air temperature (see Saha et al., 2010, for de-
tails) from 1979 through 2010. Because the CCSM3 model output is at the T31 spatial
resolution, we regridded the reanalysis data using an areal conserving remapping function
in the NCAR Command Language (NCL; UCAR/NCAR/CISL/VETS, 2013). Thus, all
analysis for this paper is done on the same grid, although we point out that this procedure
could be used in a downscaling framework, where the estimated spectral ratio is done on
a different (and likely coarser spatial resolution) grid than the observations. One possible
option would be to modify the observations based on model output at a pixel whose center
is closest to the grid cell center of the reanalysis data.

Because we are not modifying spatial or spatio-temporal dependence structure, we can
take advantage of parallel computing technologies in order to quickly produce conditional
simulations over a larger spatial domain. Figure 3 shows the estimated, pixel-wise ratio
of spectral densities at two different frequencies, and the estimated spectral density of the
reanalysis observations, with the low-frequency map corresponding to a period of around
3.2 years. The high-frequency map corresponds to a period of 2 days. The results for the
spectral densities of reanalysis show higher variability at the poles relative to the equator
for both high and low-frequency variability. The maps for the spectral ratios are not as
clear, however. In general, we see more increased low-frequency variability over land than
we do decreased low-frequency variability; however, there does appear to be increasing
low-frequency variability in parts of Canada and Alaska. In the ocean, we see both areas with increased low-frequency variability (e.g., the north Atlantic ocean and south Pacific ocean) and other areas with decreased low-frequency variability (e.g., the Equatorial Pacific ocean). Interestingly, the Equatorial Pacific shows strong differences in the direction of changing variability between the low and high frequencies, with a decrease in low-frequency variability and increase in high-frequency variability. An area with a large increase in high-frequency variability is off the western coast of Africa at roughly $0^\circ$ N and $0^\circ$ E. This provides evidence that a simple rescaling of the observations is not an appropriate method of conditional simulation. Instead, one needs to consider how variability is changing across all possible frequencies. Figure 4 shows the estimated spectral density and the spectral ratio at four locations: 168.75$^\circ$ W and 1.86$^\circ$ N (Equatorial Pacific), 38.97$^\circ$ N and 90$^\circ$ W (midwestern USA), 1.86$^\circ$ S and 0$^\circ$ E (off the western coast of Africa), and 87.16$^\circ$ S and 11.25$^\circ$ E (Antarctica). Clearly, the changes in low-frequency variability in the Equatorial Pacific and the increase in high-frequency variability off the coast of Africa are much larger in magnitude than the other two locations, which show little change in overall variability across the different frequencies. This might suggest that at certain places, the Delta method may be sufficient as a downscaling technique.

Figure 5 provides the conditional simulations and reanalysis for the two locations near the equator. We can see, in both cases, that the model clearly underestimates the seasonal variability in the location off the coast of western Africa, and gets the seasonal cycle wrong in both locations. However, we simply use the estimated seasonal cycle from reanalysis to get an estimated change in the seasonal cycle, which is then added on to the observations. This is the portion of our method that is equivalent to the Delta method.
4 Discussion

Detailed characterization of the nature in which climate is changing (mean shifting, tail behavior, spatial and temporal dependence structures) is still a relatively open area of inquiry. One of the best ways of studying how future climate might change is in first investigating the nature in which the output of general circulation models are change from the present to possible future scenarios. We have provided a method of quantifying how temporal variation about the mean is changing in these general circulation models that accounts for these changes at different temporal scales. The evidence shows that variability is in fact changing differently at different locations, and that, at a given location, the changes in high-frequency variability are not necessarily similar to the changes in low-frequency variability. We showed how we can use this estimate of changing variability to produce conditional simulations that modify the temporal dependence structure of the observations. In this way, we extend the Delta method to be able to account for changes in the mean and dependence structure.

Our method for producing conditional simulations relies on modifying the discrete Fourier transform of the observations, and as such, the length of the conditional simulations in this manner is limited by the available data. However, it is possible that by recycling old observations, one could generate conditional simulations of longer length. Another possibility is to modify the observations by phase scrambling (Davison and Hinkley, 1997), and then adding these newer pseudo-observations to the true observations.

We point out that we did not account for changing spatial and spatio-temporal dependence structures. This will be the subject of future research. Accounting for changes in spatial dependence is complicated by the nonstationarity present in the observations (due to geography, land-ocean contrasts, etc.). However, we do note that, due to the use of the observations, we are accounting for any spatial nonstationarity present in the data.
currently, and so our method is simply not accounting for changes in the nonstationarity.

Next, while we have provided a method for producing conditional simulations of daily mean temperature, most impacts models also rely on conditional simulations for daily precipitation. The presented methodology is not fit to handle the non-Gaussian, nonlinear nature of daily precipitation, but perhaps this could be incorporated on a latent or transformed Gaussian process, a popular approach in the statistics literature (see, e.g., Kleiber et al., 2012, for an example). This also creates the possibility to consider multivariate bias-correction and downscaling of temperature and precipitation together.

Perhaps most importantly, the methodology presented herein is based on the assumption of stationarity in the model output and the data. While we did incorporate the concept of a uniformly modulated process to deal with seasonal nonstationarity, this methodology is still limited to conditionally simulating equilibrium climate. Because for the foreseeable future our climate will be in a transient state, we must consider ways of extending this methodology to be able to conditionally simulate transient climate. We point out that there is the potential for this methodology to be extended by considering evolutionary spectral approach (Priestley, 1988).

In spite of these obvious limitations, we feel that this approach sets up the general framework that should be considered when thinking about conditional simulation of high-resolution future climate. First, the approach should be observation driven, using the model output to suggest how to modify the existing observations. Second, the approach should consider changes in both mean and dependence structures, and this modification of dependence structure should involve modifying quantities that are at least approximately independent. The limitations mentioned in this discussion are opportunities to extend this framework to generate more realistic conditional simulations for their uses in impacts assessment.
Acknowledgements

The authors thank Michael Glotter for helpful comments regarding the manuscript. Leeds was funded by the NSF Statistical Methods in the Atmospheric Sciences Network (Award Nos. 1106862, 1106974, and 1107046) and the NSF Decision Making Under Uncertainty Program (Award No. 0951576).

References


A Spectral ratio estimation

The following algorithm is used to estimate the ratio of spectral densities. A complete description of the algorithm is given in Section 2.3. First, set the tolerance level, $\epsilon_{\text{tol}}$ that indicates when to stop the algorithm. Then, initialize $\theta_{0,k}^{(i)}$ and $\theta_{1,k}^{(i)}$ and set $i = 1$. Follow the following steps until convergence.

1. Set $m_{1,k}^{(i)} = \theta_{1,k}^{(i)} + I_{1,k}\exp(-\theta_{1,k}^{(i)}) - 1$ and $m_{0,k}^{(i)} = \theta_{1,k}^{(i)} + I_{1,k}\exp(-\theta_{1,k}^{(i)}) - 1$.

2. Set $M_1^{(i,k)} = \frac{T}{2} \sum_{j=-T/2+1}^{T/2} (z_{1,j} - \theta_{0,j})\exp(-2\pi ijk/T)$.

3. Set $\Lambda_k^{(i)} = (1 + \delta_k^2)^{-1}M_1^{(i,k)}$.

4. Set $\lambda_k^{(i)} = \frac{1}{T} \sum_{j=-T/2+1}^{T/2} \Lambda_j^{(i)}\exp(2\pi ijk/T)$.

5. Set $\theta_{0,k}^{(i+1)} = \frac{1}{2}(m_{1,k}^{(i)} + m_{0,k}^{(i)} - \lambda_k^{(i)})$.

6. Set $\theta_{1,k}^{(i+1)} = \lambda_k^{(i)} + \theta_{0,k}^{(i)}$.

7. Stop if $\sum_{k=-T/2+1}^{T/2} \left(\theta_{1,k}^{(i)} - \theta_{1,k}^{(i-1)}\right)^2 < \epsilon_{\text{tol}}$ & $\sum_{k=-T/2+1}^{T/2} \left(\theta_{0,k}^{(i)} - \theta_{0,k}^{(i-1)}\right)^2 < \epsilon_{\text{tol}}$, otherwise set $i = i + 1$ and return to Step 1.
Figure 1: Left: Plot of CO$_2$ trajectory for a “baseline” run, where all GHG concentrations are held constant at pre-industrial values, and a “scenario” run, where CO$_2$ concentrations use historical forcings until 2010, and then increase to 700 parts per million until 2100, at which point they stabilize. Right: The corresponding annual global mean temperature for the two runs, out to 2450. The observation time series is the global mean temperature anomaly data from the Global Historical Climatological Network, with an added baseline climatology of 14° C.
Figure 2: Top row: Log of averaged periodograms for each season in the base period and scenario period for a given pixel. While different, the log periodograms are roughly parallel, suggesting we can treat the deseasonalized process as uniformly modulated. Bottom row: Log of averaged periodograms after standardizing variability across seasons. The seasonal periodograms are much closer together, suggesting most of the nonstationarity is accommodated by treating the deseasonalized process as uniformly modulated. The pixel is from the midwestern United States, centered at 38.97° N and 90° W.
Figure 3: Top: A plot of the estimated spectral density of the reanalysis data at individual pixels for two different periods, corresponding to periods of 1169 days (left) and 2 days (right). Bottom: A plot of the estimated ratio of spectral densities of the GCM output at individual pixels two different periods, corresponding to periods of 1168 days (left) and 2 days (right).

Figure 4: Left: Logarithm of the estimated spectral density using reanalysis data at four pixels centered at: 168.75° W and 1.86° N (blue), 38.97° N and 90° W (green), 1.86° S and 0° E (red), and 87.16° S and 11.25° E (black). Right: The estimated ratios in spectral densities based on CCCSM3 output for the same pixels.
Figure 5: A plot of the time series for three years for the reanalysis and conditional simulations (top row) and the seasonal cycle for reanalysis, the conditional simulations, and model output (bottom row) for 1.86°S and 0°E (left) and 1.86°N and 168.75°W (right).