An Assessment of GCM Skill in Simulating Persistence across Multiple Time Scales

FIONA JOHNSON,* SETH WESTRA, AND ASHISH SHARMA

School of Civil and Environmental Engineering, University of New South Wales, Sydney, New South Wales, Australia

ANDREW J. PITMAN

Climate Change Research Centre, University of New South Wales, Sydney, New South Wales, Australia

(Manuscript received 17 March 2010, in final form 28 February 2011)

ABSTRACT

Climate change impact studies for water resource applications, such as the development of projections of reservoir yields or the assessment of likely frequency and amplitude of drought under a future climate, require that the year-to-year persistence in a range of hydrological variables such as catchment average rainfall be properly represented. This persistence is often attributable to low-frequency variability in the global sea surface temperature (SST) field and other large-scale climate variables through a complex sequence of teleconnections. To evaluate the capacity of general circulation models (GCMs) to accurately represent this low-frequency variability, a set of wavelet-based skill measures has been developed to compare GCM performance in representing interannual variability with the observed global SST data, as well as to assess the extent to which this variability is imparted in precipitation and surface pressure anomaly fields. A validation of the derived skill measures is performed using GCM precipitation as an input in a reservoir storage context, with the accuracy of reservoir storage estimates shown to be improved by using GCM outputs that correctly represent the observed low-frequency variability.

Significant differences in the performance of different GCMs is demonstrated, suggesting that judicious selection of models is required if the climate impact assessment is sensitive to low-frequency variability. The two GCMs that were found to exhibit the most appropriate representation of global low-frequency variability for individual variables assessed were the Istituto Nazionale di Geofisica e Vulcanologia (INGV) ECHAM4 and L'Institut Pierre-Simon Laplace Coupled Model, version 4 (IPSL CM4); when considering all three variables, the Max Planck Institute (MPI) ECHAM5 performed well. Importantly, models that represented interannual variability well for SST also performed well for the other two variables, while models that performed poorly for SST also had consistently low skill across the remaining variables.

1. Introduction

Climate model evaluation studies are becoming increasingly common and a range of metrics have been developed to identify models exhibiting strong performance (Giorgi and Mearns 2002; Murphy et al. 2004; Tebaldi et al. 2004; Dessai et al. 2005; Gleckler et al. 2008). The proliferation of metrics is in part due to the increasing diversity of studies that rely on general circulation model (GCM) outputs and the recognition that the identification of which models perform well is usually conditional to the question being asked (Knutti 2008). For water resources impact assessments, precipitation is often the key variable on which such assessments are based.

Biases in the mean state of precipitation fields from GCM simulations have been identified for some time (Xu 1999; Fowler et al. 2007), and some studies have used daily precipitation (Sun et al. 2006; Perkins et al. 2007) or annual mean precipitation (Murphy et al. 2004) to assess GCM performance. However, evaluation of precipitation projections over multiple time scales (Gleckler et al. 2008) is of particular importance for the management of water resources systems, especially in parts of the world influenced by large-scale modes of low-frequency climate variability with interannual or longer periods. If future climate projections do not correctly model interannual variability, then estimates

* Additional affiliation: Climate and Water Division, Bureau of Meteorology, Sydney, New South Wales, Australia.

Corresponding author address: Ashish Sharma, School of Civil and Environmental Engineering, University of New South Wales, Kensington, NSW, 2052 Australia.
E-mail: a.sharma@unsw.edu.au

DOI: 10.1175/2011JCLI3732.1

of the security of water resources systems with multiple-year storage capacity will be biased.

We would therefore like to evaluate how well GCMs model persistence at multiple time scales in precipitation. It is well known that major drivers of interannual and interdecadal variability in global precipitation are large-scale climate modes such as the El Niño–Southern Oscillation (ENSO) phenomenon, which is the dominant coupled ocean–atmosphere mode of the tropical Pacific (Cane 1992) as well as other teleconnections including the Indian Ocean dipole (IOD) (Saji et al. 1999; Ummenhofer et al. 2008) and the interdecadal Pacific oscillation (IPO) (Power et al. 1999; Mantua and Hare 2002). These climate modes are generally defined through sea surface temperature (SST) anomalies with corresponding anomalous pressure and wind fields. Westra and Sharma (2010) found that the overall predictability of global precipitation is approximately 15% owing to variability in SST anomalies. Other sources of variability in the climate system, and in particular precipitation, include soil moisture availability (Lau 1992; Koster et al. 2006), volcanoes, and solar variability (Peixoto and Oort 1992; Collins 2007), which all can lead to variability on multiple time scales.

Most GCM evaluation studies have concentrated on specific drivers of variability and in some cases have considered teleconnections with regional rainfall anomalies (Cai et al. 2009). There do not appear to be studies that have simultaneously evaluated the modeling of large-scale climate modes in GCMs over a range of time scales at a global scale. We now propose a methodology that allows this to be carried out. The aims of the paper are to address the following questions. First, is skill in modeling SST persistence associated with corresponding skill in precipitation persistence? Second, what are the implications on water resources impact assessments of selecting GCMs on the basis of skill in representing persistence?

We first present details of the wavelet methodology used to assess persistence in GCM simulations of SST, surface pressure, and precipitation. In section 3 we present the results of the wavelet analysis and determine the best-performing models. We then validate the wavelet results by quantifying the impact of long-term persistence in modeled rainfall using a reservoir storage analogy. Finally, we discuss the practical implications of selecting GCMs based on the modeling of persistence in the context of other available performance metrics.

2. Methodology

a. Wavelets

Wavelets allow time series to be decomposed into time and frequency domains (Torrence and Compo 1998). They have been used in many climate applications including the analysis of rainfall variability over Australia (Westra and Sharma 2006), changes in the time/frequency behavior of ENSO (Jain and Lall 2001), and analysis of variability in a reanalysis sea level pressure field (Barbosa et al. 2009). Wavelet decomposition has also been suggested as a possible method to derive metrics of the modeling of ENSO in GCMs (Guilyardi et al. 2009). In this analysis, we apply a discrete wavelet transform (DWT) to time series at each grid point for SSTs, surface pressure, and precipitation using the Daubechies family of orthogonal wavelets.

Daubechies wavelets allows the time series to be decomposed into a set of mutually orthogonal details and an approximation, with each detail capturing the variability of the time series at a particular frequency and the approximation representing the remaining low-frequency residual. The orthogonality of this class of wavelets ensures that it is possible to reconstruct the original time series perfectly by adding together each of the details and the approximation, and the sum of the variance of the details and the approximation also adds up to the variance of the original time series. This means that the fraction of variability of a time series at, say, interannual periods, can be calculated simply as the variance of all the details with periods within this range divided by the total variance of the original time series (see Burrus et al. 1998 for an overview of discrete wavelet transforms, or Daubechies 1992 for a more detailed mathematical treatment). In addition, the wavelet decomposition is attractive because it requires no assumptions on the form of variability that may be present at different frequencies (Barbosa et al. 2009).

In this study, we take the observed or modeled time series at each location and decompose it into eight details, with the remaining approximation covering long-term trends and oscillations in the signal with a period greater than about 32 yr. We split the details of the decomposed time series into three groups. The first group comprises the details with subannual frequencies (with periods of 0.23, 0.47, and 0.94 yr). The second group represents the interannual frequencies, which are defined as periods greater than annual but less than decadal, and for the Daubechies 4 wavelet this corresponds to periods of 1.87, 3.73, and 7.47 yr. ENSO cycles, with a typical period of 3–6 yr (Trenberth 1997), will therefore be assessed in this group. Finally, the interdecadal group is constructed from the long-term oscillations and/or any trend that is in the data (which may be a true trend or an oscillation with period greater than the data length). For each band of frequencies, the signal is reconstructed and the variance calculated and reported as a percentage of the total variance in the original

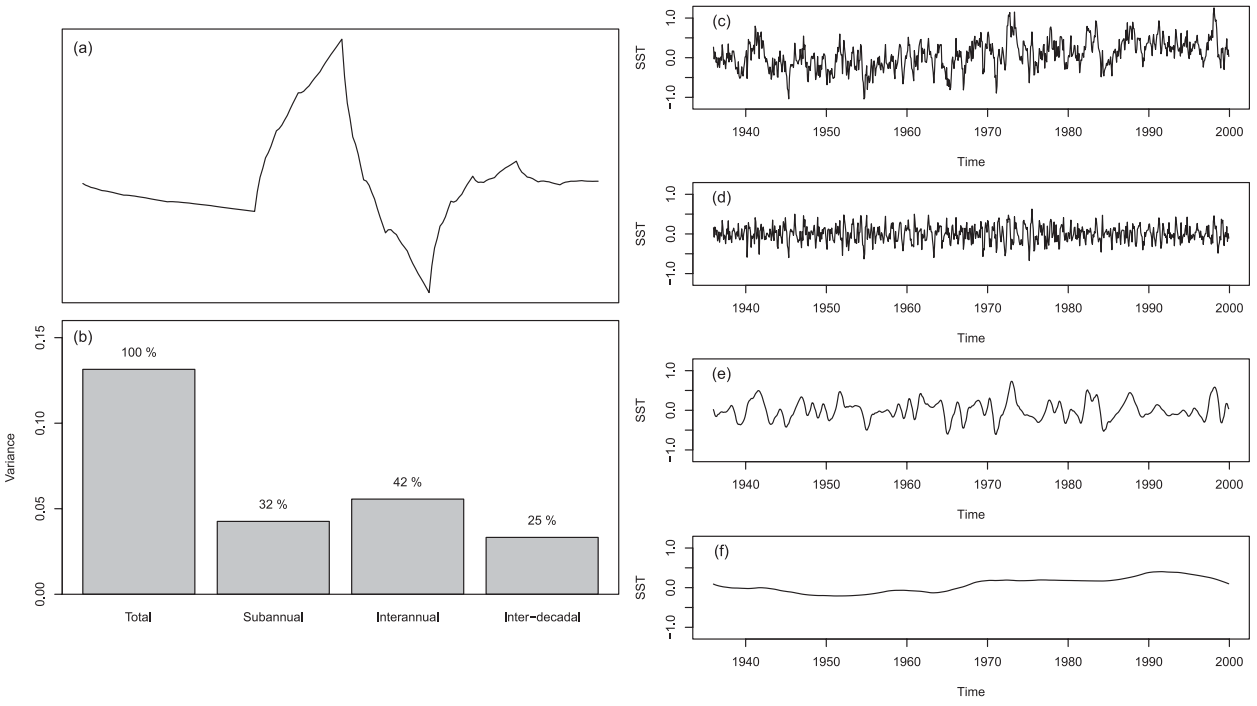

FIG. 1. Illustration of wavelet analysis methodology for observed SST anomalies over the Niño-3.4 region. (a) The Daubechies 4 wavelet is shown. (b) The contribution to the total variance from each of the frequency bands is shown. (c) The Niño-3.4 composite SST anomalies are shown with the decomposition into three frequency bands as described in the text shown for (d) subannual, (e) interannual, and (f) interdecadal frequencies, respectively.

time series. This process is illustrated in Fig. 1 for a time series of observed SST anomalies over the Niño-3.4 region, where we show the form of the Daubechies 4 wavelet and the three frequency bands of the total SST anomaly time series along with their contributions to the total variance in the time series. The results in this paper focus on the interannual variability as this is the frequency band of most importance for the study of droughts and yield assessments for large reservoirs with multiyear carry-over storage, although some results are also presented for the subannual and interdecadal timeframes. This methodology is similar to that used by Barbosa et al. (2009) in analyzing National Centers for Environmental Prediction (NCEP)–National Center for Atmospheric Research (NCAR) sea level pressure, although we have considered a wider range of frequencies and carry out the analysis to evaluate GCMs and reanalysis data compared to observations.

b. Data sources

We have used the outputs from 23 GCMs for the twentieth century available from the World Climate Research Programme (WCRP) Coupled Model Intercomparison Project phase 3 (CMIP3) multimodel dataset. Since some of the GCMs have multiple runs available, a total of 60 simulations were available for the analyses. Where multiple runs are available from a single GCM, results were averaged over all of the realizations. For each GCM we extracted monthly precipitation rates, SST, and surface pressures globally for the period 1936–99, with the 64-yr analysis period chosen to ensure consistency between the durations analyzed for the observations and models. All GCM data were interpolated to a common 5° by 5° grid across the globe, with the time series at each location converted to monthly anomalies by subtracting the monthly climatological mean.

The disadvantage of using anomaly series is that any biases in the mean state of the GCM simulations are removed and a model that performs poorly in modeling the mean seasonal cycle but captures other variability correctly may be rated better than one that simulates the mean climate state correctly and variability poorly. Which of these two hypothetical GCMs is actually the better model is contestable. However, the wavelet decomposition is such that even if anomalies were not removed from the time series, only the final wavelet approximation (i.e., the trend component) will include the means, as at all other time frequencies the wavelet details are reported, which by definition are always centered at zero. In any case, biases in the mean state can be easily removed from GCM precipitation simulations

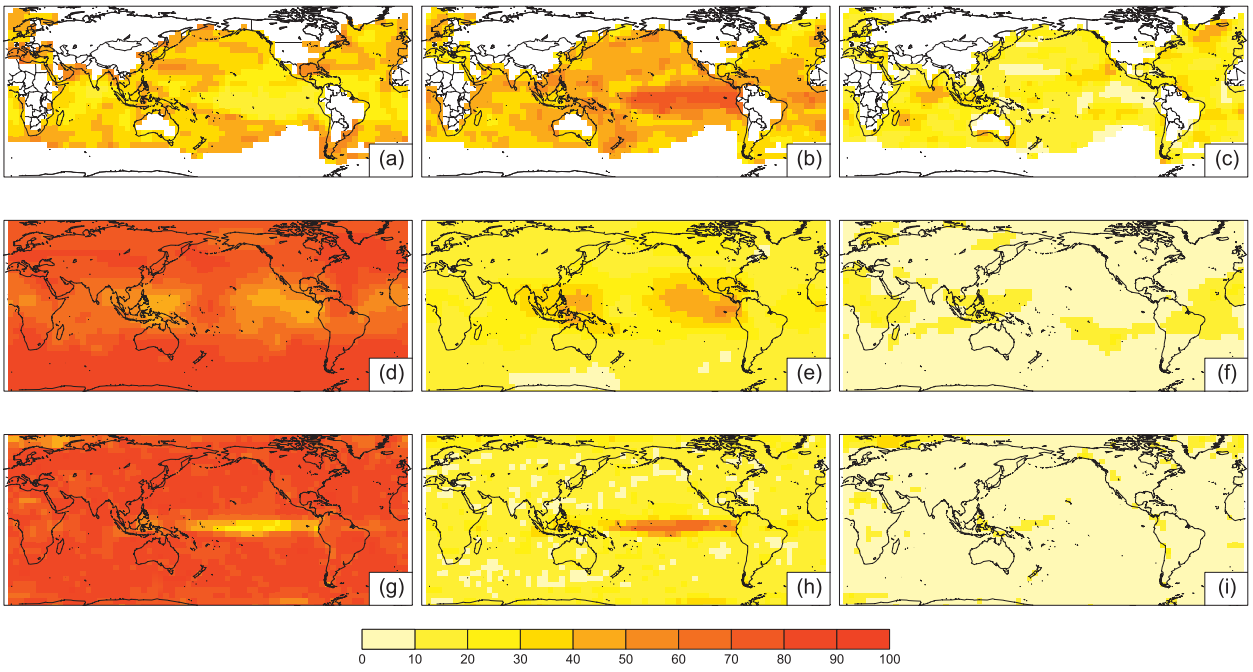

FIG. 2. Percentage variance in each frequency band (subannual, interannual, and interdecadal) for (a)–(c) SST, (d)–(f) surface pressure, and (g)–(i) precipitation derived from observations with a common color scale. (left) Subannual, (middle) interannual, and (right) interdecadal variances are shown. Reds indicate areas where a larger proportion of the total variance is contributed from that frequency band, while lighter yellows show locations with smaller contributions in that band.

using commonly applied bias correction approaches (Wood et al. 2004; Fowler and Kilsby 2007; Mehrotra and Sharma 2010), so we believe assessing the GCMs on variability alone is an acceptable approach.

Reconstructed SST anomalies are available from 1856 to 2009 (Kaplan et al. 1998) and sea level pressures from 1871 to 1998 [Hadley Centre Sea Level Pressure dataset (HadSLP1); Met Office (2009)]. For both the SST anomalies and sea level pressure we again use the final 64 yr of the historical record for the wavelet analysis. To test the sensitivity of the methodology to the chosen observed datasets we also used the National Oceanic and Atmospheric Administration (NOAA)/National Climatic Data Center (NCDC) extended reconstructed SST anomalies. For the precipitation data, the Global Precipitation Climatology Project (GPCP) data (Adler et al. 2003; Huffman et al. 2009) are the only observation dataset with global land and sea coverage, and it covers the period 1979–2007. The final dataset that has been analyzed is the NCEP–NCAR reanalysis of SST, surface pressure, and precipitation. For all datasets, the time series were extended to dyadic lengths where necessary using periodic extension, which preserves the orthogonality of the details and approximation. The impact of using a shorter length time series for the GPCP data was considered through a sensitivity analysis and is reported in the following section. In general, it has been previously noted that observational records and century-long GCM runs are relatively short when compared to climate modes that may have periodicities on the order of a decade or more (AchutaRao and Sperber 2006; Wittenberg 2009). In this study, we address the issue of short observational and model-derived datasets in a number of ways. First, where available we use multiple integrations of each GCM for the twentieth-century runs. Each integration is based on different initial conditions, which leads to different evolutions of the climate over the period of analysis. Second, we use multiple observed datasets as a way to mitigate observational uncertainty (Gleckler et al. 2008). Finally, we consider reanalysis data as an additional observational representation given its frequent use for testing and validating climate models (Reichler and Kim 2008).

3. Results

a. Spatial patterns of variance

We first present results showing the decomposition of the observed time series at each grid cell into subannual, interannual, and interdecadal components. For each grid cell we have shown the percentage variance of the original time series accounted for by each frequency band (i.e., subannual, interannual, and interdecadal). In Fig. 2, we can see the proportion of variance is reasonably

evenly split between the three frequency bands for SSTs (Figs. 2a, 2b, and 2c), while for surface pressures and precipitation greater than 70% of the variance occurs at the subannual time scale (Figs. 2d and 2g, respectively). A common color scale has been used for Figs. 2a to 2i to allow an easy comparison of the proportion of variance in each frequency band for the three different variables. However, owing to the small variance proportion in the interannual and interdecadal bands for the precipitation and surface pressure (Figs. 2f, 2h, and 2i), these maps have been repeated as Figs. 3a, 3b, and 3c with new color scales to highlight the regional variations in interannual and interdecadal variability, which are discussed below.

Regional variations in all three frequency bands show well known patterns of climate variations, which gives confidence in the applicability of the wavelet methodology. The highest interannual variance percentage for all three variables is found in the tropical Pacific Ocean. For SSTs the largest variance is in the ENSO regions and in particular the eastern tropical Pacific, while for the surface pressures, variance is highest adjacent to Central America and Southeast Asia and appears to correspond with the descending and ascending branches of the Walker circulation over the equatorial Pacific Ocean (Peixoto and Oort 1992). The band of high interannual variance in precipitation is confined to a smaller-latitude range compared to SSTs. It extends across the whole Pacific Ocean and into Indonesia and appears to be located along the intertropical convergence zone (ITCZ).

Because of the short length of the observed precipitation dataset and inherent inhomogeneity in rainfall fields, the spatial patterns of variances in the interdecadal frequency band are less smooth than for surface pressure or SST. Precipitation over oceans generally shows less interdecadal variance than over land areas. The highest proportions of interdecadal variance are apparent for parts of the world where precipitation anomalies have previously been shown to be influenced by the ENSO phenomenon and the Pacific decadal oscillation (Mantua and Hare 2002), including southeastern Africa and the western coasts of South and North America. The relatively large precipitation interdecadal variance in central North Africa may be related to sampling issues with the short observed record and low rainfalls in this region. SST interdecadal variability is found in the Indian Ocean adjacent to Madagascar and south of the Indian subcontinent. High percentages of interdecadal variability are also evident in parts of the North Atlantic Ocean and in the Southern Ocean; however, this is close to the southern boundary of the observational dataset where the estimated errors in the reconstructed SSTs are highest (Kaplan et al. 1998) so these results should be viewed with caution.

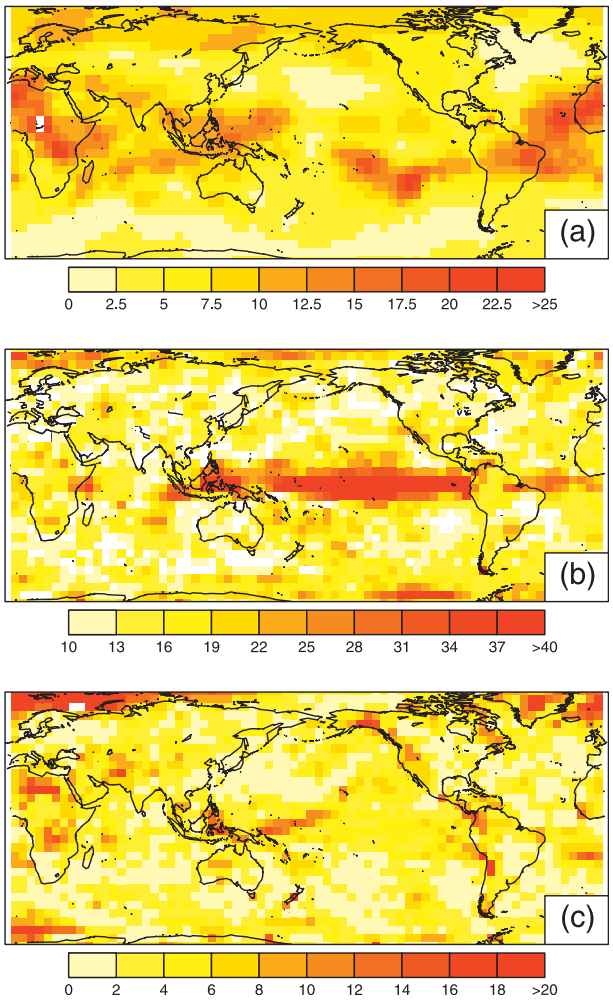

FIG. 3. As in Figs. 2f, 2h, and 2i, but with individual color scales chosen to better highlight regional variations.

We compare how closely the GCMs match the observations through the use of Taylor diagrams, which provide a summary of the match in the global patterns of the observations to model simulations in terms of correlation, root-mean-square (RMS) difference, and variance ratio (Taylor 2001). Figure 4 shows the Taylor diagrams for SST, surface pressure, and precipitation. The field that is being analyzed in each case is the proportion of variance in the interannual time period for SST, surface pressure, or precipitation, respectively. The observations are shown at the bottom of the figure as a reference point, with the distance of this reference point from the origin proportional to the overall standard deviation of the spatial pattern. Standard deviation contours from the origin are shown in blue. Contours showing the RMS difference between the model results and the observation are shown in green. The model results are then plotted based on the centered RMS distance and

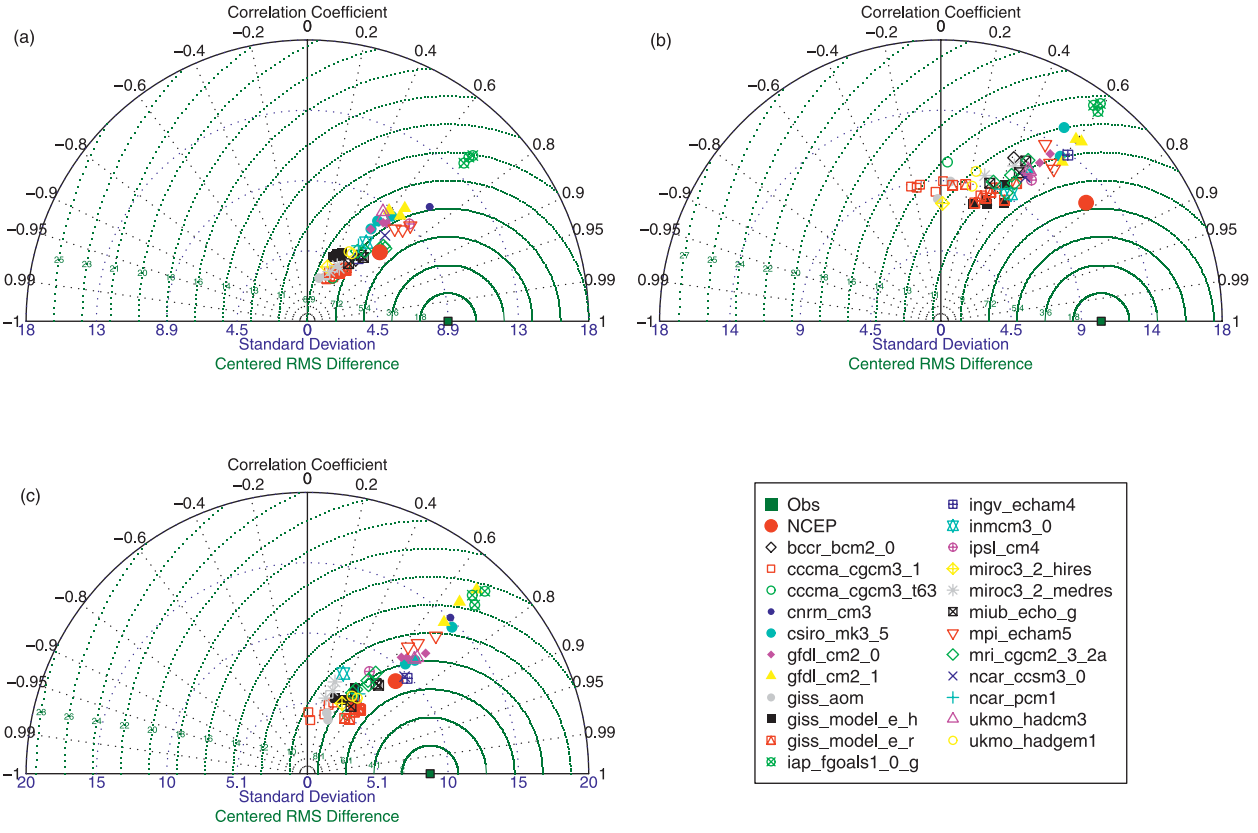

FIG. 4. Taylor diagrams showing the degree of agreement between observations, reanalysis, and GCM spatial patterns of interannual variance percentages for (a) precipitation, (b) SST, and (c) surface pressure. Multiple integrations of each model are shown with the same symbol and color.

correlation with the observations, with the azimuthal position (shown in black) representing the spatial correlation of the model and observed fields.

In examining these figures, we first note that the SST reanalysis provides the best match to the SST observations as would be expected. Conversely, the performance of the reanalysis for pressure and precipitation cannot be distinguished from that of the GCMs, with many of the GCMs reproducing the overall variance in the precipitation interannual variability as well as the reanalysis. Global mean variance in each frequency band, calculated by a weighted average of the variance at each grid point (with the weights determined by the relative surface area of each grid point, which varies by latitude), show that the reanalysis has a strong trend component in both the surface pressure and precipitation fields, which is not seen in the observations. This leads to a smaller proportion of variance being present in the other time periods compared to the observations. Simmons et al. (2004) note that the utility of reanalysis data ''for helping to document and understand climatic trends and low frequency variations is . . . a matter of some debate.'' In the case of the SSTs, the GCMs show a large range for the proportion of variance attributed to the trend component, ranging from 5% to over 20% for the SSTs, while there is better agreement between the GCMs on the proportion of variance in the trend component for precipitation and surface pressure.

Comparing the Taylor diagrams for the subannual and interdecadal frequencies (not shown) to those in Fig. 4, we find that the GCMs and reanalysis represent the interannual variability variance percentage the best of all three frequencies bands. The pattern correlations are higher in all cases for the interannual frequencies than for the other two frequency bands, and for the interdecadal band all the GCMs underestimate the standard deviation of the global field for surface pressure and precipitation, with the reanalysis providing much better estimates.

The clustering of the multiple ensemble members from individual GCMs, shown in Fig. 4, demonstrates that there are not large differences in the interannual variability component as a result of different initial

conditions. This was previously illustrated for a single GCM forced by 15 yr of SST anomalies where the interannual variations are quite similar despite differences in the time evolution of month to month fluctuations (Lau 1992). However, the results from our study show that the variation between different integrations of a single GCM is generally much smaller than the variations between first the entire multimodel ensemble and second the variations between the GCMs as a group and the observations for all three variables. It has been shown previously that model internal variability is more important at smaller spatial scales than globally (Hawkins and Sutton 2009) so these results are not unexpected. They suggest that the capacity to model interannual variability is therefore a model structure issue rather than an initial conditions issue. By implication we must therefore look at what are the differences in model structure if we are to understand why some models are better than others in simulating interannual variability, which is beyond the scope of this paper.

b. Interannual variability skill

We initially posed the question of whether the skill of a GCM in representing persistence in SST fields is related to skill in representing other manifestations of large-scale climate modes such as surface pressures and precipitation. To answer this, we use the skill score derived by Taylor (2001) as shown in (1):

$$S = \frac{4(1 + R)^4}{(\hat{\sigma}_f + 1/\hat{\sigma}_f)^2(1 + R_0)^4}, \quad (1)$$

where R is the pattern correlation coefficient between the observations and GCM under consideration, $\hat{\sigma}_f$ is the ratio of the standard deviation of the model field to the standard deviation of the observed field, and R_0 is the maximum theoretical correlation and has been taken to be 1.0. Both the correlations and standard deviations are calculated using a weighting based on grid cell area. This skill score rewards models which have good pattern correlations in preference to models which match the magnitude of the variance of the overall pattern (Taylor 2001). When there is a perfect match between the model and the observations, the score will be one (when both R and $\hat{\sigma}_f$ equal one) and for decreasing model performance, S approaches zero.

After calculating the skill score for each model, we have averaged the score over the multiple integrations of the same GCM if available to determine an aggregated score. The results for each variable for skill in matching the interannual variability of the observations are presented in Table 1, along with the rank of each model according to the skill score (with 1 being the best performing and 23 the worst performing GCM). The best GCMs from this analysis are identified as L'Institut Pierre-Simon Laplace Coupled Model, version 4 (IPSL CM4) and the Istituto Nazionale di Geofisica e Vulcanologia (INGV) ECHAM4 for individual variables; when considering performance over all three variables, the best performing models are Max Planck Institute (MPI) ECHAM5, NCAR Community Climate System Model, version 3.0 (CCSM3.0), and Centre National de Recherches Météorologiques Coupled Global Climate Model, version 3 (CNRM-CM3). Previous studies (van Oldenborgh et al. 2005; Coelho and Goddard 2009) have found that the best performing models for ENSO variance and rainfall teleconnection strength are Geophysical Fluid Dynamics Laboratory Climate Model version 2.0 (GFDL CM2.0), GFDL CM2.1, the third climate configuration of the Met Office (UKMO) Unified Model (HadCM3), and ECHAM5, all of which show good skill particularly for precipitation. As it is generally considered preferable to use multiple GCMs in impact assessments rather than selecting only the best-performing models, the skill scores presented in Table 1 could be used to weight model contributions to multimodel ensemble estimates of future changes or to select a subset of GCMs that represents interannual variability well.

Using the skill scores, we can compare the spatial patterns of the proportions of interannual variance for the best- and worst-performing GCMs to the observations. In Fig. 5, the models with the highest and lowest skills are shown for each climatic variable. The best-performing models have good matches to the observations in the tropics. For SSTs the INGV ECHAM4 model also matches the observations well in the northern Pacific, although it shows less variance adjacent to the east coast of Australia than is present in the observations. Atlantic Ocean variances are also of a similar magnitude, while Indian Ocean variances are slightly higher in the observations. On the other hand, the Canadian Centre for Climate Modelling and Analysis (CCCma) model shows almost no variation in the tropics with the highest proportion of interannual variance occurring in the Southern Ocean and northern Atlantic. The CNRM-CM3 model matches the large observed variance in precipitation in the ITCZ, while the poorly performing Goddard Institute for Space Studies Atmosphere–Ocean Model (GISS-AOM) model has generally uniform spatial variations in precipitation interannual variability. The best- and worst-performing models for surface pressure are the same as for SST anomalies. INGV ECHAM4 matches the magnitude of variations well, although the maximum westerly

TABLE 1. Skill scores and ranks for model performance in matching observed interannual variance percentages. Highest skill scores are shown in boldface.

Model	SST skill	Precipitation skill	Pressure skill	SST rank	Precipitation rank	Pressure rank	Avg skill	Avg rank
NCEP	0.61	0.49	0.51	—	—	—	0.54	
BCCR BCM2.0	0.24	0.31	0.28	15	14	16	0.28	15
CCCma Coupled General Circulation Model, version 3.1 (CGCM3.1)	0.04	0.11	0.1	23	21	23	0.08	23
CCCma CGCM3- t63	0.07	0.13	0.27	20	20	17	0.16	20
CNRM-CM3	0.37	0.54	0.37	3	2	8	0.43	3
Commonwealth Scientific and Industrial Research Organisation Mark version 3.5 (CSIRO Mk3.5)	0.33	0.39	0.45	7	9	4	0.39	9
GFDL CM2.0	0.34	0.4	0.45	6	8	4	0.40	6
GFDL CM2.1	0.38	0.45	0.34	2	5	9	0.39	8
GISS-AOM	0.07	0.07	0.12	20	23	22	0.09	22
GISS MODEL E-H (GISS E-H)	0.22	0.17	0.26	17	18	18	0.22	18
GISS Model E-R (GISS-ER)	0.19	0.18	0.34	18	17	9	0.24	16
Institute of Atmospheric Physics (IAP) Flexible Global Ocean–Atmosphere–Land System Model gridpoint version 1.0 (FGOALS-g1.0)	0.31	0.42	0.33	10	7	12	0.35	10
INGV ECHAM4	**0.39**	0.33	**0.54**	1	12	1	0.42	4
Institute of Numerical Mathematics Coupled Model, version 3.0 (INM-CM3.0)	0.3	0.36	0.2	11	11	20	0.29	14
IPSL CM4	0.35	**0.55**	0.33	4	1	12	0.41	5
MIROC3.2(hires)	0.06	0.1	0.22	22	22	19	0.13	21
MIROC3.2(medres)	0.24	0.15	0.15	15	19	21	0.18	19
MIUBECHOG	0.26	0.32	0.38	13	13	7	0.32	12
MPI ECHAM5	0.35	0.54	0.41	4	2	6	0.43	2
Meteorological Research Institute Coupled General Circulation Model, version 2.3.2a (MRI CGCM2.3.2a)	0.27	0.44	0.34	12	6	9	0.35	11
NCAR CCSM3.0	0.32	0.47	0.53	8	4	2	0.44	1
NCAR Parallel Climate Model version 1 (PCM1)	0.25	0.31	0.33	14	14	12	0.30	13
UKMO HADCM3	0.32	0.39	0.47	8	9	3	0.39	7
UKMO Hadley Centre Global Environmental Model version 1 (HADGEM1)	0.14	0.26	0.31	19	16	15	0.24	16

variance is located farther west than seen in the observations.

Generally models that perform particularly well or poorly in matching the observed patterns of interannual variability for one variable have similar performance for the other two variables. Figure 6 presents scatterplots showing the relationships between pairs of climatic variables. The correlations in the skill scores between the three pairwise combinations of all skill scores range between 0.75 and 0.85. This demonstrates that GCMs that model the variations in SSTs for the twentieth century correctly will match the interannual variability in precipitation observations better, which is important for water resources impact assessments. Santer et al. (2009) caution that GCM errors are likely to be complex and interlinked. The simple correlations of skill scores presented above may not fully account for this complex error structure. Expanding the analysis to more variables may shed further light on this issue.

c. *Use of wavelet-based skill score for model selection for reservoir storage estimates*

What are the implications of using the best and worst models identified in the previous section on climate impacts that require interannual variability to be simulated correctly? We demonstrate the impacts of incorrectly modeling interannual variability with a simple synthetic study of reservoir storage. We use observed rainfall for a point in southeastern Australia (37.5°S, 147.5°E) from the Australian Bureau of Meteorology monthly 0.25° gridded dataset, which allows a 64-yr record to be analyzed. We analyze two time series, one with the observed monthly rainfall totals and the second with the interannual variability component removed from the original observations. The interannual variability was removed by decomposing the time series using the discrete wavelet transform previously described and reconstructing the signal without the variance that

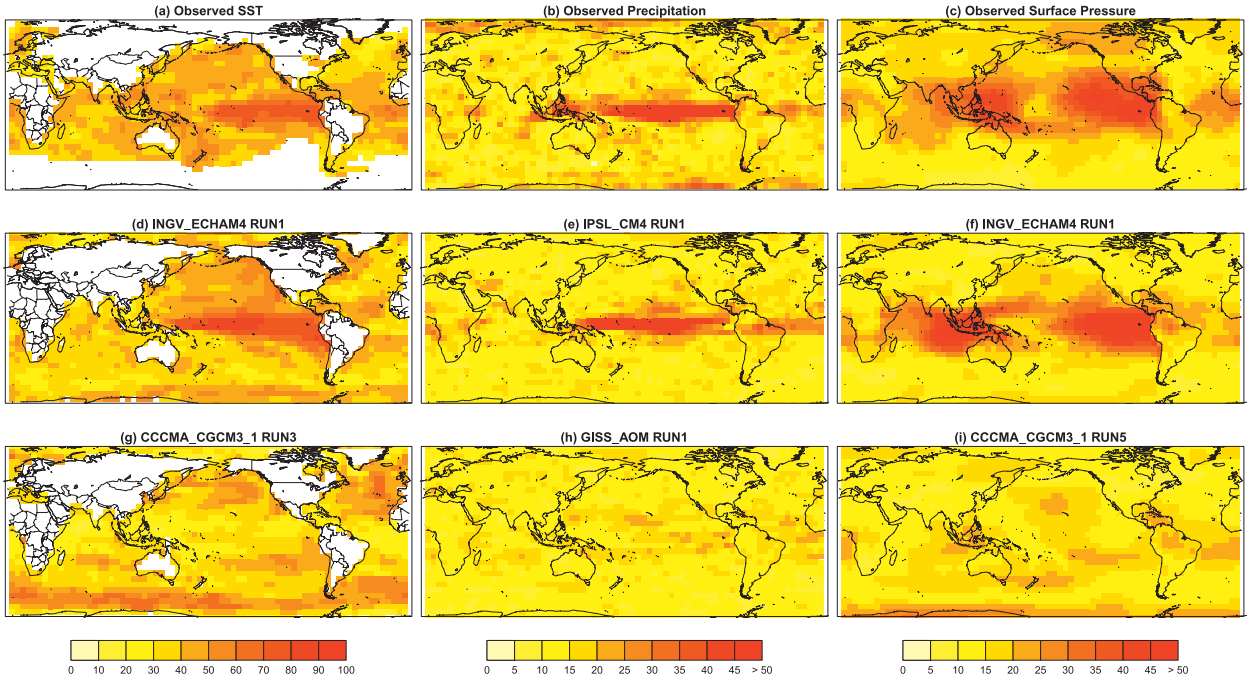

FIG. 5. Percentage variance in the interannual frequency band for (a)–(c) observations, (d)–(f) GCMs with highest skill, and (g)–(i) GCMs with lowest skill. (left) SST variability, (middle) precipitation, and (right) surface pressure are shown. Note the scale bar has a larger range for SST than for precipitation and surface pressure.

occurs with periods of approximately 2–8 yr. The second time series is then shifted and scaled such that the monthly mean and standard deviations match those of the original time series of observations (63 and 41 mm, respectively). The two time series are shown in Fig. 7a, where visually they are almost indistinguishable from each other.

To show the impact of the incorrectly specified interannual variability we use the time series to simulate reservoir storage by calculating the cumulative sum of each time series with a demand equal to the monthly mean rainfall (Koutsoyiannis 2002; Wasko and Sharma 2009). This is equivalent to a model of a reservoir providing a constant outflow equal to the mean inflow to the reservoir and operated such that there are no spills or other losses. The resulting cumulative time series are shown in Fig. 7b. These curves show the large impact of interannual variability on storage estimates. The required storage in each case is calculated as the maximum cumulative sum minus the minimum cumulative sum and as shown in Fig. 7, the required storage is underestimated by approximately 10% if we use the time series with incorrectly specified interannual variability. Substantially larger differences could be expected were the evaluation conducted for streamflows resulting from the rainfall, even for medium-sized water supply catchments.

We now extend the analysis to consider the estimates of storage from the full suite of GCMs used in the study and compare these to the wavelet estimates of variability at different time frequencies. Since the skill scores calculated in section 3 were based on the global results, we do not necessarily expect them to reflect skill at all individual locations. However, there will be GCMs that for a particular region represent the observed distribution of variance across a range of time frequencies better than other GCMs. So although the results on which GCM provides the best match to observations will be different for any given reservoir case study, we feel that the following example demonstrates a useful application of a wavelet-based analysis of variance of GCM performance.

We consider the time series from each GCM for the analysis point in southeastern Australia, and with the 64 years of precipitation data used to calculate the variance proportions for each GCM in section 3 to calculate the required storage. We have applied a monthly bias correction to each of the GCM time series such that the monthly means and standard deviations match the observations over the annual cycle. The storage calculations are presented in Fig. 8, plotted against the proportion variance in each of three wavelet frequency bands that we have considered in this paper. The observations are shown as a black triangle in each plot. It is clear that

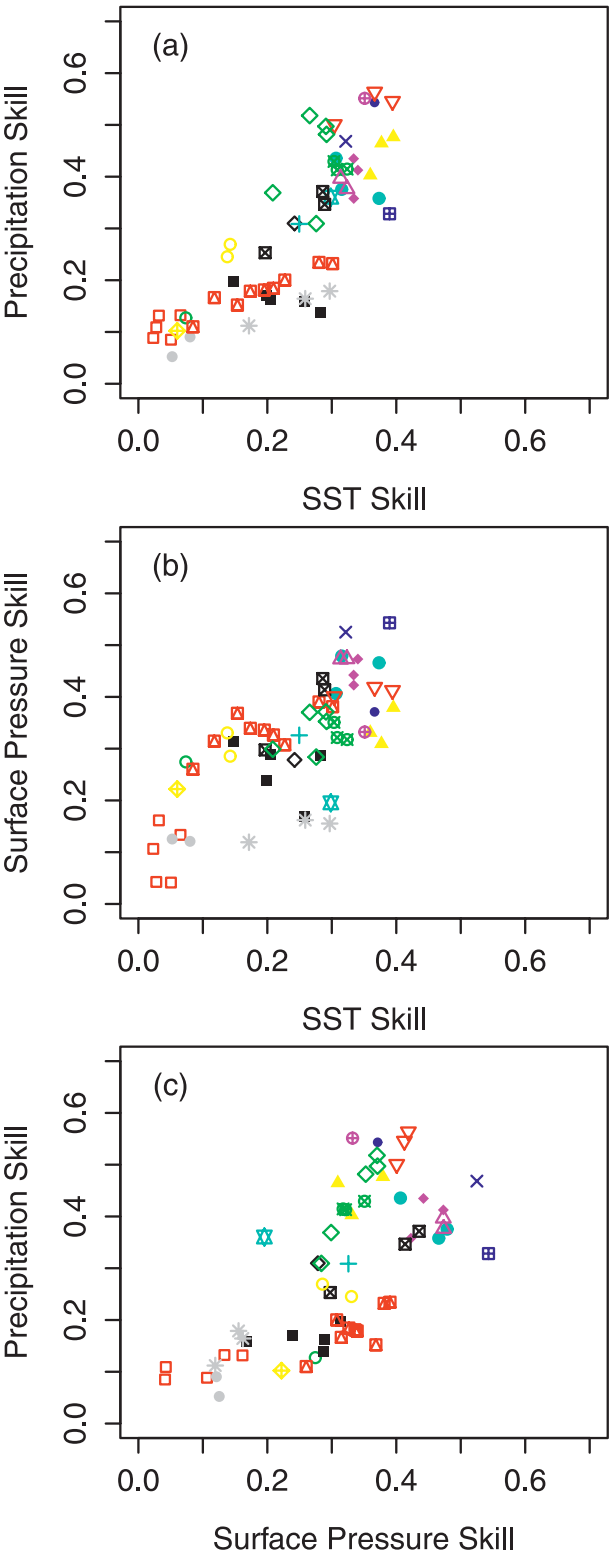

FIG. 6. Scatterplots showing the relationship between model interannual variability skill in (a) SST and precipitation, (b) SST and surface pressure, and (c) surface pressure and precipitation. Generally, models with good skill in reproducing the spatial pattern and magnitude of the observed interannual variability will also show good skill in the other variables. Refer to Fig. 4 for legend of GCM symbols and colors.

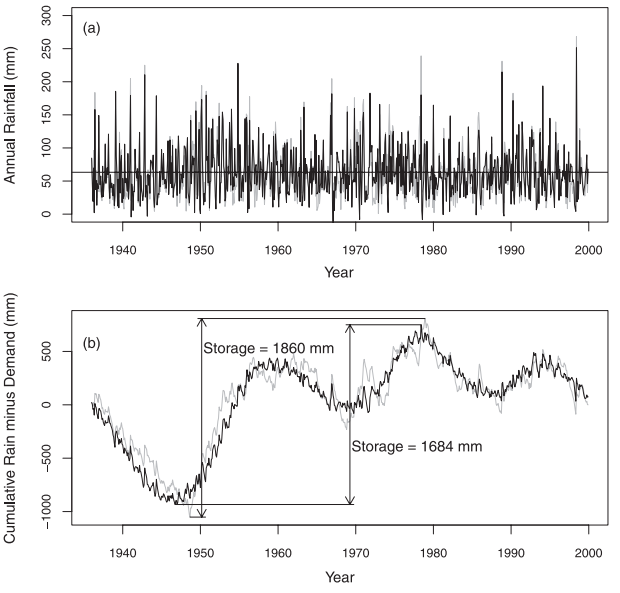

FIG. 7. (a) Observed rainfall time series at a point in southeastern Australia (37.5°S, 147.5°E) demonstrating the impact of incorrectly specified interannual variability with observed time series shown as a gray line and time series with interannual variability removed as a black line. (b) Cumulative sums of rainfall minus demand are shown with the same line types as in (a). Maximum required storage is shown for both time series.

storage calculations are most clearly affected by variance at long time scales with the strongest correlations between the storage estimates and the variance proportion for the interdecadal frequencies. This is not surprising as it is well known that storage calculations are strongly influenced by long-term dependence and trends in a time series (Koutsoyiannis 2002). But also evident is that those models that best match the proportion of variance seen in the observations (approximately 2.5%), lead to the estimates of storage that are closest to the observations (approximately 1800 mm). Models that strongly overestimate interdecadal variability, overestimate storage as well. Because we have defined variance at different frequencies as a proportion of the total variance in the time series, an inverse relationship between storage and variance is evident from Figs. 8a and 8c. Models that underestimate the proportion of variance at subannual frequencies must overestimate the proportion of variance at lower frequencies, as the total proportion variance will always be 100%. As shown above GCMs with too much low-frequency variance will have high estimates of storage, and thus a negative relationship between storage and proportion of variance will be a result as seen in Fig. 8a.

Another interesting feature of Fig. 8 is the difference that a small change in the variance proportion can have on the storage estimates. For each 1% increase in the

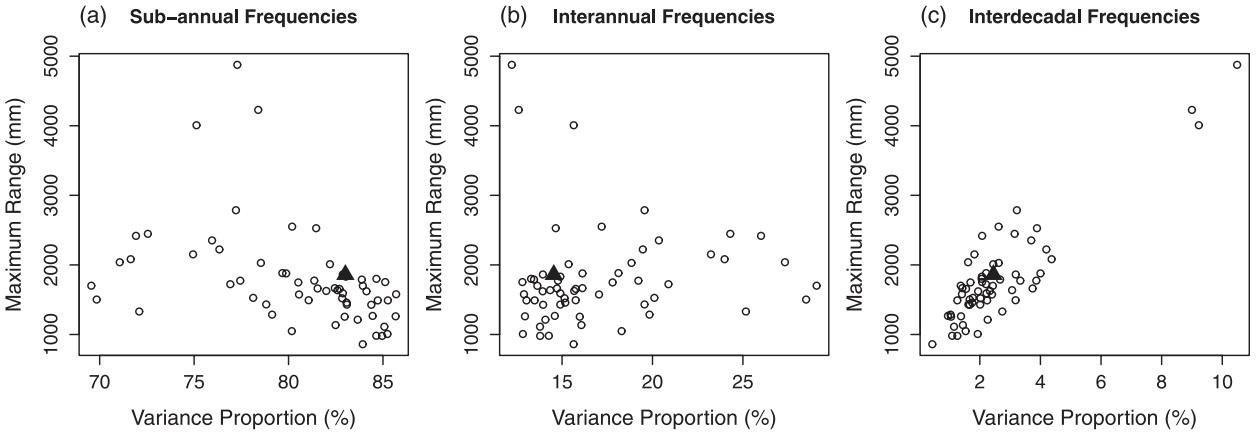

FIG. 8. Range vs proportion of variance in the (a) subannual frequency band, (b) interannual frequency band, and (c) interdecadal frequency band. GCM range and variance estimates are shown as small black dots, with the range and variance from the observed rainfall time series shown as the large filled triangle.

interdecadal variance proportion the estimated storage increases by 350 mm (based on the best-fit line through the data). This estimate is influenced by the three Model for Interdisciplinary Research on Climate 3.2, medium-resolution version [MIROC3.2(medres)] storage values, which each have an interannual variance of approximately 10% (due to a strongly decreasing trend in the precipitation time series at this location). If these models are excluded from the calculations, then a 1% increase in variance proportion equates to a 300 mm increase in storage. If, as shown in this simple example, our climate change impact assessment was considering possible changes to storage in a catchment, then we might want to change the skill score to reflect the modeling of trends and interdecadal variance rather than the interannual variance skill scores derived previously. The wavelet methodology provides sufficient information on the GCMs for this to be quickly achieved. This also highlights the importance of model evaluation metrics being designed to be appropriate for the particular impact being considered.

A bias correction step was required to allow the storage estimates from different GCMs to be compared. This illustrates one of the limitations of our study, in that we are only reporting the proportion of variability in each frequency band, as a percentage of the total variability for that time series. Therefore, if a model were to underestimate the overall variability but proportion it into the different frequency bands in the same way as the observations, it would have a higher skill score than a model that simulated the magnitude of the total variability correctly but had different frequency proportions. However, as noted before and applied for the storage estimates, simple bias correction approaches can correct the overall variability in a time series, and although methods have been developed to correct interannual variability (Johnson and Sharma 2011) in GCM precipitation time series, it is better to select models that inherently perform better in this regard.

d. Sensitivity of results to observational datasets

Our results and discussion in the previous sections are based on the assumption that the observational record is sufficiently long to capture interannual and interdecadal variability. We now examine this assumption through two alternative sensitivity tests, aiming at understanding the implications of record length and alternative data sources on our conclusions.

For both sensitivity tests, we have used SST anomalies because first the available record is the longest of the three observed datasets and second an alternative observed dataset is available, namely the NOAA/NCDC extended reconstructed SST (ERSST) anomalies (Smith et al. 2008). Figure 9 shows the impact on the estimates of global mean variance at each frequency using different periods for the analysis. In addition, the impact of using 32-yr record lengths instead of 64-yr lengths is examined. For the 64-yr length records we take 10-yr moving windows of 64 yr starting in 1901, 1911, 1921, and 1931 and compare these to the 1936–99 period used for the original analysis. We also use 9 moving windows of 32 yr starting from 1901 with the last 32-yr period starting in 1977.

The general pattern of variances for different frequencies is the same for all the different periods and two different record lengths. The subannual frequencies show the least variation among the different periods. Spatial plots of the global variation in percentage variance (not shown) show similar patterns for all periods, with the tropical Pacific having the highest percentage

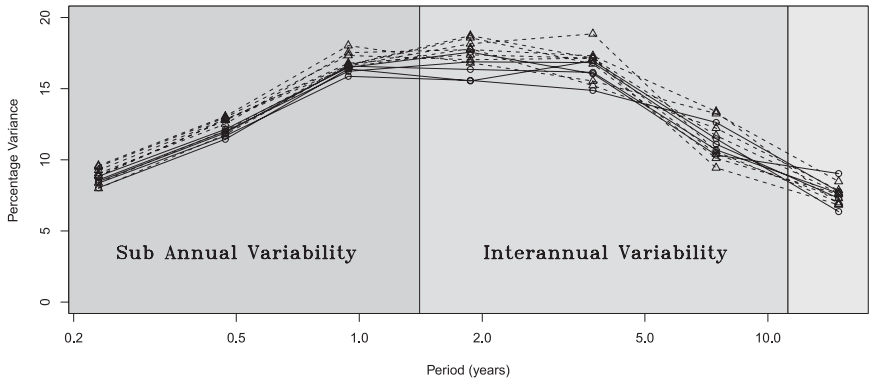

FIG. 9. Sensitivity of global mean variance to the use of different windows and 32-yr record lengths. Variances from 64-yr record lengths are shown with a solid line, and 32-yr records lengths with the dashed line.

variance in all cases for the interannual frequencies. Subannual frequencies show some variations in the percentage variance in the Southern and Atlantic Oceans. Figure 10 is a Taylor diagram showing the variation of the different periods of the observations relative to the original 64-yr period. The figure also shows the positions of the 60 GCMs, repeated from Fig. 4. The different analysis periods give variations compared to the original period that are similar in magnitude to the relationship between the reanalysis data and the original observations. There is a strong divide between the observations and the GCMs. The 64-yr records that are closest to the original data are those with the most overlap in the years analyzed; the closest has a skill score of 0.94 while the one with the least overlap has a skill score of 0.89. Also shown in Fig. 10 is the ERSST data. It falls within the range of the observations based on the Kaplan SST anomalies, with a correlation coefficient of approximately 0.9 and standard deviation of the spatial field that is similar in magnitude to the Kaplan SST anomalies. This gives us further confidence that the analysis method is insensitive to the length of record of the observations used to measure the skill of the GCMs.

4. Discussion

We have presented a skill score as a metric that can be used to assess the performance of GCMs in reproducing the observed interannual variability in climate variables of interest for water resources assessments. We now consider the utility of this metric of interannual variability in the context of model evaluation in general

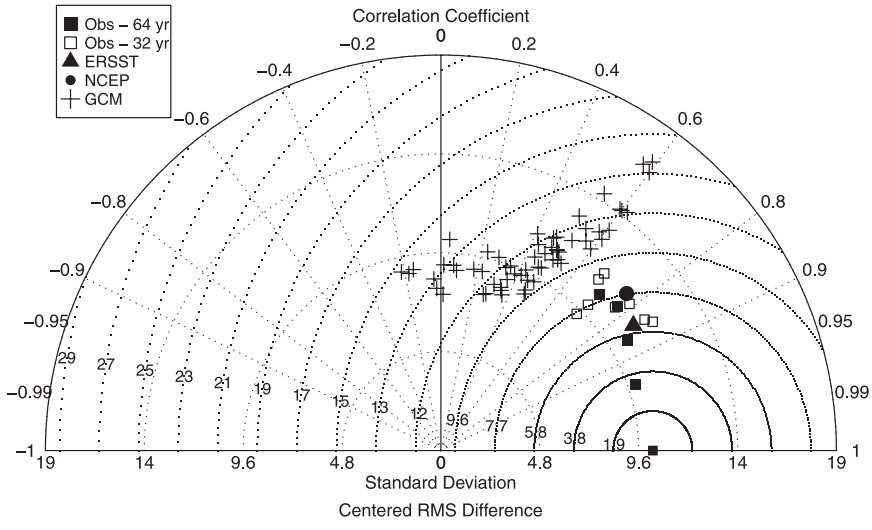

FIG. 10. Taylor diagram showing the sensitivity of the estimates to time period and observational dataset used for the global patterns of interannual variance percentage of SST anomalies.

and also by considering the advantages of our approach compared to previous studies that have evaluated the modeling of precipitation and teleconnections in GCMs.

The use of metrics and diagnostics to assess climate model outputs has been discussed by Gleckler et al. (2008) and Guilyardi et al. (2009). Metrics provide a measure of ''distance'' of a model output from an observation and are generally understood to refer to a single value that summarizes the distance in some appropriate space. Guilyardi et al. (2009) recommend that metrics ''should be concise, physically informative, societally relevant and easy to understand, compute and compare.'' Diagnostics, on the other hand, provide more information on model errors and the processes that may lead to these errors. They may take many forms including maps, time series, distributions, or power spectra. Whether a metric or diagnostic is more useful depends on the question and the user. For a climate change impact study, where we may only be able to use a subset of the available models, a metric that clearly tells us which are the best models to use for a specific application is preferable to a range of maps. But for climate modelers and process scientists, diagnostics help to improve models by highlighting processes and errors, although metrics can also be used to plot improvements (or deterioration) of different generations of a model (Gleckler et al. 2008).

There has been some historical reluctance in using a single metric to evaluate climate models because of the risk of reducing a complicated model of the climate system to a single number (Gleckler et al. 2008). It is clear that there is unlikely to be a single model that will provide the best simulations over all regions and all variables. For example, additional analysis of the results of the wavelet analysis over the ENSO regions (not presented here) indicates that the best performing GCMs will be different from those that were shown in section 3 to represent global interannual variability well. It is also important to note that ''good'' model performance in matching observed data does not provide a measure of the future reliability of a model; although the converse may be true, such that it is difficult to rely on the future simulations of a model that performs poorly in representing the twentieth-century climate. GCMs that have been shown to have similar skill in representing observations have been shown to have quite different climate sensitivities (Knutti 2008). Their responses to greenhouse gas emission scenarios will be therefore vary, which would have effects on the findings of impact assessments.

Despite these caveats, metrics can be prudently used for selecting models that are appropriate for a specific question. For water resources impact assessments, studies such as Sun et al. (2006), while helping to improve knowledge in general about flaws in climate models with respect to precipitation, do not have clear conclusions about what models perform the best, although a specific question (in this case the question of modeling daily rainfall intensity) is being considered. This may be changing as more studies start to make recommendations such as Perkins et al. (2007), who chose to identify individual models to help users of simulations ''determine models with particular strengths or weaknesses'' and second to allow modeling groups to improve their models.

How does the method proposed in this paper compare to previous studies evaluating GCM outputs? We consider here studies that have provided recommendations on the best-performing models for particular regions and using different evaluation metrics, generally related either to the modeling of precipitation or ENSO, all of which may aid in model selection for water resources climate change impact assessments. Perkins et al. (2007) limit their analysis to Australia but find that for precipitation that the best-performing models are the Bjerknes Center for Climate Research (BCCR) Bergen Climate Model version 2.0 (BCM2.0), MPI ECHAM5, and the Meteorological Institute of the University of Bonn, ECHO-G Model (MIUBECHOG). The criterion for evaluating the models in this study was the match between the probability density function (PDF) of daily rainfalls from the models and station-based observations. This methodology was motivated by the need to extend the evaluation of climate models from using just mean rainfall. However, by only examining one time scale for the PDFs the methodology does not account for biases in the GCMs at longer time scales, for example, seasonal cycle or interannual variations. Gleckler et al. (2008) and Murphy et al. (2004) use multiple variables to each create a combined evaluation index, with the rationale that ''the complexity of the models and the characteristics of the simulated fields cannot be adequately captured by a single measure of performance.'' However, given that it is clear that there is no ''best'' model and that different impact assessments require model assessments regarding a range of variables, it is hard to separate the information of interest from either of these studies. In addition, the use of monthly data (Gleckler et al. 2008) or seasonal data in the case of the climate prediction index (Murphy et al. 2004), while allowing the skill of the seasonal cycle simulation to be assessed, means that variability at other time scales is not considered. Gleckler et al. (2008) address this by proposing a model variability index, which is based on the ratio of the observed and modeled monthly variances, relative to the monthly climatology. However, they only

consider 20-yr windows, which may not properly incorporate the long time-scale variance and trends (30 yr and longer) found in some of the GCMs in this study.

Other studies assessing the modeling of ENSO in GCMs rely mainly on diagnostics to assess model performance. If users wish to pick the most suitable models for their purposes based on these diagnostics, the decisions have to be made by visual inspection. Guilyardi et al. (2009) use visual inspection to select the six GCMs that are closest to observations based on the wavelet spectra of Niño-3 SST anomalies derived by AchutaRao and Sperber (2006). A similar process could be carried out for other studies (AchutaRao and Sperber 2006; Cai et al. 2006; Joseph and Nigam 2006; Lin 2007) where we can compare, for example, maps or scatterplots to observation, but in all cases it requires a subjective decision on the best match between the observations and GCMs. The methodology presented here of the wavelet analysis combined with Taylor's (2001) proposed skill score has the advantage of leading to diagnostic maps and also an objective ranking of model performance that can be used to select models for climate change impact studies. The study could be further extended to include other variables known to influence precipitation (and its variability) such as precipitable water or outgoing longwave radiation as well as geopotential height and airflow fields (Timbal 2004; Mehrotra and Sharma 2005).

5. Conclusions

Model evaluation metrics are needed to assess the performance of GCMs over a range of variables and statistics. One important aspect, particularly for water resources applications, is how well persistence is simulated over a range of time scales. The wavelet-based skill score presented in this paper provides a useful measure of this model feature. The advantages of the approach include the ability to assess persistence on a range of different time scales and that no assumptions are required regarding the nature of the persistence in the data. We have chosen to report a skill score to summarize the wavelet variance maps to allow impact assessments to choose the best models for their purposes. The methodology can also be easily applied to find the best GCMs for smaller regions as demonstrated in the reservoir storage example provided in section 4.

Based on the skill score that incorporates the pattern correlation and overall variance in the global pattern of interannual variability, it was found that INGV ECHAM4 and IPSL-CM4 best match gridded observations for individual climatic variables, while MPI ECHAM5 has the best performance over the three variables studied. It was also found that models that are able to correctly represent variability in one climatic variable will also show good performance in other climate fields. This means we can use SSTs and surface pressure to better understand model performance in representing persistence in precipitation, which is of practical importance for water resources climate change impact assessments.

Acknowledgments. Funding for this research came from the Australian Research Council and the Sydney Catchment Authority. Their support for this work is gratefully acknowledged. The comments from three anonymous reviewers have greatly improved the presentation of the work. We acknowledge the modeling groups, the Program for Climate Model Diagnosis and Intercomparison (PCMDI) and the WCRP's Working Group on Coupled Modeling (WGCM) for their roles in making available the WCRP CMIP3 multimodel dataset. Support of this dataset is provided by the Office of Science, U.S. Department of Energy. NCEP reanalysis, NOAA ERSST V3, and GPCP data provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, from their Web site at http://www.cdc.noaa.gov/.

REFERENCES

AchutaRao, K., and K. R. Sperber, 2006: ENSO simulation in coupled ocean-atmosphere models: Are the current models better? *Climate Dyn.,* **27,** 1–15.

Adler, R. F., and Coauthors, 2003: The version-2 global precipitation climatology project (GPCP) monthly precipitation analysis (1979–present). *J. Hydrometeor.,* **4,** 1147–1167.

Barbosa, S., M. Silva, and M. Fernandes, 2009: Multi-scale variability patterns in NCEP/NCAR reanalysis sea-level pressure. *Theor. Appl. Climatol.,* **96,** 319–326.

Burrus, C. S., R. A. Gopinath, and H. Guo, 1998: *Introduction to Wavelets and Wavelet Transforms*. Prentice Hall, 268 pp.

Cai, W., D. Bi, J. Church, T. Cowan, M. Dix, and L. Rotstayn, 2006: Pan-oceanic response to increasing anthropogenic aerosols: Impacts on the Southern Hemisphere oceanic circulation. *Geophys. Res. Lett.,* **33,** L21707, doi:10.1029/2006GL027513.

——, A. Sullivan, and T. Cowan, 2009: Rainfall teleconnections with Indo-Pacific variability in the WCRP CMIP3 models. *J. Climate,* **22,** 5046–5071.

Cane, M. A., 1992: Tropical Pacific ENSO models: ENSO as a mode of the coupled system. *Climate System Modeling,* K. E. Trenberth, Ed., Cambridge University Press, 583–614.

Coelho, C. A. S., and L. Goddard, 2009: El Niño-induced tropical droughts in climate change projections. *J. Climate,* **22,** 6456–6476.

Collins, M., 2007: Ensembles and probabilities: A new era in the prediction of climate change. *Philos. Trans. Roy. Soc. London,* **A365,** 1957–1970.

Daubechies, I., 1992: *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics, 357 pp.

Dessai, S., X. F. Lu, and M. Hulme, 2005: Limited sensitivity analysis of regional climate change probabilities for the 21st century. *J. Geophys. Res.,* **110,** D19108, doi:10.1029/2005JD005919.

Fowler, H. J., and C. G. Kilsby, 2007: Using regional climate model data to simulate historical and future river flows in northwest England. *Climatic Change,* **80,** 337–367.

——, S. Blenkinsop, and C. Tebaldi, 2007: Linking climate change modelling to impact studies: Recent advances in downscaling techniques for hydrological modelling. *Int. J. Climatol.,* **27,** 1547–1578.

Giorgi, F., and L. O. Mearns, 2002: Calculation of average, uncertainty range and reliability of regional climate changes from AOGCM simulations via the ''Reliability Ensemble Averaging'' (REA) method. *J. Climate,* **15,** 1141–1158.

Gleckler, P. J., K. E. Taylor, and C. Doutriaux, 2008: Performance metrics for climate models. *J. Geophys. Res.,* **113,** D06104, doi:10.1029/2007JD008972.

Guilyardi, E., and Coauthors, 2009: Understanding El Niño in ocean–atmosphere general circulation models. *Bull. Amer. Meteor. Soc.,* **90,** 325–340.

Hawkins, E., and R. Sutton, 2009: The potential to narrow uncertainty in regional climate predictions. *Bull. Amer. Meteor. Soc.,* **90,** 1095–1107.

Huffman, G. J., R. F. Adler, D. T. Bolvin, and G. J. Gu, 2009: Improving the global precipitation record: GPCP version 2.1. *Geophys. Res. Lett.,* **36,** L17808, doi:10.1029/2009GL040000.

Jain, S., and U. Lall, 2001: Floods in a changing climate: Does the past represent the future? *Water Resour. Res.,* **37,** 3193–3205.

Johnson, F. M., and A. Sharma, 2011: Accounting for interannual variability: A comparison of options for water resources climate change impact assessments. *Water Resour. Res.,* **47,** W04508, doi:10.1029/2010WR009272.

Joseph, R., and S. Nigam, 2006: ENSO evolution and teleconnections in IPCC's twentieth-century climate simulations: Realistic representation? *J. Climate,* **19,** 4360–4377.

Kaplan, A., M. A. Cane, Y. Kushnir, A. C. Clement, M. B. Blumenthal, and B. Rajagopalan, 1998: Analyses of global sea surface temperature 1856–1991. *J. Geophys. Res.,* **103,** 18 567–18 589.

Knutti, R., 2008: Should we believe model predictions of future climate change? *Philos. Trans. Roy. Soc. London,* **A366,** 4647–4664.

Koster, R. D., and Coauthors, 2006: GLACE: The Global Land–Atmosphere Coupling Experiment. Part I: Overview. *J. Hydrometeor.,* **7,** 590–610.

Koutsoyiannis, D., 2002: The Hurst phenomenon and fractional Gaussian noise made easy. *Hydrol. Sci. J.,* **47,** 573–595.

Lau, N. C., 1992: Climate variability simulated in GCMs. *Climate System Modeling,* K. E. Trenberth, Ed., Cambridge University Press, 617–642.

Lin, J. L., 2007: Interdecadal variability of ENSO in 21 IPCC AR4 coupled GCMs. *Geophys. Res. Lett.,* **34,** L12702, doi:10.1029/2006GL028937.

Mantua, N. J., and S. R. Hare, 2002: The Pacific decadal oscillation. *J. Oceanogr.,* **58,** 35–44.

Mehrotra, R., and A. Sharma, 2005: A nonparametric nonhomogeneous hidden Markov model for downscaling of multisite daily rainfall occurrences. *J. Geophys. Res.,* **110,** D16108, doi:10.1029/2004JD005677.

——, and ——, 2010: Development and application of a multisite rainfall stochastic downscaling framework for climate change impact assessment. *Water Resour. Res.,* **46,** W07526, doi:10.1029/2009WR008423.

Met Office, cited 2009: Global mean sea-level pressure datasets (GMSLP and HadSLIP1). [Available online at http://badc.nerc.ac.uk/data/gmslp/.]

Murphy, J. M., D. M. H. Sexton, D. N. Barnett, G. S. Jones, M. J. Webb, M. Collins, and D. A. Stainforth, 2004: Quantification of modelling uncertainties in a large ensemble of climate change simulations. *Nature,* **430,** 768–772.

Peixoto, J. P., and A. H. Oort, 1992: *Physics of Climate*. American Institute of Physics, 520 pp.

Perkins, S. E., A. J. Pitman, N. J. Holbrook, and J. McAneney, 2007: Evaluation of the AR4 climate models' simulated daily maximum temperature, minimum temperature, and precipitation over Australia using probability density functions. *J. Climate,* **20,** 4356–4376.

Power, S., T. Casey, C. K. Folland, A. Colman, and V. Mehta, 1999: Inter-decadal modulation of the impact of ENSO on Australia. *Climate Dyn.,* **15,** 319–324.

Reichler, T., and J. Kim, 2008: Uncertainties in the climate mean state of global observations, reanalyses, and the GFDL climate model. *J. Geophys. Res.,* **113,** D05106, doi:10.1029/2007jd009278.

Saji, N. H., B. N. Goswami, P. N. Vinayachandran, and T. Yamagata, 1999: A dipole model in the tropical Indian Ocean. *Nature,* **401,** 360–363.

Santer, B. D., and Coauthors, 2009: Incorporating model quality information in climate change detection and attribution studies. *Proc. Natl. Acad. Sci. USA,* **106,** 14 778–14 783.

Simmons, A. J., and Coauthors, 2004: Comparison of trends and low-frequency variability in CRU, ERA-40, and NCEP/NCAR analyses of surface air temperature. *J. Geophys. Res.,* **109,** D24115, doi:10.1029/2004jd005306.

Smith, T. M., R. W. Reynolds, T. C. Peterson, and J. Lawrimore, 2008: Improvements to NOAA's historical merged land–ocean surface temperature analysis (1880–2006). *J. Climate,* **21,** 2283–2296.

Sun, Y., S. Solomon, A. Dai, and R. W. Portmann, 2006: How often does it rain? *J. Climate,* **19,** 916–934.

Taylor, K. E., 2001: Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res.,* **106,** 7183–7192.

Tebaldi, C., L. O. Mearns, D. Nychka, and R. L. Smith, 2004: Regional probabilities of precipitation change: A Bayesian analysis of multimodel simulations. *Geophys. Res. Lett.,* **31,** L24213, doi:10.1029/2004gl021276.

Timbal, B., 2004: Southwest Australia past and future rainfall trends. *Climate Res.,* **26,** 233–249.

Torrence, C., and G. P. Compo, 1998: A practical guide to wavelet analysis. *Bull. Amer. Meteor. Soc.,* **79,** 61–78.

Trenberth, K. E., 1997: The definition of El Niño. *Bull. Amer. Meteor. Soc.,* **78,** 2771–2777.

Ummenhofer, C. C., A. Sen Gupta, M. J. Pook, and M. H. England, 2008: Anomalous rainfall over southwest Western Australia forced by Indian Ocean sea surface temperatures. *J. Climate,* **21,** 5113–5134.

van Oldenborgh, G. J., S. Y. Phillip, and M. Collins, 2005: El Niño in a changing climate: A multi-model study. *Ocean Sci.,* **1,** 81–95.

Wasko, C., and A. Sharma, 2009: Effect of solar variability on atmospheric moisture storage. *Geophys. Res. Lett.,* **36,** L03703, doi:10.1029/2008GL036310.

Westra, S., and A. Sharma, 2006: Dominant modes of interannual variability in Australian rainfall analyzed using wavelets. *J. Geophys. Res.,* **111,** D05102, doi:10.1029/2005JD005996.

——, and ——, 2010: An upper limit to seasonal rainfall predictability? *J. Climate,* **23,** 3332–3351.

Wittenberg, A. T., 2009: Are historical records sufficient to constrain ENSO simulations? *Geophys. Res. Lett.,* **36,** L12702, doi:10.1029/2009gl038710.

Wood, A. W., L. R. Leung, V. Sridhar, and D. P. Lettenmaier, 2004: Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs. *Climatic Change,* **62,** 189–216.

Xu, C. Y., 1999: Climate change and hydrologic models: A review of existing gaps and recent research developments. *Water Resour. Manage.,* **13,** 369–382.