Probabilistic Estimation of Multivariate Streamflow Using Independent Component Analysis and Climate Information

SETH WESTRA AND ASHISH SHARMA

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

(Manuscript received 30 October 2008, in final form 14 May 2009)

ABSTRACT

A statistical estimation approach is presented and applied to multiple reservoir inflow series that form part of Sydney's water supply system. The approach involves first identifying sources of interannual and interdecadal climate variability using a combination of correlation- and wavelet-based methods, then using this information to construct probabilistic, multivariate seasonal estimates using a method based on independent component analysis (ICA). The attraction of the ICA-based approach is that, by transforming the multivariate dataset into a set of independent time series, it is possible to maintain the parsimony of univariate statistical methods while ensuring that both the spatial and temporal dependencies are accurately captured.

Based on a correlation analysis of the reservoir inflows with the original sea surface temperature anomaly data, the principal sources of variability in Sydney's reservoir inflows appears to be a combination of the El Niño–Southern Oscillation (ENSO) phenomenon and the Pacific decadal oscillation (PDO). A multivariate ICA-based estimation model was then used to capture this variability, and it was shown that this approach performed well in maintaining the temporal dependence while also accurately maintaining the spatial dependencies that exist in the 11-dimensional historical reservoir inflow dataset.

1. Introduction

Statistical forecasting of seasonal reservoir inflows plays an important part in reservoir management and operation, with applications including the regulation of municipal water supplies, allocation of irrigation water and environmental flows, and estimation of future hydroelectricity supplies (e.g., Lettenmaier and Wood 1993; Yao and Georgakakos 2001; Hamlet et al. 2002). These forecasts are usually developed using either a parametric or nonparametric regression or a probabilistic forecasting model (e.g., Sharma 2000a; Drosdowsky and Chambers 2001; Goddard et al. 2001; Chiew and McMahon 2002; Filho and Lall 2003), with predictors generally including a persistence component (e.g., reservoir inflows at the previous season; see Piechota et al. 2001), together with a set of exogenous predictors linked to a ''climate state.''

In the case of forecasting Australian rainfall and streamflow time series, the climate state is often represented by indices of the El Niño–Southern Oscillation (ENSO) phenomenon, which is regarded as the dominant source of interannual variability in much of Australia (Pittock 1975; Nicholls 1985; Ropelewski and Halpert 1987; Cordery and Opoku-Ankomah 1994; Nicholls et al. 1996; Chiew et al. 1998; Piechota et al. 1998; Power et al. 1998; Cordery and McCall 2000; Sharma et al. 2000; Sharma 2000b; Cai et al. 2001; Drosdowsky and Chambers 2001; Piechota et al. 2001; Drosdowsky 2002; Chiew et al. 2003; White et al. 2004; Dutta et al. 2006; Barros and Bowden 2008). Correlations with ENSO indices are usually higher for streamflow than for rainfall (Dutta et al. 2006), although the ENSO–streamflow relationship is somewhat weaker on the eastern coastal fringe, including in the Sydney catchment region considered in this paper, compared to other parts of Australia (e.g., Chiew et al. 1998).

In addition to this interannual variability, Australian rainfall and streamflow are subject to variability at decadal or longer time scales, which is often attributed to low-frequency variability in the Pacific Ocean represented by the Pacific decadal oscillation (PDO) (Mantua et al. 1997; Zhang et al. 1997; Mantua and Hare 2002) or a closely related index known as the interdecadal Pacific oscillation (IPO) (Power et al. 1999a,b), with the

Corresponding author address: Seth Westra, Water Research Centre, School of Civil and Environmental Engineering, University of New South Wales, Sydney, NSW 2052, Australia.
E-mail: s.westra@unsw.edu.au

DOI: 10.1175/2009JHM1121.1

modulating influence of this mode on Australian rainfall, streamflow, and flood risk, shown by Kiem et al. (2003) and Verdon et al. (2004).

Furthermore, Australian rainfall is also influenced by variability in the Indian Ocean, with correlations between the Indian Ocean dipole (IOD) (Saji et al. 1999; Saji and Yamagata 2003) and rainfall time series found largely in western and central Australia (Ashok et al. 2003; Drosdowsky 1993; Nicholls 1989), as well as in southeastern parts of the country (Drosdowsky 2002; Verdon and Franks 2005).

The disadvantage of the index approach for developing statistical forecasts of rainfall and/or streamflow has been highlighted in Westra et al. (2008), who suggest that indices provide only a one-dimensional representation of climate variability and therefore may not capture all the dominant sources of variability relevant for a given study region. Furthermore, since dominant sources of variability in the Pacific, Atlantic, and Indian Oceans are often strongly related to each other, the indices themselves may be highly mutually correlated, which complicates the inclusion of multiple climatic indices in a statistical model. For these reasons, in the present study we use the full global sea surface temperature anomaly (SSTA) data, with an a priori dimension reduction step to ensure that the data is in a form suitable for developing a probabilistic estimation model.

Here we present an application of a recently developed approach to developing seasonal forecasts of inflows at multiple locations (Westra et al. 2007, 2008) in the context of a major water supply system near Sydney, Australia. The nature of modeling such a large system necessitates that both temporal and spatial dependencies in the historical data are preserved and that the estimates are presented within a probabilistic framework so that estimation uncertainty can be ascertained. In line with these criteria, our approach involves (i) the identification of sources of variability for a particular region based on spectral properties of the historical reservoir inflow time series and on the correlation between reservoir inflows and measures of global climate and (ii) the development of a multivariate statistical model which links streamflow data at multiple locations (the response) to the global SSTA dataset (the predictor), with results presented in a probabilistic setting such that the full range of likely outcomes can be stochastically generated.

The remainder of this paper is structured as follows. In the following section we provide an overview of the predictor and response data used in this paper. In section 3 we highlight the complementary nature of using both correlation-based and spectral-based approaches for analyzing variability within a particular region. This information feeds into the analysis of section 4, where we present a multivariate approach to generating probabilistic seasonal streamflow estimates. Finally, a summary and conclusions are presented in section 5.

2. Data

a. Reservoir inflows

This study uses monthly inflows to the Sydney water supply system from 1909 to 2003, which were selected because they are minimally impacted by anthropogenic activities. This system is owned and operated by the Sydney Catchment Authority (available online at http://www.sca.nsw.gov.au/) and consists of nine reservoirs and two major demand zones, currently serving a population of approximately four million people. A schematic of the Sydney water supply system is presented in Fig. 1, and the 11 streamflow gauging stations used for the present analysis, together with the mean annual flows for each station, are presented in Table 1. The reservoir series exhibit little seasonality, with similar inflows across all seasons. In the analysis presented in subsequent sections, the data were aggregated to seasonal and annual inflows, with the seasons defined as autumn [March–May (MAM)], winter [June–August (JJA)], spring [September–November (SON)] and summer [December–February (DJF)].

b. Climate

A global SST anomaly (SSTA) dataset was obtained from a reconstruction of raw SST values using an optimal smoother, as described in Kaplan et al. (1998). The data is available on a 5° latitude by 5° longitude grid across the global ocean field, representing 1207 spatial locations. In the temporal dimension, the data comprises monthly averages, which we aggregated to seasonal and annual time series from 1909 to 2003 to correspond to the reservoir inflow data described in section 2a.

In addition to the global SSTA data, we consider a number of indices that describe aspects of global or regional climate variability. In the case of the Pacific Ocean, the ENSO phenomenon represents the dominant source of interannual variability, with teleconnections that often extend globally (e.g., Ropelewski and Halpert 1987; Trenberth 1997). Here we consider an index of the oceanic component of ENSO known as the Niño-3.4 index, defined as the seasonally averaged SSTA over the central Pacific Ocean (5°S–5°N, 120°–170°W; Trenberth 1997) and obtained from the International Research Institute for Climate and Society (IRI) Web site (available online at http://ingrid.ldgo.columbia.edu/SOURCES/.Indices/.nino/.KAPLAN). We also consider the Pacific

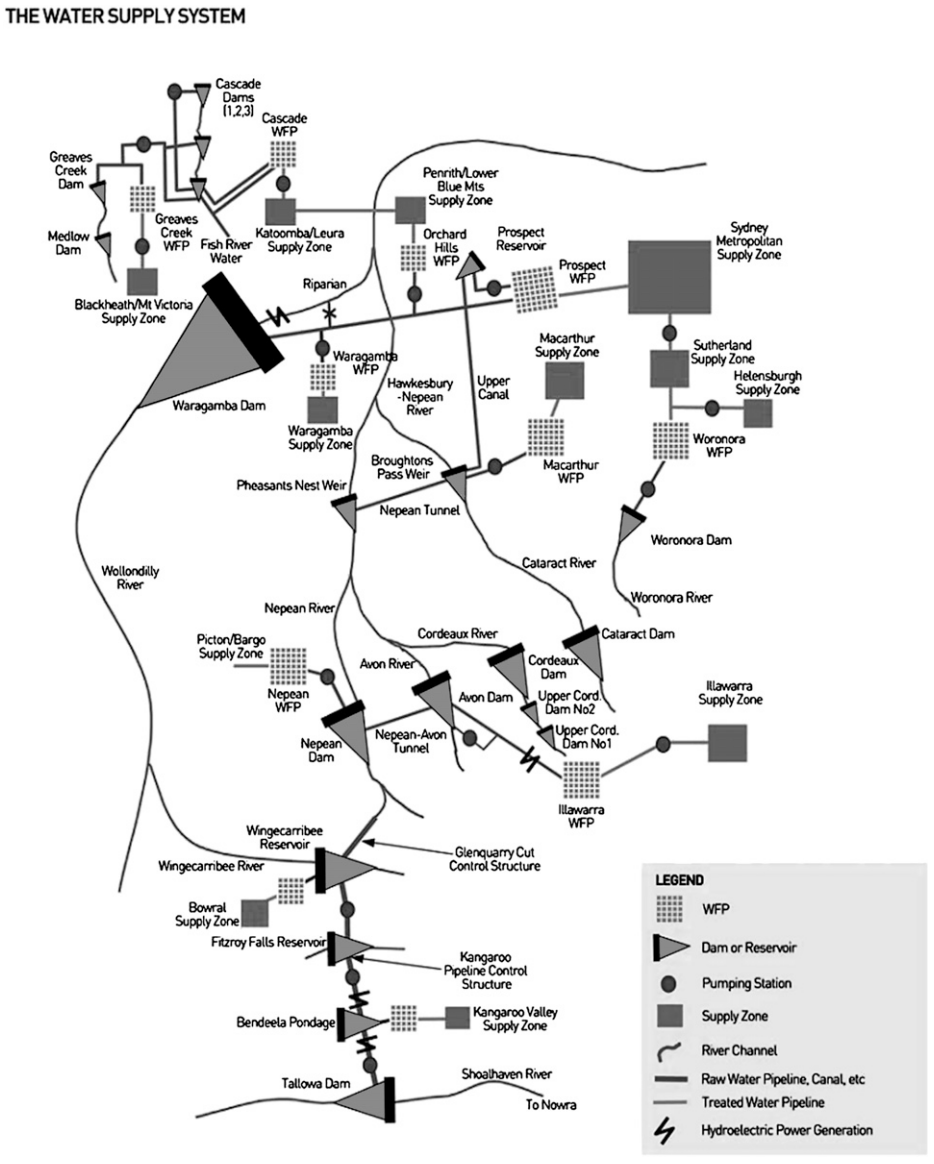

FIG. 1. The Sydney water supply system (available online at http://www.sca.nsw.gov.au).

decadal oscillation, which is derived as the leading principal component of monthly SSTAs in the North Pacific Ocean, poleward of 20°N. The monthly-mean global average SSTAs are removed to separate this pattern of variability from any global warming signal that may be present in the data (Mantua et al. 1997; Zhang et al. 1997; Mantua and Hare 2002). The data was obtained from the Department of Atmospheric Sciences' University of Washington Web site (available online at http://www.atmos.washington.edu/~mantua/abst.PDO.html).

Finally, in the Indian Ocean, we consider the Indian Ocean dipole (IOD), which is a coupled ocean atmosphere phenomenon in the Indian Ocean characterized by anomalous cooling in the southeastern equatorial Indian Ocean and anomalous warming in the western equatorial Indian Ocean. Here we use an index used by Saji et al. (1999) that represents the anomalous SST gradient between the western equatorial Indian Ocean (10°S–10°N, 50°–70°E) and the southeastern equatorial Indian Ocean (10°S–0°, 90°–110°E; index available online at http://www.jamstec.go.jp/frsgc/research/d1/iod/0).

3. Analyzing streamflow variability

Understanding sources of medium- to long-term climate variability is an important first step in generating seasonal forecasts for hydrological variables, such as

TABLE 1. List of streamflow stations.

Station	Name	Mean annual inflow (10^6 L)
1	Cataract Dam	84 236
2	Cordeaux Dam	56 590
3	Avon Dam	69 980
4	Nepean Dam	109 315
5	Woronora Dam	46 164
6	Warragamba Dam	1 110 347
7	Wingecarribee Reservoir	14 344
8	Fitzroy Falls Reservoir	16 832
9	Tallowa Dam	767 629
10	Welcome Reef	517 577
11	Penrith	196 650

rainfall and/or streamflow, since this information provides the basis for constructing and assessing the performance of the statistical models described in section 4. As discussed in the introduction, the dominant source of interannual variability in Australian rainfall and streamflow time series has been identified as the ENSO phenomenon, although recent evidence also suggests that eastern Australia is influenced by Indian Ocean variability as represented by the IOD, together with interdecadal modulation of ENSO as a result of the PDO.

The majority of studies examining the relationship between climatic phenomena and the hydrological variables of interest focus on correlation analyses between climate indices and time series of rainfall and/or streamflow. An alternative and complimentary approach considers dominant modes of variability for Australia-wide rainfall using the method of wavelets (Westra and Sharma 2006). In this analysis, it was shown that rainfall along the east coast of Australia, which had previously been shown to exhibit lower correlations with indices of ENSO compared to farther inland (Chiew et al. 1998), did not exhibit statistically significant periodicity in the 2–8-yr band typically associated with ENSO. Rather, a dominant period of 13 yr was observed, and statistically significant variability at this frequency was noted even after removing the influence of ENSO from the rainfall time series, indicating that this variability was largely independent of ENSO.

The results of these studies therefore suggest that, for the east Australian coastal region of interest for this analysis, some questions remain regarding the sources of variability, such as, how strong is the connection between Sydney's water supply and the ENSO phenomenon? Is the dominant 13-yr period identified in rainfall time series also present in streamflow data within the Sydney catchment area? and, if so, what is the source of such decadal variability? These questions will be discussed in more detail in the following sections.

a. Wavelets analysis of streamflow

Wavelets are a popular mathematical tool for the analysis of time series that have nonstationary power at a range of frequencies and, as such, are well suited to hydroclimatic time series such as rainfall and streamflow (e.g., Jain and Lall 2001). The benefit of using wavelets is that they allow a time series to be decomposed into both time and frequency components so that it is possible to capture not only the dominant frequency modes of a given time series but also how these modes change over time (Wang and Wang 1996).

Following the method described in Torrence and Compo (1998), we consider the wavelet spectra of individual and aggregate reservoir inflows at both seasonal and annual time scales. We used the Morlet wavelet as our basis function because this function is well localized in both time and frequency (Jevrejeva et al. 2003). Furthermore, the resampling logic described in Westra and Sharma (2006) was adopted for the generation of confidence intervals since this did not require the data to follow a Gaussian distribution.

The standardized aggregate annual inflows are presented in Fig. 2a. As can be observed from this time series, the data is highly skewed and is dominated by relatively low flows interspersed with a number of high-flow events three or more standard deviations away from the mean. This highlights the importance of generating confidence intervals using Monte Carlo methods for this data since the analytical approach developed in Torrence and Compo (1998) assumes a Gaussian distribution. The wavelet power is presented in Fig. 2b with the integrated spectra in time (global wavelet spectra) presented in Fig. 2c. The confidence intervals are at the 95% significance level, presented as a black contour in Fig. 2b and as a dotted blue line in Fig. 2c.

The results of the global wavelet spectra in Fig. 2c show a dominant period of approximately 13 yr, consistent with the dominant 13-yr period for Australian rainfall in this region found in Westra and Sharma (2006). Examining Fig. 2b, it is evident that this mode of variability is particularly pronounced from about 1950 onward, suggesting that since this time the high-flow events have been occurring every 13 yr. Interestingly, examination of the inflows for individual seasons and individual reservoir inflows (not shown) provide comparable results when compared with the aggregate annual statistics presented here.

We now compare the spectrum of the total annual streamflow time series with the spectrum of the Niño-3.4 time series, which is known to vary with a frequency in the 2–8-yr band. Although some research suggests the ENSO phenomenon may also vary at longer periods,

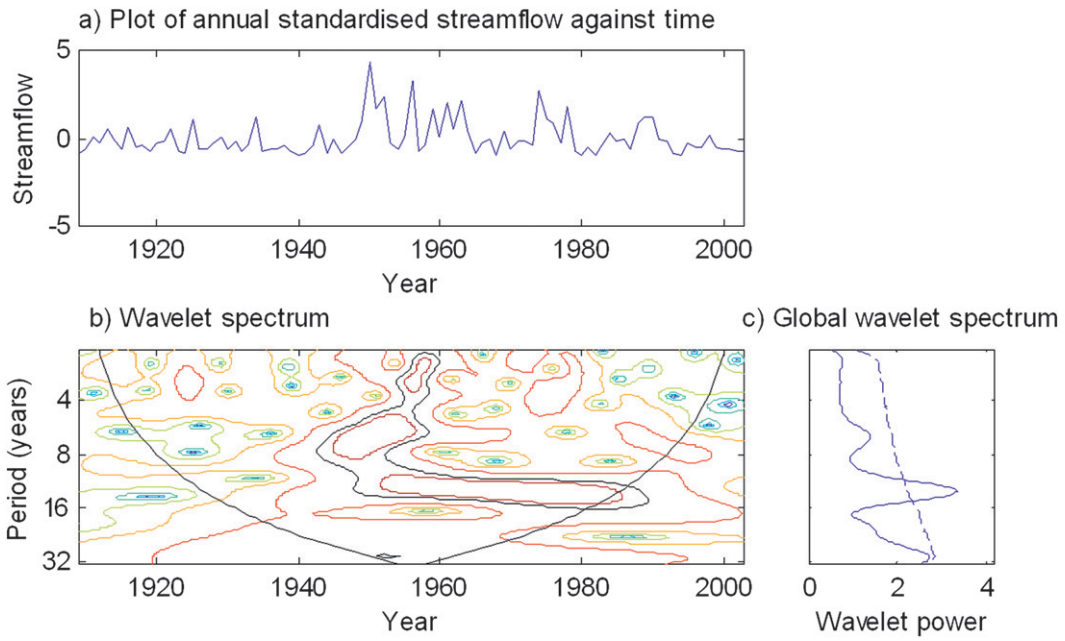

FIG. 2. Wavelet spectrum of total annual inflows comprising (a) time series of normalized streamflow from 1909 to 2003; (b) wavelet spectrum; and (c) the global wavelet spectrum defined as the integration of the wavelet spectrum over time. The cone of influence is shown as a black curve in (b), with confidence intervals at the 95% significance level shown as a black contour in (b) and dashed blue line in (c).

such a 13.9-yr cycle identified by Jevrejeva et al. (2004), it is not clear why the ENSO phenomenon should only influence streamflow variability at lower frequencies but not within the dominant 2–8-yr frequency band. Furthermore, the earlier study of rainfall variability in this region by Westra and Sharma (2006) showed that the 13-yr frequency was still present after removing the influence of ENSO from the rainfall time series, suggesting that this variability is unlikely to be captured directly using a single index of ENSO, such as the Niño-3.4 index.

Given the decadal nature of the total streamflow spectra, we then compare the spectra with the (inter)decadal mode of variability in the extratropical Pacific Ocean, which is represented by the Pacific decadal oscillation and provided in Fig. 3. Based on the global wavelet spectrum of Fig. 3c, the PDO oscillates with a 24-yr period, although it is difficult to provide a high level of confidence owing to the relatively short duration of the time series. Once again, this frequency does not correspond directly to the 13-yr period found in the Sydney reservoir inflow data.

Based on the analysis thus far, it is not possible to draw definitive conclusions on the link between Sydney reservoir inflow data and either the ENSO phenomenon or the PDO. To shed more light on the nature of this variability, we now consider correlation between reservoir inflows (both annual and seasonal) and gridded global SSTA data.

b. Comparison with oceans

The correlation plots between annual inflows and concurrent global SSTAs are provided in Fig. 4. The contours represent correlation coefficients, at intervals of 0.1 units, with blue (red) representing negative (positive) correlations. In Fig. 4, it is evident that the annual reservoir inflows are negatively correlated with tropical Pacific SSTs centered at 5°S, 175°W and positively correlated with extratropical Pacific SSTs focusing on the North Pacific poleward of 30°N and centered on 40°N, 180°. The maximum correlation coefficients at the center of these regions are of the order of ± 0.4. Examining the remaining oceans, a relatively small negative correlation can be observed in the Indian Ocean, with correlations of the same sign over the whole ocean suggesting the absence of a dipole moment. In the Atlantic, both positive and negative correlations can be observed, which, given the geographic distance from Sydney, are most likely to be due to teleconnections between Pacific and Atlantic modes of variability.

The correlation maps for each season were also computed individually (but not presented here) and generally show very similar patterns to the annual maps. In autumn, winter, and summer almost identical ocean patterns can be observed, with correlation coefficients between reservoir inflows and SSTA data of a very similar magnitude to the annual data. The main difference

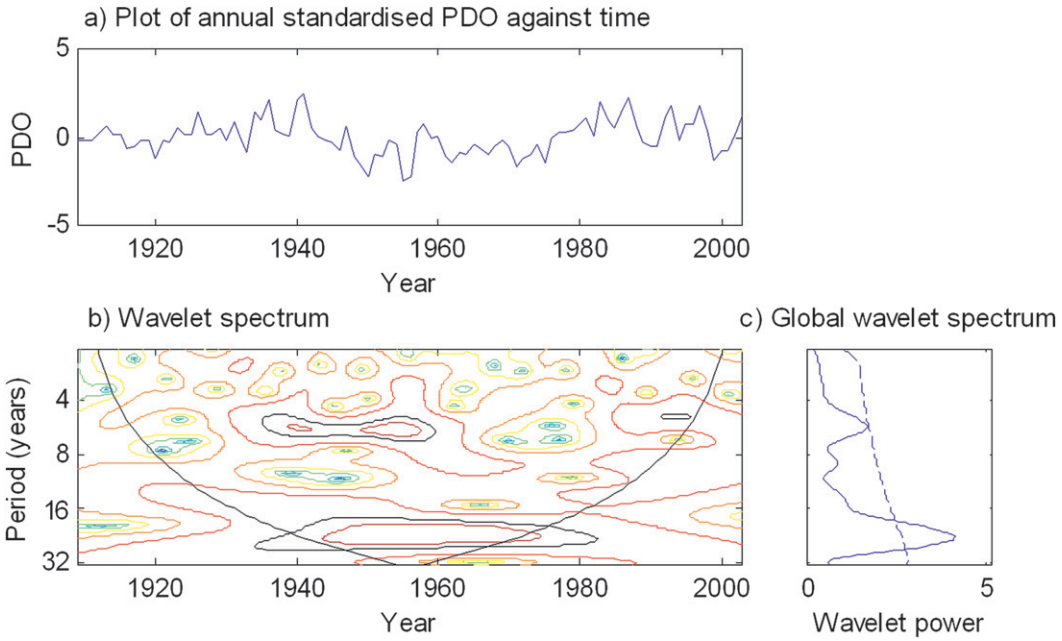

FIG. 3. Wavelet analysis of annual values of the Pacific decadal oscillation (PDO) comprising (a) normalized PDO series from 1909 to 2003; (b) wavelet spectrum; and (c) the global wavelet spectrum defined as the integration of the wavelet spectrum over time. Cone of influence as in Fig. 2.

can be observed for spring when the correlation coefficients are consistently lower than for the other seasons. The qualitative patterns (negative correlations in the tropical Pacific and the Atlantic Oceans and positive correlations in the extratropical Pacific), however, are the same.

These patterns are consistent with patterns related to both the interannual and the (inter) decadal variability

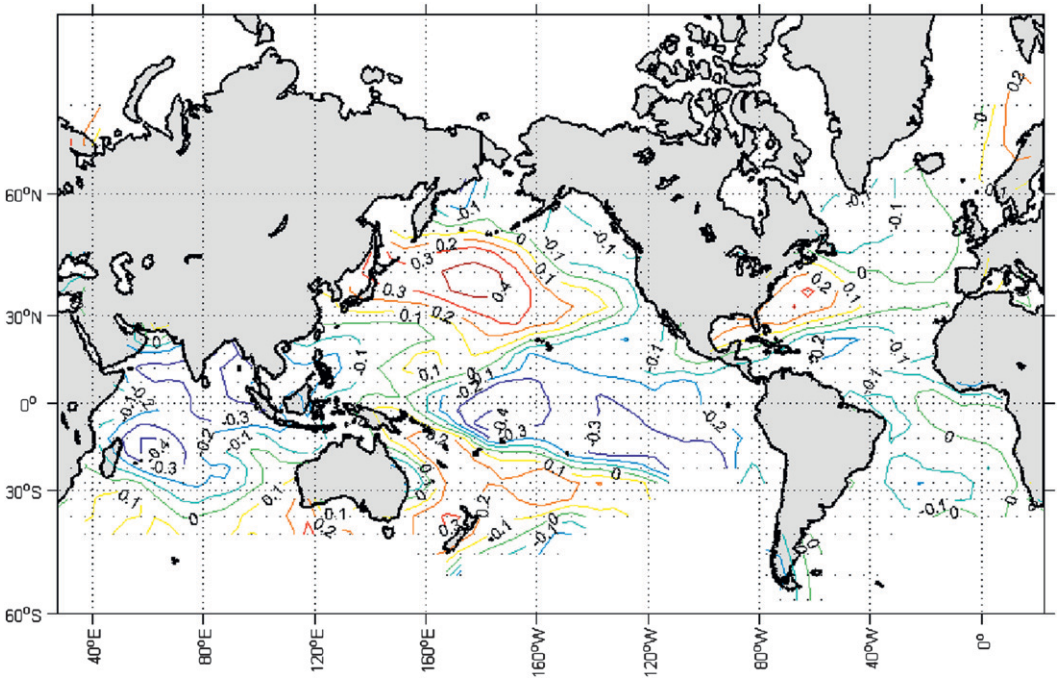

FIG. 4. Map of correlation coefficients between aggregate annual inflows and annual global SSTs. Correlations are with concurrent annual and seasonal SSTs.

TABLE 2. Correlation coefficients between annual and seasonal aggregate reservoir inflows for various climate indices.

	Index		
	Niño-3.4	PDO	IOD
Annual	−0.32	−0.43	−0.02
Autumn	−0.35	−0.26	−0.08
Winter	−0.27	−0.30	0.01
Spring	−0.17	−0.19	0.07
Summer	−0.29	−0.40	0.21

in the Pacific Ocean [e.g., compare these maps to PC2 and R-PC1 in Figs. 1 and 2 of Westra et al. (2009), which are strongly correlated to both the ENSO phenomenon and the PDO]. One important difference between Fig. 4 and the typical representation of the ENSO phenomenon is that, whereas ENSO is dominated by variability in the central tropical Pacific Ocean, our results show approximately equal importance placed on the tropical and extratropical Pacific. This again provides qualified support that the Sydney reservoir inflows are more influenced by the lower-frequency variability of the extratropical Pacific Ocean compared with the tropical ENSO phenomenon.

c. Comparison with indices

We now examine the correlations between Sydney reservoir inflows and a number of climate indices, including Niño-3.4, the PDO, and the IOD, with correlation coefficients presented in Table 2. Considering first the Niño-3.4 index, the correlation coefficients are generally relatively low, with a value of −0.32 for the annual series. This concurs with the preceding analysis, indicating correlation coefficients between reservoir inflows and eastern equatorial Pacific were typically in the range of −0.3 to −0.4. The PDO shows slightly higher correlation coefficients—up to −0.43 for the annual data. These correlation coefficients should be contrasted with statistically significant confidence bounds of ±0.2 at the 5% significance level. In both cases, the correlation coefficients are lowest for the spring season, which reflects the results of section 3b that show correlation coefficients between aggregate reservoir inflows and global SSTAs are lowest in spring.

In contrast to results from the Pacific Ocean, correlation coefficients between reservoir inflows and the IOD are much lower, with an annual value of −0.02. The only season that is statistically significant is summer, with a correlation coefficient of 0.21.

4. Developing estimates of multivariate streamflow

a. Approach

The estimation approach used for the present study is similar to the method presented in Westra et al. (2008), which was developed for the generation of seasonal forecasts in the multivariate setting. The focus of the research is on the development of multivariate predictor–response relationships that maintain the appropriate joint dependencies in the response field and has been illustrated using concurrent relationships between reservoir inflows and SSTAs. Development of a forecast model can be achieved either by developing statistical or dynamical forecasts of the future evolution of the SSTA field, or by assuming lagged predictor–response relationships.

This approach uses a technique known as independent component analysis (ICA) to transform an n-dimensional multivariate dataset ($\mathbf{X}$) of length l into a set of n-dimensional independent univariate time series ($\mathbf{S}$) also of length l, which are related by

$$\mathbf{S} = \mathbf{A}^{-1}\mathbf{X}, \tag{1}$$

where $\mathbf{A}$ is an $n \times n$ orthogonal matrix (throughout this paper we use upper case bold san serif for matrices and lower case bold regular for vectors). In cases where the dimension of $\mathbf{X}$ is large, an intermediate dimension reduction step using principal component analysis (PCA) is often incorporated into the algorithm such that a higher m-dimensional data matrix $\mathbf{X}$ is transformed into an n-dimensional subspace $\mathbf{X}_{\mathrm{PCA}}$ prior to the application of the ICA model of Eq. (1). Further details related to this dimension reduction step are provided in Westra et al. (2008).

The benefit of finding the independent components $\mathbf{S}$ for the present study lies in the property that the joint density of the independent components can be factorized into the product of their marginal densities:

$$f(\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_n) = f_1(\mathbf{s}_1) f_2(\mathbf{s}_2) \ldots f_n(\mathbf{s}_n), \tag{2}$$

where $f_i(\mathbf{s}_i)$, $i = 1, \ldots, n$ represents the marginal probability density function of each $\mathbf{s}_i$ and $f(\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_n)$ represents the joint probability density function of all $\mathbf{S}$. In practical terms, after the ICA transformation is found, it is possible to develop univariate statistical models for each $\mathbf{s}_i$ without needing to consider joint dependencies of the other independent components in $\mathbf{S}$. The benefits of this approach for statistical modeling are illustrated through a synthetic example presented in Fig. 3 of Westra et al. (2007).

The objective of ICA therefore is to find a matrix $\mathbf{W}$, which is an estimator of the inverse of the true mixing matrix $\mathbf{A}^{-1}$, through

$$\mathbf{Y} = \mathbf{W}\mathbf{X}, \tag{3}$$

where $\mathbf{Y}$ represents an estimator of the independent components $\mathbf{S}$. The estimated independent components

$\mathbf{Y}$ can be found up to a scalar multiple by finding a $\mathbf{W}$ that maximizes the non-Gaussianity of the $\mathbf{y}_i$, with further discussion on the links between non-Gaussianity and independence in Hyvärinen et al. 2001 and Westra et al. 2007. For the present study we use kurtosis as a measure of non-Gaussianity so that we now simply need to find the orthogonal rotation $\mathbf{W}$ that maximizes the kurtosis of the $\mathbf{y}_i$.

The algorithm adopted for the present study can now be presented as follows:

ALGORITHM: STATISTICAL ESTIMATION OF MULTIVARIATE STREAMFLOW

Step 1: We start by developing a model of reservoir inflow persistence at each location i as follows:

$$\mathbf{x}_{q,i,t} = \beta_i^{(1)} \mathbf{x}_{q,t-1} + \boldsymbol{\varepsilon}_{i,t}^{(1)}, \tag{4}$$

for $i = 1, \ldots, m$, where the subscript q represents reservoir inflows, t and $t - 1$ represent the present and preceding season, and $\boldsymbol{\varepsilon}_{i,t}^{(1)}$ represents the model residual. The above model is developed to simulate Markovian dependence; however, rather than a standard autoregressive process that considers the previous time step of a given time series, this model uses a common predictor of previous season aggregate reservoir inflows across all locations. As such, a subscript i is not included in $\mathbf{x}_{q,t-1}$. Furthermore, whereas a Markov process considers the previous time step in a single time series, this model considers each season separately, such that $\mathbf{x}_{q,i,t}$ may represent, say, a time series of summer inflows and $\mathbf{x}_{q,t-1}$ would represent the time series of aggregate spring inflows.

Step 2: We now rotate the residuals $\boldsymbol{\varepsilon}_{i,t}^{(1)}$ from step 1 to obtain a set of independent components that can be used to generate a statistical model linking these residuals to climate state. The rotation is performed as follows:

$$\mathbf{Y}_{q(\varepsilon),t} = \mathbf{W}_{q(\varepsilon),t} \mathbf{E}_t^{(1)}, \tag{5}$$

where $\mathbf{E}_t^{(1)}$ represents the m-dimensional multivariate error term from step 1 and $\mathbf{Y}_{q(\varepsilon),t}$ represents the n-dimensional estimate of the independent components, with the subscript $q(\varepsilon)$ indicating that calculations are performed on the error term of Eq. (4). We use the FastICA algorithm (available online at http://www.cis.hut.fi/projects/ica/) to derive $\mathbf{W}$, with details provided in Hyvärinen et al. (2001). As discussed in relation to Eq. (1) and in more detail in Westra et al. (2008), in certain cases it is necessary to reduce the dimension of the predictand [in this case $\mathbf{E}_t^{(1)}$ from Eq. (4)] to ensure the ICA algorithm is able to find a satisfactory solution. In these cases, an additional error term $\mathbf{E}_t^{(2)}$ of dimension m is included and represents the variance of $\mathbf{E}_t^{(1)}$ not explained by the reduced dimension subset $\mathbf{E}_{t,\mathrm{PCA}}^{(1)}$. As PCA searches for directions of maximum variance, the residuals $\mathbf{E}_t^{(2)}$ are expected to be largely random noise and are unlikely to be coherent spatially. As such this term can factorize into individual $\boldsymbol{\varepsilon}_{i,t}^{(2)}$ such that, when generating probabilistic estimates of the residuals, it is possible to treat error terms as univariate time series.

Step 3: Now that the independent components have been found, it is possible to develop a regression model that links the response to a set of predictors that represents climate state. Such a model may be linear, nonlinear, or nonparametric, although for the present case we focus only on the linear model, which can be written as

$$\mathbf{y}_{q(\varepsilon),i,t} = \sum_j \beta_j^{(2)} \mathbf{y}_{s,j,t} + \boldsymbol{\varepsilon}_{i,t}^{(3)}, \tag{6}$$

where the $\mathbf{y}_{s,j,t}$ represents the set of $j = 1, \ldots, p$ predictors, which for the present case are the independent components of the global SSTA dataset as indicated by the subscript s and were derived in a similar fashion to the independent components of the streamflow residuals of step 2. We use concurrent SSTA data for the present case, and to develop forecasts it would be necessary to modify this step by using either forecasts of the SSTA field or, alternatively, to use SSTA data at previous time steps to develop lagged relationships. One important constraint to this model is that each predictor $\mathbf{y}_{s,j,t}$ cannot be used for more than one $\mathbf{y}_{q(\varepsilon),i,t}$, since it is necessary that the estimated $\hat{\mathbf{y}}_{q(\varepsilon),i,t}$ are mutually independent. The approach used to ensure this is that predictor ICs are selected with a best subset selection approach and removed from the available set of predictors, such that a given $\mathbf{y}_{s,j,t}$ cannot be used for more than one $\hat{\mathbf{y}}_{q(\varepsilon),i,t}$. More details on predictor selection can be found in Westra et al. (2008).

Step 4: The $\hat{\mathbf{y}}_{q(\varepsilon),i,t}$ can now be rotated back to the original space using $\mathbf{W}_{q(\varepsilon)t}^{-1}$ to produce $\hat{\boldsymbol{\varepsilon}}_{i,t}^{(1)}$, and this can be added onto the fitted persistence model [Eq. (4)] in step 1 to yield the estimator of streamflow $\hat{\mathbf{x}}_{q,i,t}$. To generate results in a probabilistic setting such that confidence limits can be estimated, one simple and intuitive approach is to bootstrap (with replacement) the error terms $\boldsymbol{\varepsilon}_{i,t}^{(2)}$ and $\boldsymbol{\varepsilon}_{i,t}^{(3)}$, which owing to the construction of the model can be treated as univariate. Furthermore, the time between samples in both the predictor and response dataset is one year, approximately the same as the decorrelation time for global sea surface temperatures (Goddard et al. 2001), so the effects of autocorrelation are not likely to be significant.

In the application of Sydney's reservoir inflows, m represents the 11 inflow stations described in Table 1, and the reduced dimension n was selected to be three based on results of previous work (Westra et al. 2009), where it was found that, for geophysical data of the length used in the present study, a dimension of three resulted in a sensible compromise between maximizing the variance explained while generating components that are physically meaningful. This additional dimension reduction step may represent a limitation of the ICA technique for very high dimensional problems where spatial correlation is not strong; however, in our case the reduced dimension explained approximately 90% of the total streamflow variance and therefore is not expected to significantly degrade performance. The seasons were treated separately so that each time series consisted of 95 data points (e.g., spring inflows from 1909 to 2003). The only other preprocessing step involved standardizing each reservoir inflow time series.

b. Results

The performance of the model described in section 4a should be measured by looking at both the capacity of the model to provide forecasts of future inflows (the temporal performance) as well as the capacity of the model to maintain the historical dependence between sites (the spatial performance). These are discussed in turn below.

1) TEMPORAL PERFORMANCE

The temporal performance is measured by examining the correlation coefficients between historical inflows and ''hindcasts,'' developed by applying the model in section 4a to the historical data. Cross-validation was not required owing to the parsimonious nature of the model, which was fitted on 95 years of data and has as its predictors the previous season's aggregate inflows, and between 0 and 2 predictors of climate state for each independent component. It is furthermore noted that for the purposes of the present paper, we are using SSTA data with concurrent reservoir inflows to develop the model. The concurrent analysis is considered to be sufficient to compare the performance of statistical approaches in linking information contained in the predictor (SST) dataset with the response (streamflow) dataset and to assess the ability of the model to maintain the spatial dependence in the multivariate response. In a true forecast setting, it would be necessary to use either SSTA data, which themselves have been forecast (e.g., Colman and Davey 2003; Van den Dool et al. 2003), or alternatively to use the longer decorrelation times of the SSTA dataset (Goddard et al. 2001) to generate lagged relationships.

The correlation coefficients for the model applied to the historical data are provided in Table 3 for (i) the persistence or autoregressive order-one (AR1) model; (ii) the ICA model using global SSTAs as measures of climate state; and (iii) the full model that combines persistence with the SSTA-based predictors. The results should be compared with a statistically significant correlation coefficient of 0.2 at the 5% significance level, assuming 95 data points and no more than three predictors.

Considering first the persistence-only model in Table 3a, the greatest persistence is evident in winter with an average correlation coefficient across all sites of 0.44, followed by spring (0.35), and then autumn (0.25). The persistence from spring to summer is generally low, with an average correlation coefficient of 0.12, which is not statistically significant at the 5% level. Second, we consider the ICA model with global SSTAs as predictors in Table 3b. In contrast to the persistence-only model, the best performance occurs for summer with a correlation coefficient of 0.32, followed by spring (0.24), winter (0.21), and finally autumn (0.12). Last, the results for the full forecast are provided in Table 3c, and it can be seen that the best performance is in winter with a mean correlation coefficient of 0.49, followed by spring (0.41), summer (0.35), and autumn (0.28).

To benchmark the performance of our model, we compare these correlation coefficients with the correlation coefficients between the aggregate seasonal inflows and indices of climate presented in Table 2. With the exception of the Niño-3.4 index in autumn and the PDO in summer, our model appears to slightly outperform the index-based approach. Also, examining correlation coefficients between aggregate seasonal inflows and the SSTA dataset presented in Figs. 4b–e, it can be seen that the highest correlation coefficients are approximately 0.4 in both the equatorial and extratropical Pacific Ocean. As a result, we believe the performance of the full ICA model to be comparable to other linear regression-based models that do not take spatial dependence into account.

The forecast performance for three time series—Cordeaux Dam, Nepean Dam, and Warragamba Dam—is shown in Fig. 5. Here the forecasted time series are represented by the solid gray line, with 5% and 95% confidence limits presented as dashed gray lines. The historical time series from 1909 to 2003 is presented as a solid black line. Although there is significant difference between the historical and forecast data, as expected the historical time series is within the 90% confidence interval for the majority of the time. Furthermore, when considering the peaks in historical streamflow variability, it can be seen that the forecast data is able to capture the direction of change (e.g., above average or below average inflows), although the forecasts generally underestimate the magnitude of the extreme historical inflows.

TABLE 3. Correlation coefficients between estimated reservoir inflows and historical reservoir inflows between 1909 and 2003.

Station	Autumn	Winter	Spring	Summer
	(a) Estimated with the persistence (AR1)-only component in Eq. (4)			
1	0.18	0.40	0.26	0.12
2	0.28	0.42	0.26	0.10
3	0.22	0.38	0.23	0.06
4	0.30	0.51	0.39	0.12
5	0.22	0.45	0.27	0.07
6	0.30	0.55	0.49	0.18
7	0.19	0.39	0.30	0.16
8	0.14	0.47	0.25	0.01
9	0.17	0.38	0.35	0.07
10	0.29	0.45	0.48	0.23
11	0.50	0.49	0.54	0.23
Mean	0.25	0.44	0.35	0.12
	(b) Estimated with the residual component using ICA in Eq. (4)			
1	0.12	0.16	0.19	0.31
2	0.12	0.18	0.24	0.33
3	0.08	0.17	0.23	0.29
4	0.11	0.15	0.17	0.36
5	0.03	0.18	0.20	0.33
6	0.09	0.24	0.36	0.33
7	0.14	0.42	0.25	0.31
8	0.16	0.08	0.17	0.30
9	0.19	0.29	0.31	0.37
10	0.11	0.21	0.32	0.40
11	0.15	0.18	0.15	0.19
Mean	0.12	0.21	0.24	0.32
	(c) Estimated with both the persistence (AR1) and residuals components using ICA in Eq. (4)			
1	0.22	0.43	0.32	0.33
2	0.31	0.45	0.34	0.35
3	0.23	0.42	0.32	0.30
4	0.31	0.52	0.41	0.38
5	0.21	0.48	0.33	0.34
6	0.31	0.58	0.56	0.37
7	0.23	0.53	0.43	0.35
8	0.21	0.47	0.29	0.30
9	0.25	0.46	0.45	0.37
10	0.30	0.49	0.55	0.46
11	0.51	0.52	0.55	0.30
Mean	0.28	0.49	0.41	0.35

2) SPATIAL DEPENDENCE

We now examine the spatial performance of the model. As discussed earlier, the primary application of the ICA-based approach is that, by splitting the multivariate data into a set of univariate series that do not exhibit dependence on one another, it is possible to develop a model that performs comparably with univariate methods in the temporal dimension while simultaneously maintaining the spatial dependence in the historical data.

To test for spatial performance we use our model to create 100 random samples, each with 95 data points, by bootstrapping both $\varepsilon_{i,t}^{(2)}$ and $\varepsilon_{j,t}^{(3)}$ in Eqs. (5) and (6), respectively. This allows us to examine the full probabilistic set of multivariate model outcomes. We then develop a kernel density estimate to examine the joint density of the data, following the methods of Sharma et al. (1998) and Sharma (2000b) and using a Gaussian reference bandwidth. Before applying the kernel density estimate, we applied a log transform to the synthetically generated data followed by standardization.

Although the estimation model maintains spatial dependence across all 11 dimensions, we considered only bivariate kernel density estimates, both for ease of visualization and also owing to the computational difficulty of estimating kernel density estimates in 11 dimensions (consider, e.g., that to develop probability density functions that integrate to 1, we needed to use a grid size of 80

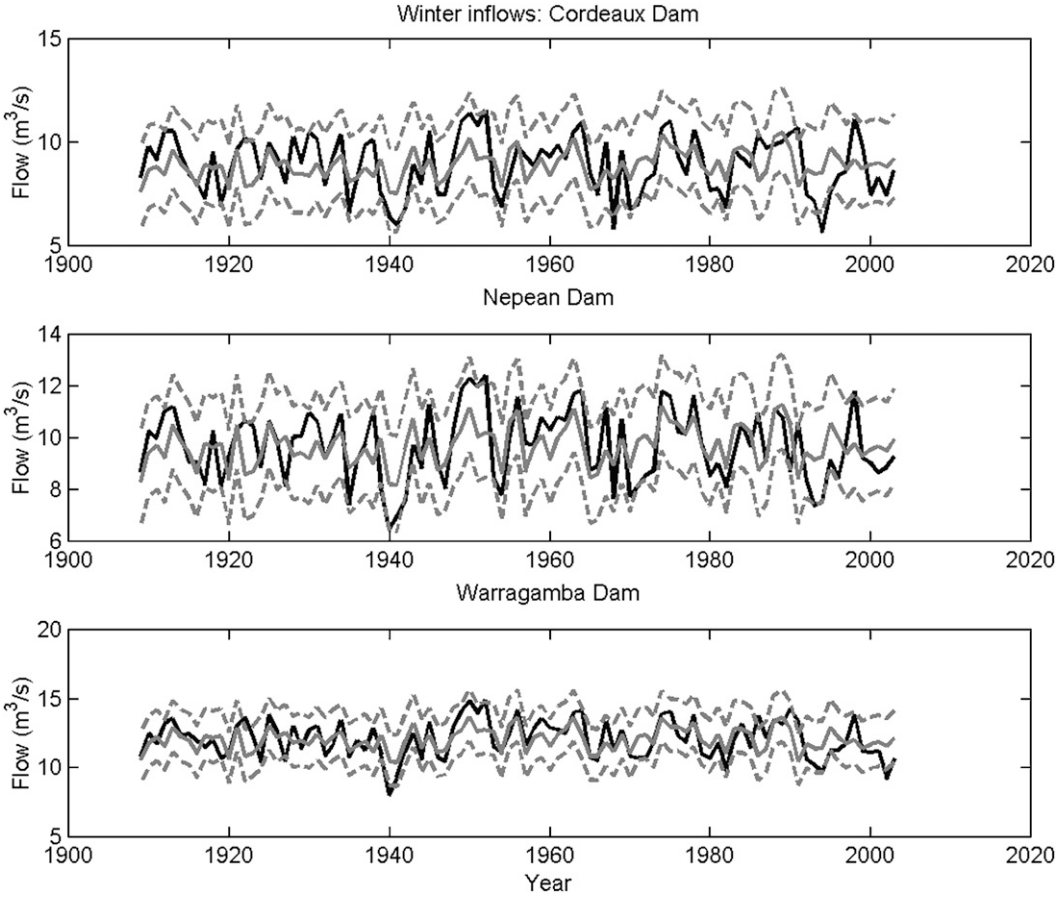

FIG. 5. Time series of historical flows (black line) and estimated flows based on multivariate ICA regression model (gray solid line) for inflows to (top) Cordeaux Dam, (middle) Nepean Dam, and (bottom) Warragamba Dam. Dotted gray lines represent 5% and 95% significance levels.

to estimate the pdf and extending that to 11 dimensions would require a grid of 80^{11}).

The results of one such bivariate plot are presented in Fig. 6 for inflows to the Cordeaux Dam and the Nepean Dam for the winter season. The gray dots in both plots represent the historical inflows, and the contours in the left plot represent the kernel density estimate based on this historical data. The contours in the right plot represent the kernel density estimate of the forecast results, which were generated as described above. The results show a close correspondence in the joint density estimates between the historical data and the forecasts, even though all regression modeling was performed on univariate time series. These results were replicated in numerous other bivariate plots (not shown) and reflect the conclusions of the more rigorous analysis of joint density by Westra et al. (2007, 2008) in which the joint density estimates of the ICA-based approach were compared using the mean integrated squared error (MISE) criteria and found to outperform both PCA and canonical correlation analysis (CCA) in capturing spatial dependence, due to the emphasis on higher-order dependencies not accounted for in covariance- or correlation-based statistics.

Note that for the bivariate plot in Fig. 6 we did not transform the logged data back to the original space. Owing to the nature of the simplistic bootstrapping procedure used, it is possible that there will be a small minority of estimated flows that are negative and therefore physically implausible. This is inevitable for datasets that are highly skewed with generally low flows interspersed with several very high flow events. During implementation of such models, the treatment of negative flows is left to the discretion of the user and may involve setting negative flows to zero or, alternatively, discarding negative flows and drawing new values randomly from the joint density estimates until all the results are positive.

5. Conclusions

The purpose of the research presented in this paper is to provide an approach for identifying likely climatic

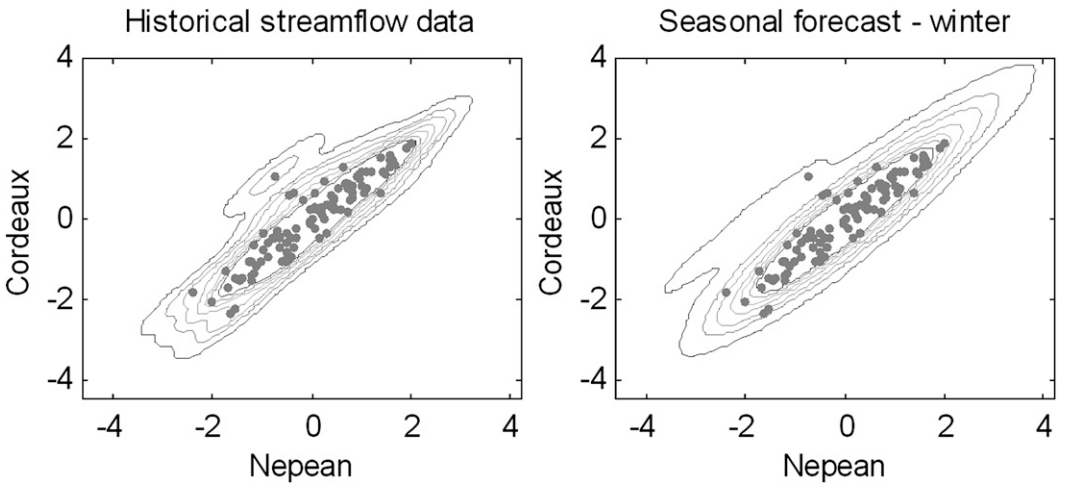

FIG. 6. Spatial dependence between historical streamflow and the multivariate streamflow forecasts using the method described in algorithm 1 at the Cordeaux Dam and the Nepean Dam. The gray dots represent historical winter streamflow, and the contours represent kernel density estimates based on the (left) historical data and (right) forecast data using the multivariate ICA model presented in section 4a.

impacts on reservoir inflows and then using this information to generate probabilistic multivariate seasonal forecasts for future reservoir inflows.

A review of the literature on sources of rainfall and streamflow variability in the Sydney water supply system demonstrates that the ENSO phenomenon strongly influences rainfall in much of the eastern third of the Australian continent, although the same literature shows a somewhat lower (but still statistically significant) correlation on the eastern coastal fringe, east of the Great Dividing Range, which is the region where Sydney sources its water supply. The wavelets analysis used in Westra and Sharma (2006) showed that rainfall in this region varied with a period of 13 yr, and this result has been reproduced for the reservoir inflow data analyzed in the present paper. This period is longer than the period of 2–8 yr commonly associated with the ENSO phenomenon and shorter than the interdecadal variability associated with the Pacific decadal oscillation, which was shown to vary with a period of 24 yr.

Examining the correlation between the aggregate annual and seasonal reservoir inflows and the global SSTA dataset, statistically significant correlations were observed in both the equatorial Pacific and extratropical Pacific of opposing sign, with an upper bound of approximately ± 0.4 depending on the season. The spatial patterns were coherent across all seasons, although correlations were generally lower for spring. The spatial patterns contained similarities to the spatial patterns associated with ENSO except that, whereas ENSO is dominant in the central and eastern equatorial Pacific, the correlations with aggregate reservoir inflows focus more on the central and western equatorial Pacific as well as the extratropical Pacific. As such, variability in the Sydney reservoir inflow region shares some characteristics with both ENSO and the PDO, with this conclusion being confirmed in the low but statistically significant correlation coefficients between aggregate inflows and indices of both these phenomena.

For this analysis it is evident that, to develop seasonal statistical forecasts for the Sydney reservoir inflows, it is necessary to use a method that captures as much of the global SSTA variability as possible rather than restricting the predictors to indices of climate. Furthermore, such a method also needs to capture the spatial dependence between reservoir inflows at each of the 11 data collection sites presented in Table 1. As such, we extended the statistical methods developed in Westra et al. (2008, 2009), which use ICA to transform the multivariate data into a set of independent components, such that it becomes possible to use univariate statistical methods (linear regression in this case) to develop forecasts based on persistence and exogenous climate predictors. The forecast performance was assessed to be similar to the linear regression-based methods that did not include a spatial component discussed in section 3. When examining spatial dependence, however, it was observed that the ICA-based method provides an excellent representation of spatial dependence from one location to the next and for this reason is a significant improvement to other methods that are currently available that either do not take spatial dependence into account or require the full model to be developed in the multivariate setting rather than breaking the model

down into univariate (independent) components. As such, applications of this method are well suited to any probabilistic reservoir study in which interactions between different reservoirs are important.

A logical extension of the present work is to further develop the linear-regression-based models of Eqs. (4) and (6) for incorporating persistence and ''climate state,'' respectively, to a nonlinear or nonparametric setting (e.g., Barros and Bowden 2008). Since these models need only be developed in one dimension, such extensions are relatively straight forward, although care needs to be taken that such models are appropriately validated using some form of cross-validation. This work will be future research.

Acknowledgments. Funding for this research came from the Australian Research Council and the Sydney Catchment Authority. Their support for this work is gratefully acknowledged. We also wish to thank the two anonymous reviewers for providing valuable feedback.

REFERENCES

Ashok, K., Z. Guan, and T. Yamagata, 2003: Influence of the Indian Ocean Dipole on the Australian winter rainfall. *Geophys. Res. Lett.,* **30,** 1821, doi:10.1029/2003GL017926.

Barros, A. P., and G. J. Bowden, 2008: Toward long-lead operational forecasts of drought: An experimental study in the Murray-Darling River Basin. *J. Hydrol.,* **357,** 349–367.

Cai, W., P. H. Whetton, and A. B. Pittock, 2001: Fluctuations of the relationship between ENSO and northeast Australian rainfall. *Climate Dyn.,* **17,** 421–432.

Chiew, F. H. S., and T. A. McMahon, 2002: Global ENSO-streamflow teleconnection, streamflow forecasting and interannual variability. *Hydrol. Sci. J.,* **47,** 505–522.

——, T. C. Piechota, J. A. Dracup, and T. A. McMahon, 1998: El Niño/Southern Oscillation and Australian rainfall, streamflow and drought: Links and potential for forecasting. *J. Hydrol.,* **204,** 138–149.

——, S. L. Zhou, and T. A. McMahon, 2003: Use of seasonal streamflow forecasts in water resources management. *J. Hydrol.,* **270,** 135–144.

Colman, A. W., and M. K. Davey, 2003: Statistical prediction of global sea-surface temperature anomalies. *Int. J. Climatol.,* **23,** 1677–1697.

Cordery, I., and Y. Opoku-Ankomah, 1994: Temporal variation of relations between tropical sea-surface temperatures and New South Wales rainfall. *Aust. Meteor. Mag.,* **43,** 73–80.

——, and M. McCall, 2000: A model for forecasting drought from teleconnections. *Water Resour. Res.,* **36,** 763–768.

Drosdowsky, W., 1993: Potential predictability of winter rainfall over southern and eastern Australia using Indian Ocean sea-surface temperature anomalies. *Aust. Meteor. Mag.,* **42,** 1–6.

——, 2002: SST phases and Australian rainfall. *Aust. Meteor. Mag.,* **51,** 1–12.

——, and L. E. Chambers, 2001: Near-global sea surface temperature anomalies as predictors of Australian seasonal rainfall. *J. Climate,* **14,** 1677–1687.

Dutta, S. C., J. W. Ritchie, D. M. Freebairn, and G. Y. Abawi, 2006: Rainfall and streamflow response to El Niño Southern Oscillation: A case study in a semiarid catchment, Australia. *Hydrol. Sci. J.,* **51,** 1006–1020.

Filho, F. A. S., and U. Lall, 2003: Seasonal to interannual ensemble streamflow forecasts for Ceara, Brazil: Applications of a multivariate, semiparametric algorithm. *Water Resour. Res.,* **39,** 1307, doi:10.1029/2002WR001373.

Goddard, L., S. J. Mason, S. E. Zebiak, C. F. Ropelewski, R. Basher, and M. A. Cane, 2001: Current approaches to seasonal-to-interannual climate predictions. *Int. J. Climatol.,* **21,** 1111–1152.

Hamlet, A. F., D. Huppert, and D. P. Lettenmaier, 2002: Economic value of long-lead streamflow forecasts for Columbia River hydropower. *J. Water Resour. Plann. Manage.,* **128,** 91–101.

Hyvärinen, A., J. Karhunen, and E. Oja, 2001: *Independent Component Analysis*. John Wiley, 481 pp.

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

Jevrejeva, S., J. C. Moore, and A. Grinsted, 2003: Influence of the Arctic Oscillation and El Niño-Southern Oscillation (ENSO) on ice conditions in the Baltic Sea: The wavelet approach. *J. Geophys. Res.,* **108,** 4677, doi:10.1029/2003JD003417.

——, ——, and ——, 2004: Oceanic and atmospheric transport of multiyear El Niño–Southern Oscillation (ENSO) signatures to the polar regions. *Geophys. Res. Lett.,* **31,** L24210, doi:10.1029/2004GL020871.

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** (C9), 18 567–18 589.

Kiem, A. S., S. W. Franks, and G. Kuczera, 2003: Multi-decadal variability of flood risk. *Geophys. Res. Lett.,* **30,** 1035, doi:10.1029/2002GL015992.

Lettenmaier, D. P., and E. F. Wood, 1993: Hydrologic forecasting. *Handbook of Hydrology,* D. R. Maidment, Ed., McGraw-Hill, 26.1–26.30.

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

——, ——, Y. Zhang, J. M. Wallace, and R. C. Francis, 1997: A Pacific interdecadal climate oscillation with impacts on salmon production. *Bull. Amer. Meteor. Soc.,* **78,** 1069–1079.

Nicholls, N., 1985: Towards the prediction of major Australian droughts. *Aust. Meteor. Mag.,* **33,** 161–166.

——, 1989: Sea surface temperatures and Australian winter rainfall. *J. Climate,* **2,** 965–973.

——, B. Lavery, C. Frederiksen, W. Drosdowsky, and S. Torok, 1996: Recent apparent changes in relationships between the El Niño-Southern Oscillation and Australian rainfall and temperature. *Geophys. Res. Lett.,* **23,** 3357–3360.

Piechota, T. C., F. H. S. Chiew, J. A. Dracup, and T. A. McMahon, 1998: Seasonal streamflow forecasting in eastern Australia and the El Niño–Southern Oscillation. *Water Resour. Res.,* **34,** 3035–3044.

——, ——, ——, and ——, 2001: Development of exceedance probability streamflow forecast. *J. Hydrol. Eng.,* **6,** 20–28.

Pittock, A. B., 1975: Climatic change and the patterns of variation in Australian rainfall. *Search,* **6,** 498–504.

Power, S., F. Tseitkin, S. Torok, B. Lavery, R. Dahni, and B. McAvaney, 1998: Australian temperature, Australian rainfall and the Southern Oscillation, 1910-1992: Coherent variability and recent change. *Aust. Meteor. Mag.,* **47,** 85–101.

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

——, F. Tseitkin, V. Mehta, S. Torok, and B. Lavery, 1999b: Decadal climate variability in Australia during the twentieth century. *Int. J. Climatol.,* **19,** 169–184.

Ropelewski, C. F., and M. S. Halpert, 1987: Global and regional scale precipitation patterns associated with the El Niño/Southern Oscillation. *Mon. Wea. Rev.,* **115,** 1606–1626.

Saji, N. H., and T. Yamagata, 2003: Possible impacts of Indian Ocean Dipole mode events on global climate. *Climate Res.,* **25,** 151–169.

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

Sharma, A., 2000a: Seasonal to interannual rainfall probabilistic forecasts for improved water supply management: Part 1—A strategy for system predictor identification. *J. Hydrol.,* **239,** 232–239.

——, 2000b: Seasonal to interannual rainfall probabilistic forecasts for improved water supply management: Part 3—A nonparametric probabilistic forecast model. *J. Hydrol.,* **239,** 249–258.

——, U. Lall, and D. G. Tarboton, 1998: Kernel bandwidth selection for a first order nonparametric streamflow simulation model. *Stochastic Hydrol. Hydraul.,* **12,** 33–52.

——, K. C. Luk, I. Cordery, and U. Lall, 2000: Seasonal to interannual rainfall probabilistic forecasts for improved water supply management: Part 2 — Predictor identification of quarterly rainfall using ocean-atmosphere information. *J. Hydrol.,* **239,** 240–248.

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.

Van den Dool, H. M., J. Huang, and Y. Fan, 2003: Performance and analysis of the constructed analogue method applied to U.S. soil moisture over 1981–2001. *J. Geophys. Res.,* **108,** 8617, doi:10.1029/2002JD003114.

Verdon, D. C., and S. W. Franks, 2005: Indian Ocean sea surface temperature variability and winter rainfall: Eastern Australia. *Water Resour. Res.,* **41,** W09413, doi:10.1029/2004WR003845.

——, A. M. Wyatt, A. S. Kiem, and S. W. Franks, 2004: Multidecadal variability of rainfall and streamflow: Eastern Australia. *Water Resour. Res.,* **40,** W10201, doi:10.1029/2004WR003234.

Wang, B., and Y. Wang, 1996: Temporal structure of the Southern Oscillation as revealed by waveform and wavelet analysis. *J. Climate,* **9,** 1586–1598.

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.

——, C. Brown, U. Lall, and A. Sharma, 2007: Modeling multivariable hydrological series: Principal component analysis or independent component analysis? *Water Resour. Res.,* **43,** W06429, doi:10.1029/2006WR005617.

——, A. Sharma, C. Brown, and U. Lall, 2008: Multivariate streamflow forecasting using independent component analysis. *Water Resour. Res.,* **44,** W02437, doi:10.1029/2007WR006104.

——, C. Brown, U. Lall, I. Koch, and A. Sharma, 2009: Interpreting variability in global SST data using independent component analysis and principal component analysis. *Int. J. Climatol.,* doi:10.1002/joc.1888, in press.

White, W. B., A. Gershunov, J. L. Annis, G. McKeon, and J. Syktus, 2004: Forecasting Australian drought using Southern Hemisphere modes of sea-surface temperature variability. *Int. J. Climatol.,* **24,** 1911–1927.

Yao, H., and A. Georgakakos, 2001: Assessment of Folsom Lake response to historical and potential future climate scenarios: 2. Reservoir management. *J. Hydrol.,* **249,** 176–196.

Zhang, Y., J. M. Wallace, and D. S. Battisti, 1997: ENSO-like interdecadal variability: 1900–93. *J. Climate,* **10,** 1004–1020.