INTERNATIONAL JOURNAL OF CLIMATOLOGY
Int. J. Climatol. **27**: 1389–1402 (2007)
Published online 22 January 2007 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/joc.1474

Seasonal sea-level forecasts by canonical correlation analysis – an operational scheme for the U.S.-affiliated Pacific Islands

MD. Rashed Chowdhury,[a,*] Pao-Shin Chu,[b] Thomas Schroeder[c] and Nicole Colasacco[d]

[a] *Pacific ENSO Applications Center, University of Hawaii at Manoa, Honolulu, Hawaii*
[b] *Department of Meteorology, School of Ocean and Earth Science and Technology (SOEST), University of Hawaii at Manoa, Honolulu, Hawaii*
[c] *Department of Meteorology, School of Ocean and Earth Science and Technology (SOEST), Joint Institute for Marine and Atmospheric Research (JIMAR), University of Hawaii at Manoa, Honolulu, Hawaii*
[d] *Pacific ENSO Applications Center, University of Hawaii at Manoa, Honolulu, Hawaii*

Abstract:

The objective of this study is to develop an operational canonical correlation analysis (CCA) statistical model for sea-level forecasts in the U. S.-affiliated Pacific Islands (USAPI) with lead times of several months or longer. The El Niño-Southern Oscillation (ENSO) climate cycle and the sea-surface temperatures (SSTs) in the tropical Pacific Ocean are taken as the primary factors in modulating sea-level variability on the seasonal time scales.

Observations revealed that the sea-level variations in the USAPI are sensitive to ENSO cycle with low sea level during El Niño and high sea level during La Niña events. The correlation between the sea-level variability and the fluctuations of tropical Pacific SSTs has been found to be strong. The cross-validated results indicated that the SST-based CCA model is potentially useful in predicting seasonal sea-level variations in the USAPI. For all target seasons at 1- and 2-season lead times, the average correlation skill has been found to be 0.50 or better. Based on this operational CCA model, the real-time forecasts for seasonal sea-level variations (i.e. deviations with respect to climatology) are published at the official web site of Pacific ENSO Applications Center (PEAC) (http://lumahai.soest.hawaii.edu/Enso/peu/update.html) for planning and decision options regarding hazard management in the USAPI.

KEY WORDS sea level; El Niño-southern oscillation (ENSO); sea-surface temperature (SST); canonical correlation analysis (CCA); and US – affiliated Pacific Islands (USAPI)

Received 21 September 2005; Revised 1 November 2006; Accepted 6 November 2006

INTRODUCTION

This study is aimed at providing a basis for the development of an outlook for seasonal sea-level variability in the U.S-affiliated Pacific Islands (USAPI). At present, the USAPI consisting of Guam, Republic of Palau, Commonwealth of the Northern Mariana Islands, Federated States of Micronesia, Marshall Islands, and American Samoa are provided with an El Niño–Southern Oscillation (ENSO)-based seasonal rainfall prediction. This information is significant to hazard preparedness actions in these islands. For example, the advance information on ENSO during the year of 1997–1998 helped the USAPI islanders to prepare, collectively and individually, a real-time hazard mitigation response.

In addition to seasonal rainfall products, there is a strong demand for forecasts of sea-level variations with a lead time of a season or longer for the USAPI. This demand was raised by the representatives of each of these islands in the last Pacific ENSO Applications Center (PEAC) review meeting (http://research.eastwestcenter.org/climate/PEAC/) held in Honolulu, Hawaii (USA), June 1–3, 2004. Their motivation stems from the necessity to plan for extreme events and to facilitate basin-wide planning. Such forecasts would be expected to have far-reaching economic ramifications. Therefore, our primary intentions are to identify the nature and strength of possible teleconnections between variations in sea level and the ENSO events, and to develop a statistical scheme that can capture the natural variability of seasonal sea level, quantifying the skill at lead times of one to three seasons.

ENSO has a significant impact on the climate variability in the Pacific Islands. Bjerknes' (1966, 1969) pioneering studies indicated that tropical climate was strongly influenced by ENSO episodes. Subsequent empirical studies (e.g. Ropelewski and Halpert, 1987; Chu, 1995; Chu and Chen, 2005) supported the results of Bjerknes. Lau's (1985) global climate model experiments indicate

* Correspondence to: MD. Rashed Chowdhury, Research Scientist, Pacific ENSO Applications Center, University of Hawaii at Manoa, 2525 Correa Road, HIG 350, Honolulu, Hawaii 96822.
E-mail: rashed@hawaii.edu

that much of the atmospheric response to ENSO is associated with the changes in sea surface temperatures (SSTs) in the Pacific. Pacific SSTs can thus be used to forecast regional climate fluctuations, especially in the tropical Pacific (Barnston and He, 1996; Yu *et al.*, 1997). The state of the ENSO not only directly affects the climate in the tropical Pacific, but also affects the climate over many large regions of the world far removed from the Pacific through a chain of teleconnections that take only a few weeks to occur once an El Niño or La Niña has established itself (Tribbia, 1991).

While the monsoon subsystems in the western Pacific are connected to variations in SSTs in the Pacific, research specifically addressing the ENSO climate cycle and sea-level variation in the USAPI is rather limited. Xue *et al.* (2000) provided some initial findings on ENSO prediction and its impact on sea-level variability, concluding that the sea level contains the most essential information for ENSO because it represents the filtered response of the ocean to noisy wind forcing (also see Xue and Leetmaa, 2000). Our study, however, examines the robustness of the ENSO and sea-level relationship by analyzing the composites of seasonal variations and by correlating the SST time series at each geographical grid-point with sea levels at various gauges. Also, different from Xue and Leetmaa's Markov model for sea-level forecasts, we employed CCA to develop an operational forecasting scheme for sea-level variability.

U.S.-AFFILIATED PACIFIC ISLANDS – ENVIRONMENTAL SETTINGS

The U.S-affiliated Pacific Islands include Territory of Guam, Republic of Palau, Commonwealth of the Northern Mariana Islands, Republic of the Marshall Islands, Federated States of Micronesia, and American Samoa (Figure 1). A brief summary of environmental settings of these Islands is as follows:

Guam is the largest Micronesian island, with a land area of 212 square miles. The Republic of Palau (RPalau) is the westernmost jurisdiction in Micronesia, less than 500 miles from the Philippines. The Commonwealth of the Northern Mariana Islands (CNMI) forms a chain of 17 volcanic islands, with a land area of 181 square miles. The Republic of the Marshall Islands (Marshalls) consists of two chains of 29 coral atolls and five low islands stretching several hundred miles from north to south with a total land area of 70 square miles. The Federated States of Micronesia (FSM) consist of four states: Chuuk, Kosrae, Yap, and Pohnpei. American Samoa (ASamoa), a group of islands in the mid-South Pacific, has a land area of 76 square miles.

DATA, BASIC INDICES, AND METHOD

Sea-level data

The University of Hawaii Sea Level Center (UHSLC) provides three online databases: research quality data, fast delivery data, and map data. The quality assessment is mostly based on the residuals (observed data minus predicted tides) of the hourly data. This assessment also applies to the daily and monthly data since they were derived from the quality-controlled hourly data. In order to maintain the required quality of data, the UHSLC emphasizes three main aspects: 1) the linking of the data to a reference level (tidal datum), 2) the inspection of the timing quality, and 3) the replacement of short gaps and spikes. Daily values are obtained by using a two-step filtering operation. First, the dominant diurnal and semidiurnal tidal components are removed from the quality-controlled hourly values. Secondly, a 119-point convolution filter (Bloomfield, 1976) centered on noon is applied to remove the remaining high-frequency energy and to prevent aliasing when the data are computed to daily values. The 95, 50, and 5% amplitude points

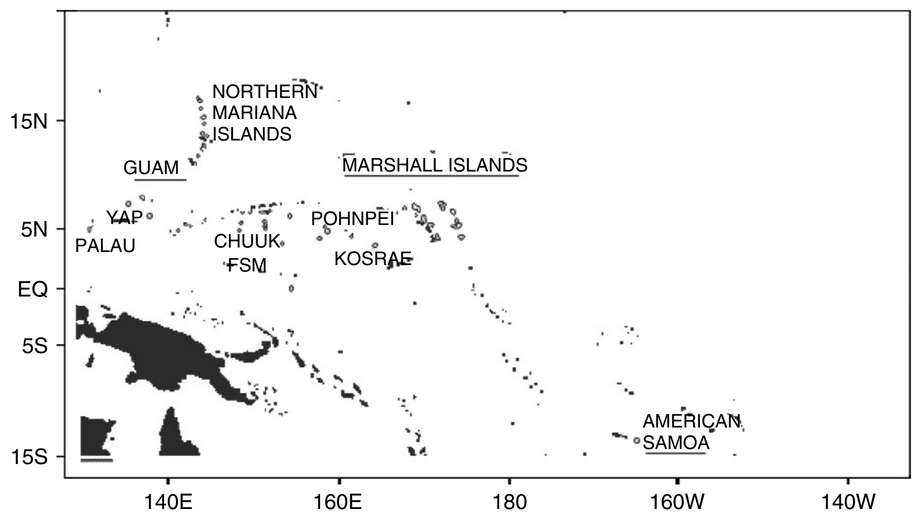

Figure 1. Geographical locations of the U.S-affiliated Pacific Islands (Note that tide gauge stations *Marianas* from Guam, *Kwajalein* from Marshalls, and *Pago_Pago* from American Samoa (underlined) are taken for comprehensive composite analyses). See Table I for geographical details (latitude, longitude) of these tide gauge stations.

DOI: 10.1002/joc

are 124.0, 60.2, and 40.2 h, respectively. The Nyquist frequency of the daily data is at a period of 48 h, which has a response of about 5% amplitude; thus, aliasing is minimal. The primary tidal periods have a response of less than 0.1% amplitude.

The technical aspects of quality-control procedures have been well documented in Kilonsky and Caldwell (1991) and Caldwell and Kilonsky (1992). This research quality data of the UHSLC is the largest global collection of quality-controlled sea-level data. More comprehensive data-related discussions are also available at http://ilikai.soest.hawaii.edu/UHSLC/jaslpr2/slman2.html (also see references therein). We simply downloaded the research quality monthly sea-level data for some of the specific tide gauge stations from the UHSLC web site (available at http://ilikai.soest.hawaii.edu/uhslc/rqds.html). The monthly sea-level values available in the UHSLC are calculated from the daily data with a simple average. Geographical details (latitude, longitude) and length of data records of the tide gauge stations are listed in Table I. Tide gauge stations located at Guam, Marshalls, and American Samoa have the longest data records and, in order to maintain consistency, detailed analyses are conducted for all the above three stations using data records from 1950 to 2003. For the other three stations with shorter records in Table I, some analyses are also performed.

While most statistical techniques require that data are stationary, most climate data exhibit cyclostationary features due to the seasonal nature of climate. In order to remove nonstationarity from the data, the periodic mean function of sea-level variation, which is estimated using data from 1975 through 1995, has been subtracted. Although the annual maximum sea level shows an increasing trend, the increase being concentrated in the latter half of 20th century (see *Pacific ENSO Update* 12 : 2, 2006), the annual variability of mean did not show any significant trend.

For a number of reasons, missing values existed in some of the years in the UHSLC time series. For example, the sea-level data of Guam in 1998 were missing because the tide gauge was severely damaged by tornado. In this case, missing values were replaced with the values of the best 'nearest neighboring stations'. In this case, the missing values of Guam were replaced by the values obtained by fitting a linear regression ($r^2 = 0.56$, *significant at 0.01 level*) equation with the observed values of Saipan, after observing that the two time series are highly correlated to each other and also statistically significant. Previous studies by Chowdhury *et al.* (2006) have already confirmed that, with the exception of American Samoa, the monthly sea-level fluctuations of the USAPIs are significantly correlated with each other's. This finding also justifies our initial choice to utilize two of the north Pacific Islands – one from the west (Guam) and the other from the east (Marshalls) – and one of the South Pacific Islands (American Samoa) for detailed analyses. Data records for these three gauges are longer and their locations are also geographically representative. It is also worth noting here that currently the CCA analyses for real-time forecasts are conducted from a shorter time series data (1975–2005). The shorter time series exhibited more data consistency for all the available gauges in the USAPI and provided better management capability to cyclostationary climate data.

Observed climate fields of SST and atmospheric circulation data

The National Center for Environmental Prediction (NCEP) historical monthly fields of the global SST have been used in this study. In this NCEP data set, the monthly SST data for the period of 1982–1994 have been established by blending real-time *in situ* (ship and buoy) and satellite measurements (Reynolds, 1988; Reynolds and Marsico, 1993). For this study, the Extended Reconstructed SST (ERSST) (ERSST: version 2) data (Smith and Reynolds, 2002) from the National Oceanic and Atmospheric Administration (NOAA) – National Climate Data Center (NCDC) have been used. The mean has been subtracted from the original data. The ERSST data are based on the Comprehensive Ocean-Atmosphere Data (COAD) data.

The SST data for the tropical region covered by latitudes (30S $\sim$ 30N) and longitudes (100°E $\sim$ 60°W) have been downloaded from the web link of the International Research Institute for Climate Prediction

Table I. Geographical details (latitude, longitude) and length of data records of each tide gauge station.

Islands	Tide gauge stations (#)	Latitude	Longitude	Years of data records
Guam	Marianas (# 053)	13.44°N	144.65°E	1948–2003
Rpalau	Malakai – B (# 007)	7.33°N	134.47°E	1969–2003
CNMI	Saipan (# 028)	15.23°N	145.75°E	1978–2003
Marshalls	Kwajalein (# 055)	8.73°N	167.73°E	1946–2003
FSM	Chuk (# 054)	7.45°N	151.85°E	1963–1991
ASamoa	Pago Pago (# 056)	14.29°S	170.69°W	1948–2004

Note: RPalau stands for Republic of Palau, CNMI for Commonwealth of the Northern Mariana Islands, FSM for Federated States of Micronesia, and ASamoa for American Samoa (*underlined stations are taken for comprehensive analyses*).

(http://iridl.ldeo.columbia.edu/expert/SOURCES/.NOAA/.NCDC/.ERSST/.SST/). The time record is from January 1950 to December 2003. For atmospheric circulation data (zonal wind at 850-hPa), the NCEP/National Center for Atmospheric Research (NCAR) reanalysis has been used (Kalnay *et al.*, 1996). This 850-hPa level is known to be a good indicator of broad-scale features of low-level winds.

Harmonic analysis

Harmonic analysis has commonly been used to determine the annual fluctuations of geophysical time series (see Chowdhury *et al.*, 2006). In this study, the amplitude, phase, and percentage explained by the annual and semi-annual variations at these long-term sea-level stations are analyzed.

Linear correlations and scatterplots

The standard approach to study the relationship among the climate fields (such as SST) by correlating time series at each geographical grid-point with the variable of interest (such as sea level of each year) has been conducted. The Pearson correlation was adopted and the resulting correlations of seasonally averaged variables have been plotted. Scatter plots were drawn to examine whether the relationship between SST anomalies and sea-level variations are linear.

Canonical correlation analysis (CCA) and cross validation

The CCA technique was introduced by Hotelling (1935), and since its introduction, prediction of climate variations using the method has received wide attention for its many success stories. CCA is a procedure for assessing the relationship between two fields of variables. Specifically, this analysis allows us to investigate the relationship between *two sets* of basis vectors, one for **x** and the other for **y**, such that the correlations between the *projections* of the variables onto these basis vectors are mutually maximized. Spatial loading patterns of empirical orthogonal function mode 1 (EOF) (henceforth, EOF1) and empirical orthogonal function mode 2 (henceforth, EOF2) for SST were plotted. Finally, the CCA cross-validation skills results are provided. This popular method has been described extensively in the literature (see Barnston, 1994; Chu and He, 1994; Barnston and He, 1996; Barnston and Smith, 1996; Yu *et al.*, 1997), and therefore not discussed here further. As in many cases the operational CCA forecasts are usually derived by applying linear statistics to a nonlinear system, therefore several authors have pointed out that caution needs to be applied when interpreting CCA (Cherry, 1996; Cherry, 1997; Newman and Sardeshmukh, 1995).

In CCA, it is implicitly assumed that the first three modes of the predictand field are the ones that are most highly correlated with the first eight modes of the predictor fields. However, there is no guarantee that one of the higher modes of variations in the predictor set (e.g. higher than the eighth mode) will not be strongly associated with the predictand set. To overcome this potential problem, an independent approach – Principal Component Regression (PCR) (Draper and Smith, 1981) – has been used. The cross-validation scheme is also employed to assess the *PCR* model validity and adequacy.

The Climate Predictability Tool (CPT) software (available at http://iri.columbia.edu/outreach/software/) has been used to generate the CCA and PCR analyses. The CPT provides a Windows package for constructing a seasonal climate forecast model, performing model validation, and producing forecasts, given updated data. Although the software is specifically tailored for these applications, it can be used in more general settings to perform CCA or PCR on any data, and for any application. The help pages provide guidance for using the software only in the applications for which it was specifically designed, but should be sufficiently general to provide guidance for other uses.

HARMONIC, CORRELATION, AND EMPIRICAL ORTHOGONAL FUNCTION (EOF) ANALYSIS

Annual cycle of sea level

The climatology of the annual cycle of sea level has been investigated, details of which are reported elsewhere (*Pacific ENSO Update* 11 : 1, 2005; also see Chowdhury *et al.*, 2006). In brief, the sea-level variability for most of the northern Pacific Islands, by and large, displayed a strong annual cycle with a gradual increase of sea level from January to July and recession from July to December. On the other hand, the lone South Pacific island (American Samoa), tended to show several peaks in the annual cycle. Results of harmonic analyses showed that, except for the Marshalls, the annual cycle (i.e. first harmonic) explained a considerable percentage (64–88%) of variances of the sea-level variability for the other North Pacific Islands (e.g. Guam, RPalau, CNMI, and FSM). The variance explained for Marshalls is 44% and for American Samoa it is only 20%. However, the semiannual cycle (i.e. second harmonic) adds considerably to the variances in these two cases −17% for Marshalls and 11% for American Samoa respectively.

Linear correlation of seasonal averages of SSTs and sea level

The 3-month average sea level for the target season in July–August–September (JAS) was correlated with the 3-month moving average of SSTs starting from the preceding season. In Guam, the sea level in JAS displays strong and positive correlations with the SSTs of the preceding April–May–June (AMJ) in the tropical western Pacific and negative correlation with SSTs in the Niño 3.4 region (Figure 2(a)). The positive correlation implies that warmer sea waters or more heat content with a deeper thermocline in the tropical western North Pacific correspond to higher sea levels in Guam. Conversely, negative sea-level anomalies are associated with cooler

DOI: 10.1002/joc

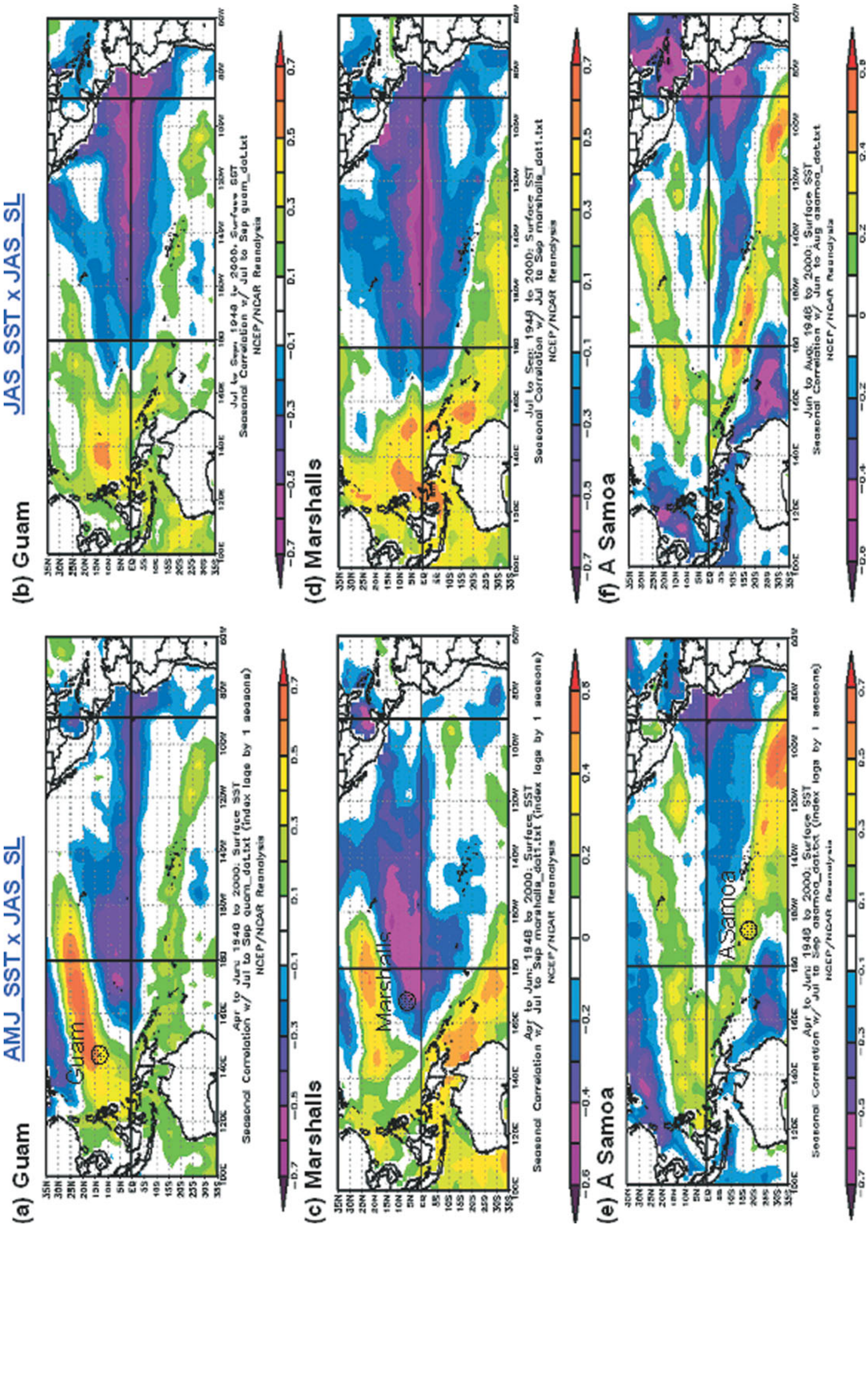

Figure 2. Linear correlation between seasonal SSTs and sea levels across the tropical Pacific. This figure is available in colour online at www.interscience.wiley.com/ijoc

sea waters or less heat content in the western North Pacific. By JAS, the area of positive correlation expands equatorward and the negative correlation in the equatorial eastern Pacific strengthens (Figure 2(b)). Similar patterns were also observed in the case of the Marshall Islands (Figure 2(c),(d)). This correlation result between the sea level for the target season JAS and the SSTs in the tropical Pacific indicates a strong link with ENSO. However, differing from Guam, local SSTs do not seem to be in phase with sea-level variations at the Marshall Islands. American Samoa showed a correlation map in AMJ that is similar to that for Guam and the Marshalls (Figure 2(e)). However, in the following JAS, the positive correlation in the southwestern Pacific strengthened, but the negative correlation in the Niño 3.4 region weakened and, as a result, the ENSO signal in the central Pacific eroded (Figure 2(f)).

Scatterplots for SST anomalies and seasonal sea level variations are shown in Figure 3. These plots are prepared on the basis of the anomalous SSTs of the region that displayed highest correlation with the sea-level variations (Figure 2). For example, the sea-level variations of Guam displayed highest correlation (+0.6–0.7) with the anomalous SSTs in the region enveloped by longitude 140E–180 and latitude 10–20N (Figure 2(a)). Similarly, the highest correlated (−0.5–0.6) region for Marshalls lies between longitude 170E–140W and latitude is 0–20N (Figure 2(c)). For American Samoa, the correlation is +0.5–0.7 and the region lies between longitude 170–100 °W and latitude 15–30 °S (Figure 2(e)).

The scatterplots tended to show that a moderate-to-strong linear relationship exists between the anomalous SSTs (in January-February-March (JFM), AMJ, and JAS) and sea-level variations (in AMJ, JAS, and October-November-December (OND)) (see Figure 3). It may be mentioned here that, other than these seasons, a very weak linear relationship has been found to exist (not shown in Figure 3) between SSTs in OND and sea level in JFM. This relationship is very strong for ASamoa and relatively weak for Marshalls. Other than for the season OND in Guam (Figure 3 (iii)), this relationship is also strong. It is important to note here that some nonlinearity among these data were observed, which was somewhat expected. However, this did not cause any serious barriers for CCA forecasts, because there are examples that operational empirical forecasts are usually derived by applying linear statistics to a nonlinear system and assumes, rather than proves, causality (see Murphy *et al*., 2001).

EOF analysis of SST and sea-level records

The EOF analyses of the SST data were carried out to minimize problems of collinearity and to generate relatively independent, contiguous SST indices. Leading empirical orthogonal function mode 1 (EOFs) are selected as independent variables for the subsequent CCA model. This is a recommended way of handling a prediction problem when the candidate predictor is a spatially co-varying field (Jolliffe *et al*., 1986). The EOF spatial loading summarizes the variances of SST and sea level respectively. The leading EOFs of SST anomalies (henceforth, X-EOFs) have been calculated for each season. A total of 8 eigenmodes have been retained in SST analysis as they explained 73–76% of the total variance (Table II). The EOFs for sea levels (henceforth, Y-EOFs) provide 83–96% of variances with the first three modes retained (Table III).

The spatial structure X-EOF1 resembles those of the leading eigenmodes presented in past studies (for example, see Kawamura, 1994 and references therein). For X-EOF1, negative loadings exist over the tropical western Pacific extending to the subtropical latitudes, and large positive loadings exist over the central and eastern equatorial Pacific (Figure 4 *left panel*). In this EOF1, the large positive loadings over the central and eastern equatorial Pacific (Figure 4) and relatively weak loading in the tropical Indian Ocean (not shown in Figure 4) resemble slightly different features from other similar studies (e.g. Hsiung and Newell, 1983; Nitta and Yamada, 1989). Furthermore, it can also be seen in this EOF1 that negative loadings over the central North Pacific around 20°–30 °N are not very dominant. Thus, like Kawamura (1994), it is concluded that this mode accounts for the fundamental SST fluctuations over the equatorial Pacific and is not very strongly linked to those over the tropical Indian Ocean or the central North Pacific.

Table II. Percentage of variance explained by eigenvectors for the Pacific sea-surface temperatures (values in parentheses are cumulative variance by the ***k*** largest eigenvalues).

Ek/SST	JFM SSTs	AMJ SSTs	JAS SSTs	OND SSTs
E1	30.5 (30.5)	26.2 (26.2)	29.0 (29.0)	31.5 (31.5)
E2	15.5 (46.0)	17.1 (43.3)	17.5 (46.5)	17.5 (49.0)
E3	10.5 (56.5)	9.5 (52.8)	8.1 (54.6)	8.1 (57.1)
E4	6.2 (62.7)	6.5 (59.3)	7.0 (61.6)	5.2 (62.3)
E5	5.1 (67.8)	5.0 (64.3)	5.0 (66.6)	4.0 (66.3)
E6	4.5 (72.3)	4.5 (68.7)	4.0 (70.6)	3.8 (70.1)
E7	3.5 (75.8)	3.8 (72.5)	3.2 (73.8)	3.0 (73.1)
E8	–	3.0 (75.5)	2.8 (76.0)	

(Note: JFM: January–February–March, AMJ: April–May–June, JAS: July–August–September, and OND: October–November–December).

Table III. Percentage of variance explained by eigenvectors for sea levels (values in parentheses are cumulative variance by the ***k*** largest eigenvalues).

Ek/Sea-level	JFM Sea level	AMJ Sea level	JAS Sea level	OND Sea level
E1	66.0 (66.0)	58.0 (58.0)	52.0 (52.0)	54.0 (54.0)
E2	18.0 (84.0)	25.0 (83.0)	32.0 (84.0)	34.0 (88.0)
E3	7.0 (91.0)	–	–	8.0 (96.0)

(Note: See footnote in Table II).

DOI: 10.1002/joc

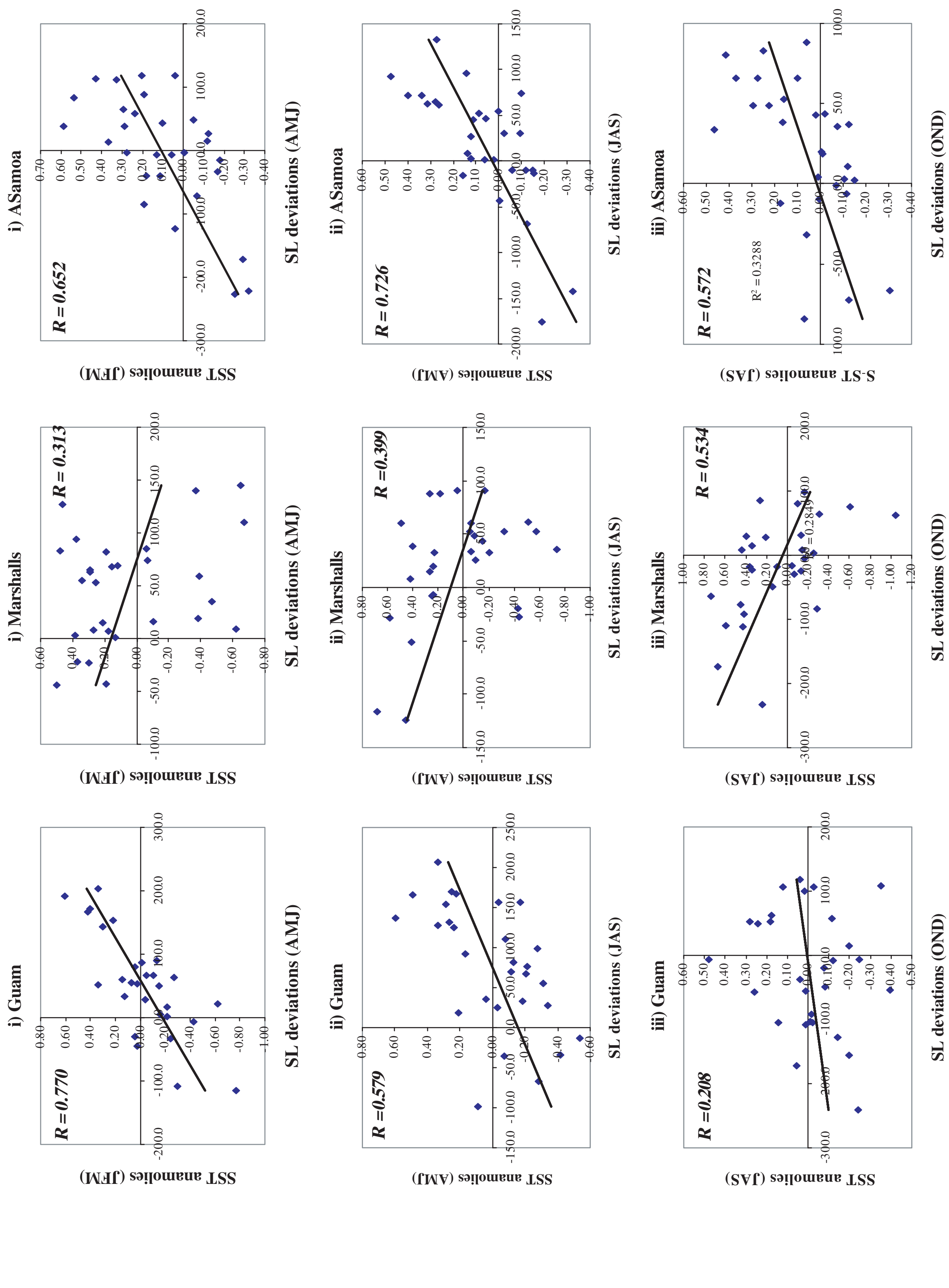

Figure 3. Scatterplot of sea-level and SST anomalies in Guam *(left)*, Marshalls *(middle)*, and American Samoa *(right)*. Note that x-axis is the sea-level deviations in mm and y-axis is the SST anomalies in °C. In parenthesis, names of the seasons are mentioned. This figure is available in colour online at www.interscience.wiley.com/ijoc

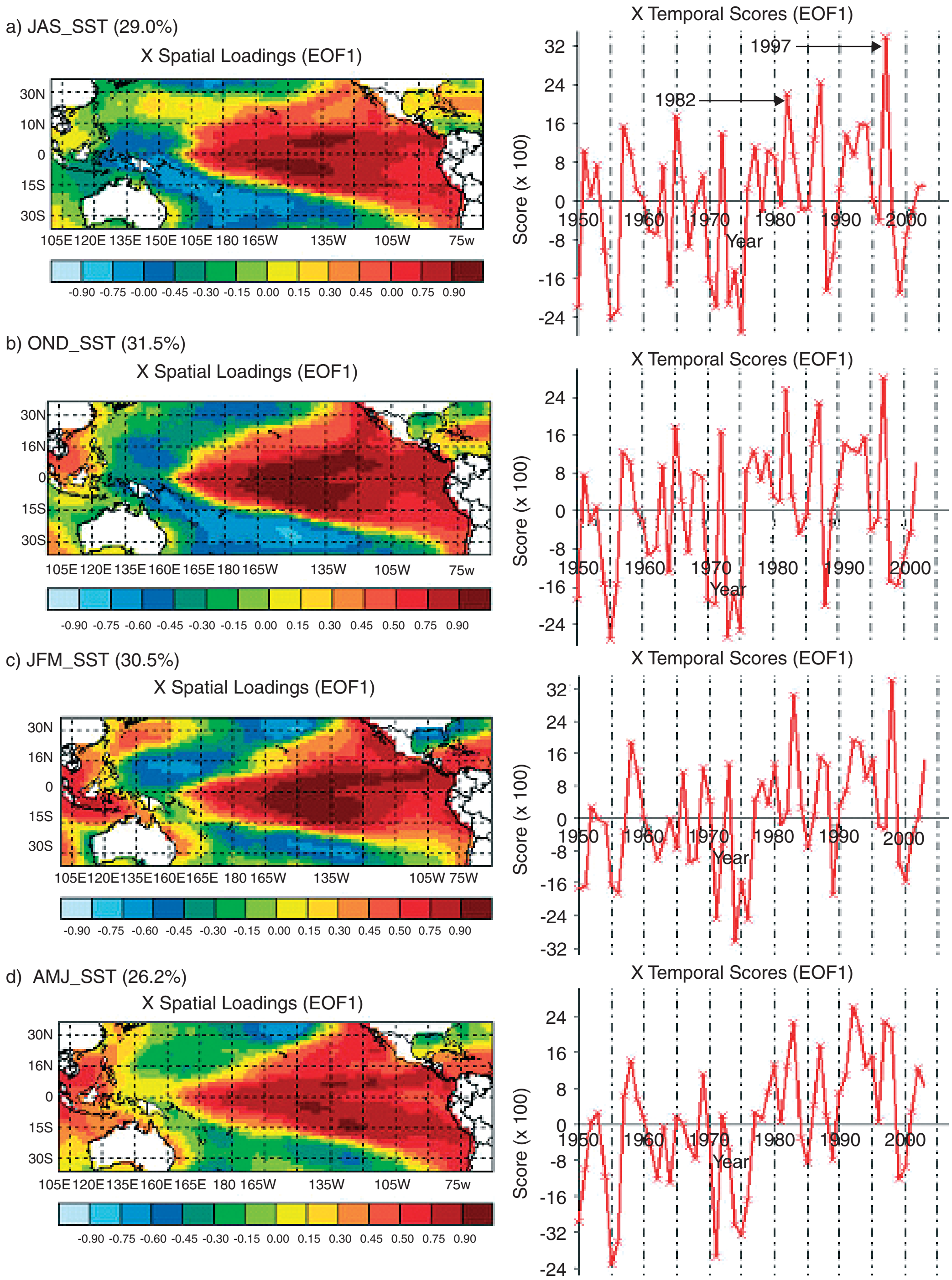

Figure 4. The principal loading patterns of SST anomalies for EOF1 (*left panel*). Temporal score time series data (*right panel*) (Note: Figures in parenthesis are percentage of variances; see notes in Table II for abbreviations of OND, JFM, AMJ, and JAS).

The temporal variability of EOF1 coincides quite well with the occurrence of ENSO events having a quasi periodicity of 2–5 years (Figure 4 *right panel*). There is also an indication of interdecadal variability in the EOF1 time series, particularly in JFM and AMJ SST modes (Figure 4(c) and (d)). From the time-dependent EOF series, major years of El Niño such as 1982–1983 and 1997–1998 stand out. However, an expanded El Niño signal (well off the tropics) is visible in EOF1 which could be a combination of El Niño and Pacific Decadal Oscillation (PDO). Therefore, although EOF1 has been identified with the ENSO mode, at this stage, it is difficult to conclude that no other eigenmodes include the ENSO signal.

The spatial structure of EOF2 shows some noteworthy differences to that of EOF1 (Figure 5). In EOF2, large positive loadings are located over the western Pacific and the South China Sea, whereas negative loadings exist in the mid-latitude of the north Pacific (Figure 5(a)). Also weak positive loadings exist in low-latitude regions of the

DOI: 10.1002/joc

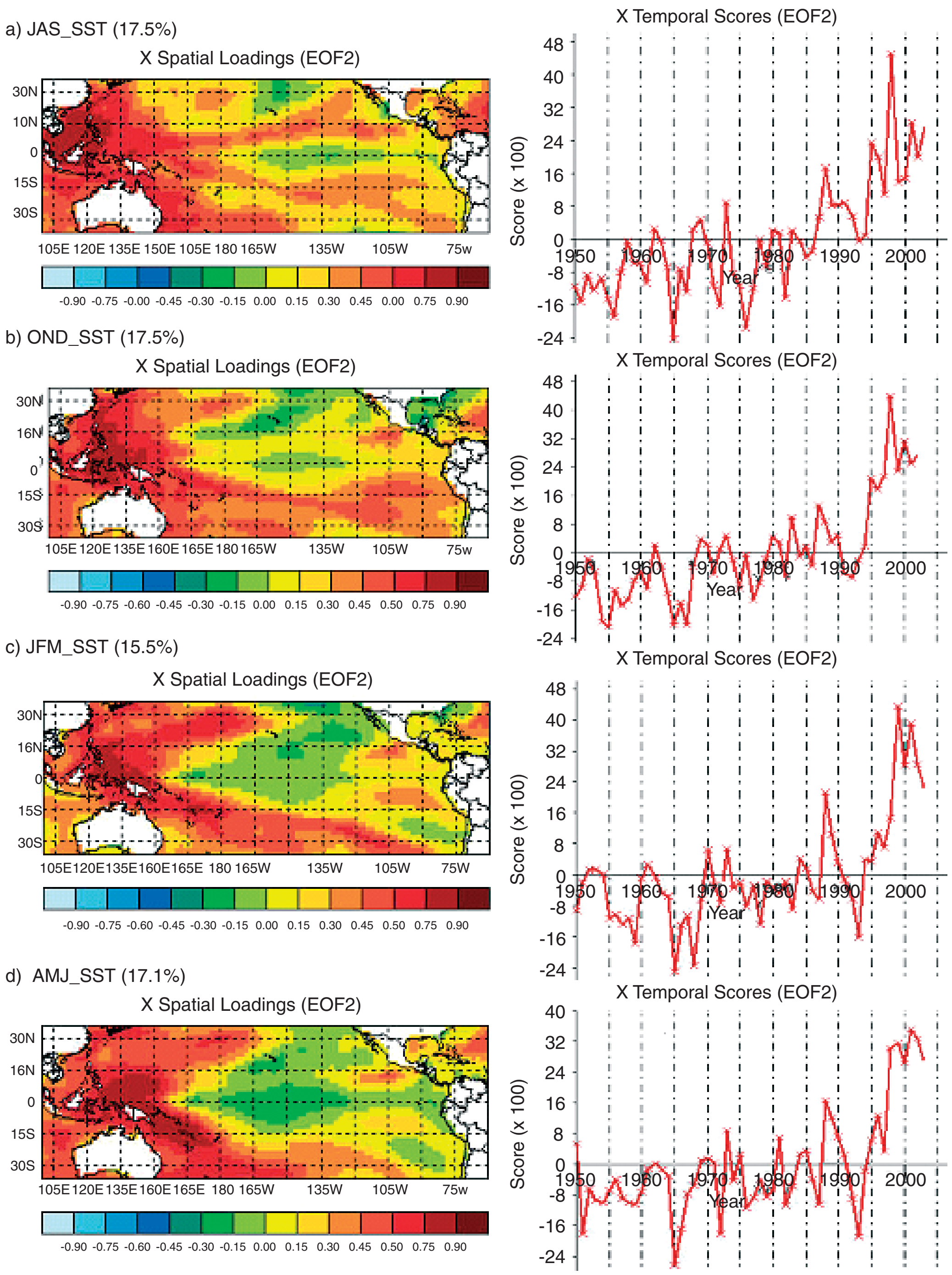

Figure 5. See as Figure 4, except for EOF2.

eastern and central Pacific. The time component of EOF2 is in contrast with that of EOF1; it shows a long-term upward trend with pronounced signal in the last 20 years (Figure 5 *right panel*). Similar features have also been found in Kawamura (1994) and Allan and Slingo (2002).

RESULTS

ENSO and seasonal sea-level variability

Results revealed that the sea-level variations in the northwestern tropical Pacific islands (Guam and Marshalls Islands) have been found to be sensitive to the ENSO cycle, with low sea level during El Niño and high sea level during La Niña events. Consistent with north-western Pacific islands, American Samoa also displayed pronounced fall and rise of sea levels during El Niño and La Niña years. However, as compared to Guam and Marshalls, American Samoa did show these variations with a 4–6 month delay. While the sea-level variations in Guam/Marshalls start showing up from July, it is usually January when the sea level in American Samoa starts to fluctuate. A comprehensive analysis of El Niño/La Niña events and the seasonal sea-level deviations for the USAPI has been reported in Chowdhury *et al.* 2006.

DOI: 10.1002/joc

CCA model forecast and Hindcast skill

The hindcast skill of the CCA model for 1950–2003 is estimated using a cross-validated scheme. Cross-validation is a generalization of the common technique of repeatedly omitting a few observations from the data, reconstructing the model, and then making forecasts for the omitted cases (e.g. Stone, 1974; Chu and He, 1994). Cross-validation is conducted to evaluate the overall forecasting skill of the CCA model. The cross validation is nonparametric and provides an unbiased estimate of forecast skill.

In this study, only one observation was removed at a time for each case. This is justified because interannual autocorrelations in the data are small. Therefore all data available were used except that for the season for which the prediction was targeted. For example, to forecast the summer (JAS) sea level of 1990 with 0-season lead time, a 53-year JAS time series of sea level (1950–89, 1991–2003) and a 53-year AMJ SST series (1950–1989, 1991–2003) were used to build a CCA model. Then this resulting CCA model was used to forecast sea-level values in summer 1990 using the independent SST values of spring 1990. The climatology is re-computed when each year is held out, and the anomaly of the target year is redefined in terms of the means of the other years. By doing this repeatedly, we obtain 54 forecasts of sea level which can be compared to the observed sea level. The moving-average season of three consecutive months was then used to identify the season with best predictability, yielding a total of 12 target seasons (JFM, FMA, MAM, AMJ, MJJ, JJA, JAS, ASO, SON, OND, NDJ, and DJF).

Table IV provides cross-validation correlation skills for the 12 moving seasons at 0-season lead time. Overall, sea-level forecasts for all the 12 seasons are reasonably well predicted (with 0-season lead) with a mean correlation skill of 0.635, 0.693, and 0.612 for Guam, Marshalls, and American Samoa, respectively.

The aforementioned result is applied to a zero season lead time. Now the question arises as to whether there is any predictability of sea level at time scales one to three seasons lead time. To provide a predictive skill at longer lead time, CCA cross-validation skills up to three seasons in advance for Guam, Marshalls, and ASamoa are presented in Figure 6 (*top panel*) and the average skills of these stations at 0 to 3 seasons lead time are presented in Figure 6 (*bottom panel*). As indicated, different islands show different levels of predictive skill. In general the forecast skills for JFM and AMJ target seasons in Guam, Marshalls, and American Samoa are well predicted with an average correlation skill of 0.583, 0.632, 0.604 (for JFM), and 0.607, 0.604, 0.642 (for AMJ) respectively. As the season advances, forecast skills fluctuate and the predictions skills in JAS are not as skillful as those of JFM and AMJ. For instance, JAS predictions skills in Guam and American Samoa are 0.490 and 0.567 respectively, when averaged from 0 to 3 seasons ahead.

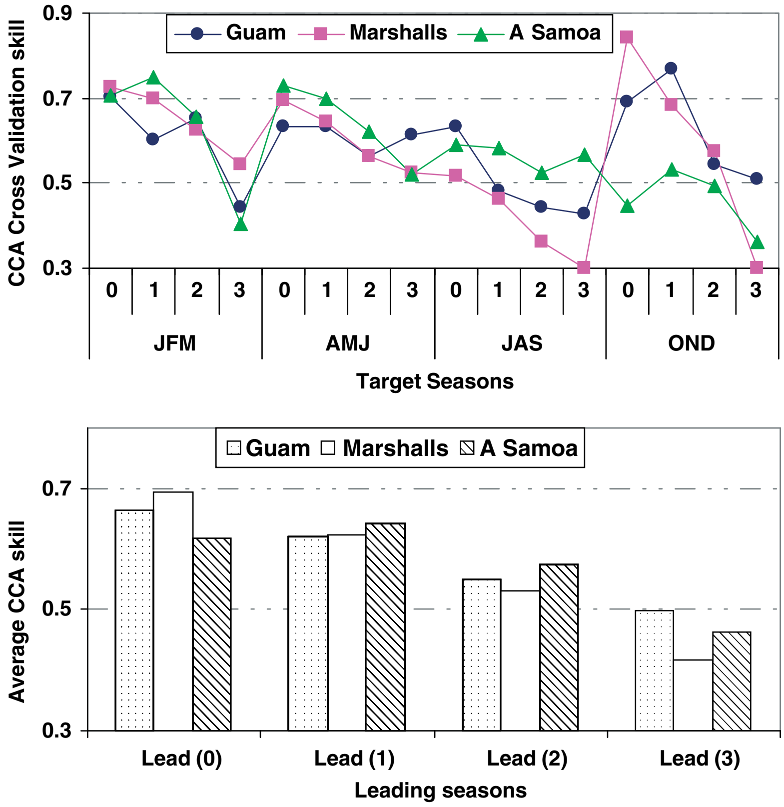

Figure 6. CCA cross-validation hindcast skills for Guam, Marshalls, and ASamoa (*top Panel*) and average skills at 0 to 3 seasons lead time (*bottom panel*). Note that 0, 1, 2, 3 represent zero, one, two, and three seasons lead time respectively. *For example, JFM-0, 1, 2, and 3 means 'sea-level' of target season JFM based on SSTs of previous OND, JAS, AMJ, and JFM*. Also see foot note in Tables II and IV for some useful definitions and abbreviations. This figure is available in colour online at www.interscience.wiley.com/ijoc

DOI: 10.1002/joc

Table IV. Zero season lead CCA cross-validation skill for 12 moving seasons (from JFM SST × AMJ SL to DJF SST × MAM SL) in Guam, Marshalls, and A Samoa.

Target season (Sea-level)	Predictor period (SST)	Guam	Marshalls	A Samoa	*Average*
AMJ	JFM	0.631	0.695	0.730	0.685
MJJ	FMA	0.619	0.699	0.708	0.675
JJA	MAM	0.624	0.638	0.646	0.636
JAS	AMJ	0.634	0.516	0.591	0.580
ASO	MJJ	0.592	0.528	0.421	0.514
SON	JJA	0.641	0.658	0.418	0.572
OND	JAS	0.691	0.841	0.448	0.660
NDJ	ASO	0.556	0.858	0.593	0.669
DJF	SON	0.710	0.844	0.663	0.739
JFM	OND	0.701	0.726	0.708	0.712
FMA	NDJ	0.680	0.654	0.716	0.683
MAM	DJF	0.540	0.660	0.705	0.635
	Average	0.635	0.693	0.612	

(Note: Lead time is the time interval between the end of the initial period and the beginning of the forecast period. Also see footnote in Table III for other abbreviations. *Forecasts are thought to be of* ***useful skill*** *(or at least fair skill) if the CCA cross-validation value is greater than 0.3. Higher skills correspond to greater expected accuracy of the forecasts. Skill levels greater than 0.4 and 0.5 are thought to be moderate and good, while skill levels greater than 0.6 are thought to be strong*).

The average predictive skill for Guam has been found to be strong (0.661 and 0.613) at 0 to 1-season lead time, the skill remained considerably high (0.551) at 2-season lead time, and the skill is still reasonably good (0.486) at 3-season lead time (Figure 6, *bottom panel*)'. Both the Marshalls and American Samoa displayed similar correlation skills −0.686, 0.631, 0.518, and 0.411 for Marshalls at 0 to 3-seasons lead time, and 0.619, 0.642, 0.593, and 0.440 for American Samoa at 0 to 3-seasons lead time respectively (Figure 6). If the average skills for all four seasons are averaged, Guam exhibits the highest skill in OND (0.631) and the lowest in JAS (0.490). The highest and lowest 4-seasons average skills for Marshalls is 0.632 in JFM and 0.403 in JAS, while American Samoa displayed the highest and lowest skills in AMJ (0.642) and OND (0.481) respectively.

One reason why JAS has weaker predictability is probably due to the effect of the spring predictability barrier. Previous studies by Yu *et al.* (1997) found it to be very difficult to generate accurate rainfall forecasts for the USAPI using spring SSTs as predictors. However, unlike rainfall prediction, where the spring-barrier effect has contributed to the difficulty of rainfall predictions, this was a somewhat weaker obstacle for SST-based sea-level predictions. The reason why JFM and AMJ have better predictability is that ENSO responses are most pronounced during boreal winter as the Pacific SSTs and Southern Oscillation index anomalies reach their peaks during boreal winter/spring.

Principal component regression (PCR) model forecast

The same cross-validation scheme was employed to assess the PCR model validity and adequacy. Figure 7 (*top panel*) shows the lead time-target season cross-section of PCR cross-validation skill for the tide gauge

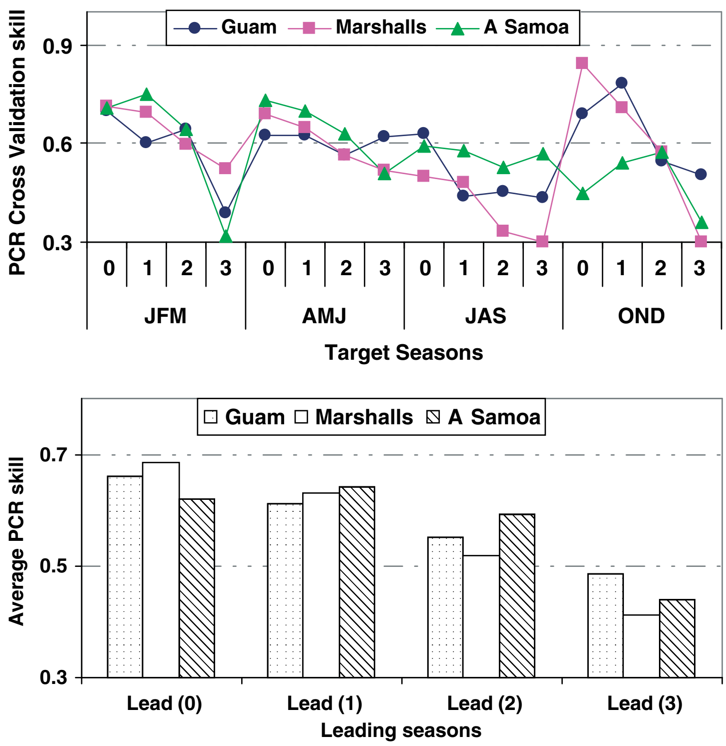

Figure 7. Same as Figure 6, for PCR cross-validation hindcast skills for Guam, Marshalls, and ASamoa (*top panel*) and average skills of these stations at 0 to 3 seasons lead (*bottom panel*). This figure is available in colour online at www.interscience.wiley.com/ijoc

DOI: 10.1002/joc

stations in Guam, Marshalls, and American Samoa. These three stations have provided almost identical value with the previous CCA analyses. The other three stations at RPalau, CNMI, and FSM (not reported here) also provided similar skills. The PCR model also provided reasonably skillful prediction with one to three seasons lead (Figure 7, *bottom panel*).

Skills of CCA, PCR, and other sea-level forecast models – comparative perspective

When comparing the skills of CCA and PCR models, both the models provided similar skillful results. While the CCA model-based average cross-validation correlation skills for Guam, Marshalls, and American Samoa are 0.659, 0.628, 0.552, and 0.460 at 0-, 1-, 2-, and 3-seasons lead respectively, the corresponding PCR model skills are found to be 0.655, 0.628, 0.554, and 0.446 respectively. Other than some sampling effects, the CCA and PCR model are found to be equally skillful in handling sea-level forecasting problems for the USAPI.

The skills of CCA and PCR models are compared with the skills of persistence forecasts. It has been found that at 0-season lead the correlation skill for persistence forecast are comparable to CCA and PCR models. While the average correlation skill for CCA and PCR models at 0-seasons lead are 0.659 and 0.655, the average skills for persistence forecasts for the same time frame is 0.586. However, as the lead time increases to 1 to 3-seasons, the correlation skill for persistence forecasts appeared to be weaker and were found to be poor when compared to CCA and PCR models. Therefore, the CCA and PCR models are more skillful than persistence forecasts in handling sea level prediction schemes.

It is, however, worth comparing the skill of the CCA model with other models forecasting the sea level in the western Pacific region. One of the available models is the Markov model for sea-level forecasts (see Xue and Leetmaa, 2000; available at the web site: http://www.cpc.ncep.noaa.gov/products/people/yxue/MKmodel_sl_clim7100_godas/corr_SLfore_TG27_weaver.gif). As compared to the Markov model, our CCA forecast model provides almost comparably skillful results for the 0-season and 1-season lead time. As the lead time increases, for example with 2-seasons and 3-seasons lead time, our model provides a slightly better skill than the Markov model. Another important finding is that, as compared to the Markov process, our model captured a much better skill for the Marshall Islands. While the Markov model provides skill levels of $0.3 \sim 0.4$ and $0.2 \sim 0.3$ at 2- and 3-seasons lead for Marshall Islands, the skill of our CCA model has been found to be 0.532 and 0.418 respectively. While skill levels for tide gauge stations are likely to be slightly lower than the sea level anomalies (see Xue and Leetmaa, 2000), our CCA model provided skillful results almost similarly comparable to the Markov model of sea-level anomalies. This is because our CCA model focuses on a particular region of the Pacific (USAPI); we have identified the SSTs of the region that displays the highest correlation with the sea-level variation for some specific stations and, therefore a better skill is visible.

It may be mentioned here that, among the USAPIs, the Marshall Islands are extremely sensitive to sea-level variability and change. The 1997–98 droughts – when the sea level in Majuro (Marshalls) was considerably lower (i.e. $2 \sim 12$ inches) than the average – required the installation of large desalination plants to meet the drinking water needs of the densely packed urban populace. We do not have a model-based study to correlate drought and sea-level variations in Marshalls; however, various published/unpublished reports indicate that, in addition to deficient rainfall, this drought was also caused by the sea-level variations (see Shea *et al.*, 2001 and references therein). The stakeholders therefore demand accurate information on the height, time, and duration of sea-level forecasts in specific climate sensitive locations. On the basis of the regional climate dynamics, the prime responsibility of PEAC is to prepare scenarios for these sensitive locations about the probable impact of the extreme event for the upcoming season.

In addition to these three islands, the CCA model has also been tested with other islands (RPalau, CNMI, and FSM). The skill level has been found to be similarly high for these islands. Also, as part of our efforts to develop an operationally seasonal sea-level forecasting scheme, we have also tested our model with some other islands located in the U.S Trust Territory (*Wake, Johnston*) and some Hawaii stations (*Hilo, Sand, Kahului, Mokuoloe*, and *Nawiliwili*). The CCA cross-validated skills for these islands were not found to be as good as those observed in the USAPI (skill level lies between 0.3 and 0.5). Therefore, a CCA model, based on the tropical Pacific SST alone, may not be able to accurately capture the sea-level fluctuations for these islands in the US Trust and Hawaii. Perhaps a complete overhaul with a different SST region (i.e. other than in the tropics) and with additional oceanic/atmospheric variables (indices) would be necessary to raise the predictive skill of sea-level forecasts for these islands. Research related to this issue is in progress and will be reported when available.

SUMMARY AND CONCLUSION

Following an overview of the annual cycle – which is very strong in the north Pacific Islands and relatively weak in the South Pacific Islands – and evaluating the impact of ENSO on the seasonal sea-level variability – which resembles El Niño for low sea level and La Nina for high sea level – composite analyses for seasonal variation, and the effects of SST on sea level have been studied by linear correlation. The CCA and PCR methods are employed to forecast sea level. Findings are summarized as follows:

(1) There is a pronounced lead/lag correlation between sea-level variability in JAS and the fluctuations of the tropical SST. The correlation intensifies as the

DOI: 10.1002/joc

season advances from AMJ to JAS. For Guam and the Marshalls, the correlation indicates a linkage that is related to ENSO. But for American Samoa, this correlation indicates only moderate to weak linkage. Sea level variations at Guam and American Samoa are positively correlated with in situ SST, indicating positive sea-level anomalies with more heat content and vice versa. Such a relationship, however, is not found at Marshall Islands.

(2) The CCA model provides useful skill in predicting sea level in the Pacific Islands. When all stations are considered, overall sea-level forecasts for all the 12-moving seasons displayed a mean skill of 0.635, 0.693, and 0.612 (with 0-season lead) for Guam, Marshalls, and American Samoa respectively. Overall, sea-level forecasts for Guam, Marshalls, and American Samoa are well predicted with a strong and average skill level of 0.6 to 0.7 at zero to one season lead time. At two seasons lead, the average skill level slightly drops to between 0.5 and 0.6. At three seasons lead time, this skill level drops to between 0.4 and 0.5. In every case, the observed sea level shows large enough seasonal variability and the predicted (hindcast) sea level fits to a large degree to such variability. Therefore, given the well-known ENSO impact on USAPI rainfall, the Pacific SSTs thus offer additional advantages in our predictions of Pacific island sea-level variability.

(3) A problem with this CCA model is that some of the ENSO events – like the years of 1951, 1958, 1972, 1982–83, 1987–89, and 1997–98 – have been found to be underestimated in some of the locations. Barnston *et al*. (1999) showed that this problem is common among ENSO forecast models in the prediction of the 1997–98 warm events. It has been suggested that atmospheric high-frequency variabilities, e.g. westerly wind bursts and Madden-Julian Oscillations, played critical roles in the timing and strength of the 1997–98 warm event (Yu and Rienecker, 1998). Xue and Leetmaa (2000) observed that the warm events in 1982–83 and 1997–98 and the cold event in 1988–89 are seriously underestimated in their Markov model. They also suggested that the atmospheric high-frequency variabilities played important roles in the timing and strength of the 1982–83 warm and 1988–89 cold events. We have not made any additional efforts to investigate this matter further, but we also assume that similar causes, as identified by other authors, are responsible for underestimation in some of the major ENSO years in this research too.

Pacific Island communities are most vulnerable to climate variability. Advance information on sea-level variability can contribute significantly to hazard preparedness actions of the people of these islands. Based on this operational forecasting technique by the CCA model, the PEAC has already started publishing the real-time forecast of sea-level deviations at the official web site of PEAC (http://lumahai.soest.hawaii.edu/Enso/peu/update.html) with the caption 'Experimental Sea level Forecasts (*deviations w.r.t. climatology*) for the USAPI. This information has also been distributed through the printed issue of *Pacific ENSO Update* newsletter. Significant efforts are being made to add more new tide gauge stations from the USAPI so that a better picture of sea-level variations is available. The enhancement of current forecasting and warning capabilities with seasonal sea-level information and forecasts offers the potential for greater latitude in planning and decision options regarding hazard management in the USAPI.

ACKNOWLEDGEMENTS

This project was funded by cooperative agreement NA17RJ1230 between the Joint Institute for Marine and Atmospheric Research (JIMAR) and the National Oceanic and Atmospheric Administration (NOAA). The views expressed herein are those of the authors and do not necessarily reflect the views of NOAA or any of its subdivisions. This is SOEST contribution 6573.

Grateful acknowledgements are expressed to Tony Barnston (International Research Institute for Climate Prediction) for his careful reading and reviewing of this paper. We are also grateful to Dr. Mark Merrifield, Professor Michael Hamnett, Mr. James Weyman, Ms. Eileen Shea, and M/s. Ousmane Ndiaye and Zin Zhao for providing valuable comments during the entire period of this research. Comments of other PEAC officials M/S Chip Guard and Mark Lander are also highly appreciated. Special thanks are due to Diane Henderson for proof editing. We express our gratitude to the Sea Level Center (University of Hawaii) and the IRI for Climate Prediction (IRI) for providing us easy access, manipulation, and visualization of earth science data.

REFERENCES

Allan RP, Slingo A. 2002. Can current climate forcings explain the spatial and temporal signatures of decadal OLR variation? *Geophysical Research Letters* **29**(7): 45, 1–4.

Barnston AG. 1994. Linear statistical short-term climate predictive skills in the Northern Hemisphere. *Journal of Climate* **7**: 1513–1564.

Barnston AG, He Y. 1996. Skill of CCA forecasts of 3-month mean surface climate in Hawaii and Alaska. *Journal of Climate* **9**: 2020–2035.

Barnston AG, Smith TM. 1996. Specification and prediction of global surface temperature and precipitation from global SST using CCA. *Journal of Climate* **9**: 2660–2697.

Barnston AG, Glantz MH, He Y. 1999. Predictive skill of statistical and dynamical climate models in SST forecasts during the 1997–98 El Niño episode and the 1998 La Niña onset. *Bulletin of the American Meteorological Society* **80**: 217–243.

Bjerknes J. 1966. A possible response of the atmospheric Hadley circulation to equatorial anomalies of ocean temperature. *Tellus* **18**: 820–829.

Bjerknes J. 1969. Atmospheric teleconnections from the equatorial Pacific. *Monthly Weather Review* **97**: 163–172.

Bloomfield P. 1976. *Fourier Analysis of Time Series: An Introduction*. John Wiley and Sons: New York; 129–137.

Caldwell P, Kilonsky B. 1992. Data processing and quality control at the TOGA Sea Level Center. Joint IAPSO-IOC Workshop on Sea Level Measurements and Quality Control, IOC Workshop Report No. 81. UNESCO: Paris, 12–13 October, 1992, 122–135.

Chowdhury MR, Chu P-S, Schroeder T. 2006. ENSO and seasonal sea-level variability – A Diagnostic discussion for the U.S.-Affiliated Pacific Islands. *Theoretical and Applied Climatology* (In press). available at www.springerlink.com/content/1434-4483.

Cherry S. 1996. Singular value decomposition analysis and canonical correlation analysis. *Journal of Climate* **9**: 2003–2009.

Cherry S. 1997. Some comments on singular value decomposition analysis. *Journal of Climate* **10**: 1759–1761.

Chu P-S. 1995. Hawaii rainfall anomalies and El Niño. *Journal of Climate* **8**: 1697–1703.

Chu P-S, He Y. 1994. Long-range prediction of Hawaiian winter rainfall using canonical correlation analysis. *International Journal of Climatology* **14**: 659–669.

Chu P-S, Chen H. 2005. Interannual and interdecadal rainfall variations in the Hawaiian Island. *Journal of Climate* **18**: 4796–4813.

Draper NR, Smith H. 1981. *Applied Regression Analysis*, 2nd edn. John Wiley & Sons Inc.: New York: 709.

Hotelling H. 1935. The most predictable criterion. *Journal of Educational Psychology* **26**: 139–142.

Hsiung J, Newell RE. 1983. The principal nonseasonal modes of variation of global sea surface temperature. *Journal of Physical Oceanography* **13**: 1957–1967.

Jolliffe IT, Jones B, Morgan B. 1986. Comparison of cluster analyses of the english personal social services authorities. *Journal of the Royal Statistical Society. Series A (Statistics in Society)* **149**: 253–270.

Kalnay, Co-authors. 1996. The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society* **77**: 437–472.

Kawamura R. 1994. A rotated analysis of global sea surface temperature variability with interannual and interdecadal Scales. *Journal of Physical Oceanography* **24**: 707–715.

Kilonsky BJ, Caldwell P. 1991. In pursuit of high-quality sea level data. *IEEE Oceans Proceedings* **2**: 669–675.

Lau NC. 1985. Modeling the seasonal dependence of the atmospheric response to observed El Niño. *Monthly Weather Review* **113**: 1970–1996.

Murphy SJ, other Co-authors. 2001. Seasonal forecasting for climate hazards: prospects and response. *Natural Hazards* **23**: 171–196, Springer-Netherlands.

Newman M, Sardeshmukh PD. 1995. A caveat concerning singular value decomposition. *Journal of Climate* **8**: 352–360.

Nitta T, Yamada S. 1989. Recent warning of tropical sea surface temperature and its relationship to the Northern Hemisphere circulation. *Journal of the Meteorological Society of Japan* **67**: 375–383.

Pacific ENSO Update (A Bulletin of the Pacific ENSO Applications Center, ed. Nicole Colasacco). 2005. *ENSO and Sea Level Variability – Climatology of Annual Cycle*, 11, 1.

Pacific ENSO Update (A Bulletin of the Pacific ENSO Applications Center, ed. Nicole Colasacco). 2006. *Trends in Sea Level Extremes in the USAPI*, 12 : 2, (also available at http://lumahai.soest.hawaii.edu/Enso/peu/update.html).

Reynolds RW. 1988. A real-time global sea surface temperature analysis. *Journal of Climate* **1**: 75–86.

Reynolds RW, Marsico DC. 1993. An improved real-time global sea surface temperature analysis. *Journal of Climate* **6**: 114–119.

Ropelewski CF, Halpert MS. 1987. Global and regional scale precipitation patterns associated with the El Niño–Southern Oscillation. *Monthly Weather Review* **115**: 1606–1626.

Shea EL, Co-authors. 2001. Preparing for a changing climate: The potential consequences of climate variability and change. A report of the Pacific Islands Regional Assessment Team, East-West Center, Honolulu, Hawaii. East-West Center: 102.

Smith TM, Reynolds RW. 2002. Extended reconstruction of global sea surface temperatures based on COADS Data (1854–1997). *Journal of Climate* **7**: 1495–1510.

Stone M. 1974. Cross-validatory choice and assessment of statistical predictors. *Journal of the Royal Statistical Society Series B* **36**: 111–147.

Tribbia JJ. 1991. *The Rudimentary Theory of Atmospheric Teleconnections Associated with ENSO. Teleconnections Linking World Wide Climate Anomalies*, Glantz MH, Katz RW, Nicholls N (eds). Cambridge University Press; 285–308.

Xue Y, Leetmaa A. 2000. Forecasts of tropical Pacific SST and sea level using a Markov model. *Geophysical Research Letters* **27**: 2701–2704.

Xue Y, Leetmaa A, Ji M. 2000. ENSO prediction with Markov models: the impact of sea level. *Journal of Climate* **13**: 849–871.

Yu L, Rienecker MM. 1998. Evidence of an extratropical atmospheric influence during the onset of the 1997–98 El Niño. *Geophysical Research Letters* **25**: 3537–3540.

Yu Z-P, Chu P-S, Schroeder T. 1997. Predictive skills of seasonal to annual rainfall variations in the U. S. affiliated Pacific islands: Canonical correlation analysis and multivariate principal component regression approaches. *Journal of Climate* **10**: 2586–2599.

DOI: 10.1002/joc