INTERNATIONAL JOURNAL OF CLIMATOLOGY, VOL. 18, 215–224 (1998)

SHORT-TERM CLIMATE PREDICTION OF MEI-YU RAINFALL FOR TAIWAN USING CANONICAL CORRELATION ANALYSIS†

PAO-SHIN CHU*

Department of Meteorology, School of Ocean and Earth Science and Technology, University of Hawaii, Honolulu, Hawaii, USA
email: chu@soest.hawaii.edu

Received 29 November 1996
Revised10 July 1997
Accepted 12 July 1997

ABSTRACT

Canonical correlation analysis (CCA) is used for predicting Mei-Yu (May–June) rainfall for eight major stations in Taiwan based on the antecedent November–December sea-surface temperatures (SSTs) over the Pacific Ocean (50°N–40°S, 120°E–90°W). To reduce the large dimensionality of the SST data set, an empirical orthogonal function analysis is first performed and the leading nine eigenmodes of SST are retained as predictors. The root-mean-square error and the Pearson product-moment correlation coefficient are used to serve as a yardstick in overall forecast evaluation.

Forecasts are made for the period 1986–1995, which is independent from the developmental data sets. A moderate skill is achieved for most stations. In particular, Mei-Yu rainfall is more predicable for the last 3 years (1993–1995), when the island experienced a long spell of deficient rainfall. A cross-validation technique is used to estimate the overall hindcast skill of the CCA model for the period of 1956–1995 and results suggest that certain stations have more skill than others. Likewise, a CCA 'climatological prediction' is conducted in a cross-validated mode. © 1998 Royal Meteorological Society.

KEY WORDS: Mei-Yu rainfall prediction for Taiwan; canonical correlation analysis; sea-surface temperature over the Pacific; cross-validation; Taiwan

INTRODUCTION

Climate prediction is vital to many human activities and drought has been a recurrent and troublesome problem for Taiwan. Water resource managers need to know, for instance, rainfall forecasts well before the season so that they can make a scientifically based decision regarding whether to issue water conservation or water rationing for the regions expected to be affected. This would in turn effectively minimize the hardship incurred to customers (e.g. farmers, industry users) and/or residents during periods of drought. By incorporating the forecast information into long-range resources planning and management, decision-makers can take more pro-active instead of reactive action.

Climatologically, Mei-Yu rainfall in Taiwan occurs from mid-May to mid-June, 1 month earlier than its counterpart over central and eastern China (Tao and Chen, 1987; Ding, 1994). Figure 1 shows Mei-Yu rainfall climatology as well as the location of first-order stations in Taiwan. In the central and southern part of Western Taiwan, where population is dense, Mei-Yu rainfall makes up more than 30 per cent of the annual rainfall totals. Because of the large percentage of annual rainfall being produced in a short time period, there is considerable interest in Mei-Yu rainfall prediction. This situation became critical when the Mei-Yu season was heralded by a prolonged drought, as exemplified by the dry summer and autumn of 1995 and the successive, dry winter of 1996. Whether the dry condition would have persisted into the Mei-Yu season of 1996 posed a great challenge to decision-makers during the first few months of 1996.

*Correspondence to: P.-S. Chu, Department of Meteorology, School of Ocean and Earth Science and Technology, University of Hawaii, Honolulu, Hawaii, 96822, USA.
† SDEST Contribution Number 4533

CCC 0899-8418/98/020215-10 $17.50

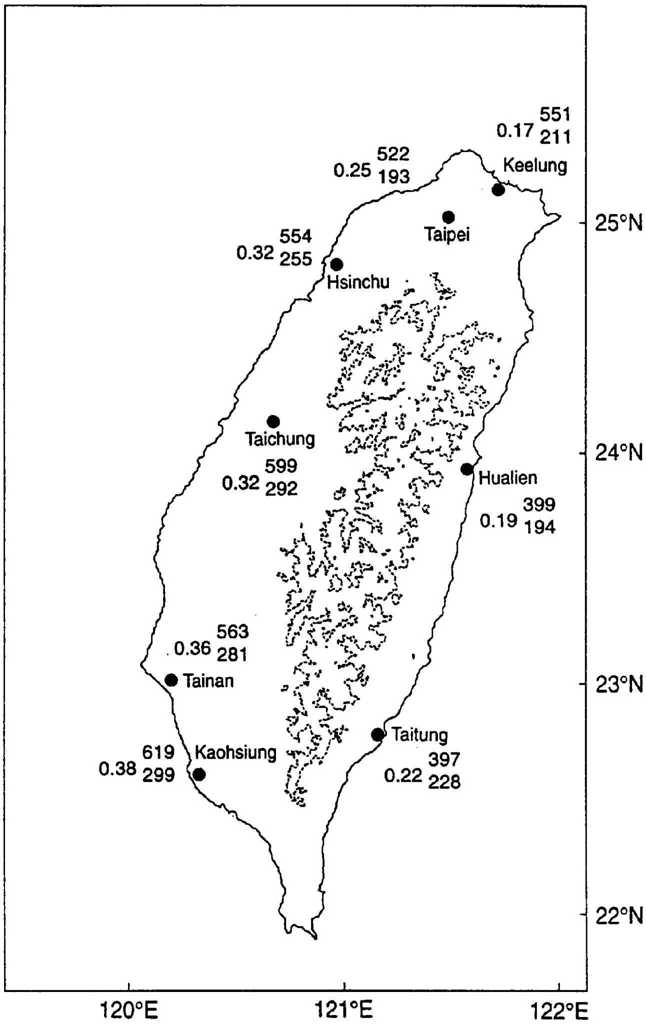

Figure 1. Mei-Yu rainfall climatology and the locations of eight major stations in Taiwan. Numbers on the left at each station denote the ratio of May–June rainfall to the total annual rainfall. Numbers in the upper (lower) right denote mean (standard deviation) of May–June rainfall (mm). The base period is 1956–1995. Dotted line denotes the terrain contour of 1500 m

In this study, we attempt to forecast Taiwan Mei-Yu rainfall using the Pacific sea-surface temperatures (SSTs) with an advanced technique called the canonical correlation analysis (CCA). Many studies have shown that SST is the most important boundary forcing on the atmospheric circulation (e.g. Hastenrath and Heller, 1977). Canonical correlation analysis has been used as a major method to operationally forecast the surface climate for a number of stations in the continental USA (Barnston, 1994), to forecast winter rainfall in Hawaii (Chu and He, 1994) and seasonal precipitation in other tropical Pacific islands (He and Barnston, 1996; Yu *et al.*, 1997, hereafter referred to as YCS).

DATA AND METHODS

The monthly mean SST data over the Pacific (50°N–40°S, 120°E–90°W) from 1955 to 1994 are used. This data set is obtained from the U.S. National Centers for Environmental Prediction. It has a resolution of 10° latitude by 10° longitude, yielding 125 grids over the Pacific. Long-term monthly rainfall records for eight first-order stations

(Figure 1) are supplied by the Taiwan Central Weather Bureau. To be consistent with SST data, we extract rainfall records from these eight stations for the period from 1956 to 1995.

Data from these respective months are first summed to produce a bi-monthly total. The bi-monthly data are then normalized by taking the difference between the individual bi-monthly value and the long-term average of the bi-monthly value, and dividing this difference by the standard deviation of the bi-monthly series. This normalization method is applied to each grid-point of the SST field and to each rainfall station. The normalization procedure ensures a zero mean value and unit variance for each series.

Canonical correlation analysis is a relatively advanced statistical technique, permitting the decomposition of circulation patterns between two fields into modes. Put another way, CCA identifies new variables that maximize the interrelationships between the two data sets (e.g. Nicholls, 1987) or two fields (e.g. YCS). For mathematical details on CCA, see Tatsuoka (1988) and Wilks (1995). Chu and He (1994) give a demonstration on how the CCA prediction equation is constructed.

BACKGROUND CLIMATE OF RAINFALL

Long-term averaged May–June rainfall totals and their standard deviation for eight stations are also shown in Figure 1. Climatologically, Mei-Yu rainfall is higher on the western side than the eastern side of the island. This spatial distribution of rainfall pattern reflects an interplay between the Mei-Yu front, the western Pacific subtropical high pressure ridge, the low-level south-westerly flows from the Bay of Bengal, and the island's orography.

The Mei-Yu front is a rain-bearing synoptic/mesoscale system which constantly moves from southern China toward Taiwan durng this pre-summer monsoon season. Because of the blocking effect by the Central Mountain Range, which is basically orientated from north to south with an average height of 2 km, rainfall during the Mei-Yu season is much less on the east side (leeward) of the island than that on the west side (windward). The position and intensity of the western Pacific subtropical high pressure ridge are related to the anomalous SST warming or cooling in the equatorial Pacific (Ding, 1994); if the ridge extends far enough westward towards Taiwan over a long period of time, subsidence prevails and a dry Mei-Yu season ensues. The role of the south-westerly flows from the Bay of Bengal is to provide a major and widespread moisture source for southern China and Taiwan.

In Figure 1, standard deviations of May–June rainfall are higher in regions where the mean values are higher. Accordingly, higher values (greater than 250 mm) are found in Hsinchu, Taichung, Tainan and Kaohsiung. It is in this region (western Taiwan) where changes in Mei-Yu rainfall are expected to be large from one year to another (i.e. larger interannual variability).

DATA PREPROCESSING, CCA PATTERNS AND PROCEDURES FOR PREDICTION

As in Chu and He (1994), we have computed the correlations between the bi-monthly SST in the Pacific and the following May–June rainfall and found the values and patterns are most conspicuous when the preceding November–December SST is correlated with Mei-Yu rainfall. The November–December SST values are thus used as predictor variables and the following May–June rainfall records are used as predictand variables. In order to compact the predictor data field (125 grids times 40 years) and to filter out the small noise not important in the prediction experiments, the SST field over the Pacific is subjected to empirical orthogonal function (EOF) analysis.

Figure 2(a) shows the spatial pattern of the first SST eigenmode, which accounts for about 28 per cent of the total variance. The spatial pattern reveals large, positive loadings over the equatorial Pacific, with a centre near the south-central to eastern Pacific. Negative loadings are found in the subtropics and mid-latitudes, on both flanks of the area of positive loadings. Temporal variability of this mode (Figure 2(b)) is characterized by extremes during El Niño and La Niña years; for example, positive excursions occur during El Niño events (e.g. 1982, 1986–1987, 1991–1994) and negative excursions during La Niña events (e.g. 1975, 1988). Interestingly, time series of the eigen coefficient of this mode also show a tendency of long-term shift toward positive coefficients starting from 1976, probably reflecting a decadal-scale variability in the coupled, tropical ocean–atmosphere system (e.g. Trenberth and Hurrel, 1994; Graham, 1995). The first nine modes of the Pacific SST,

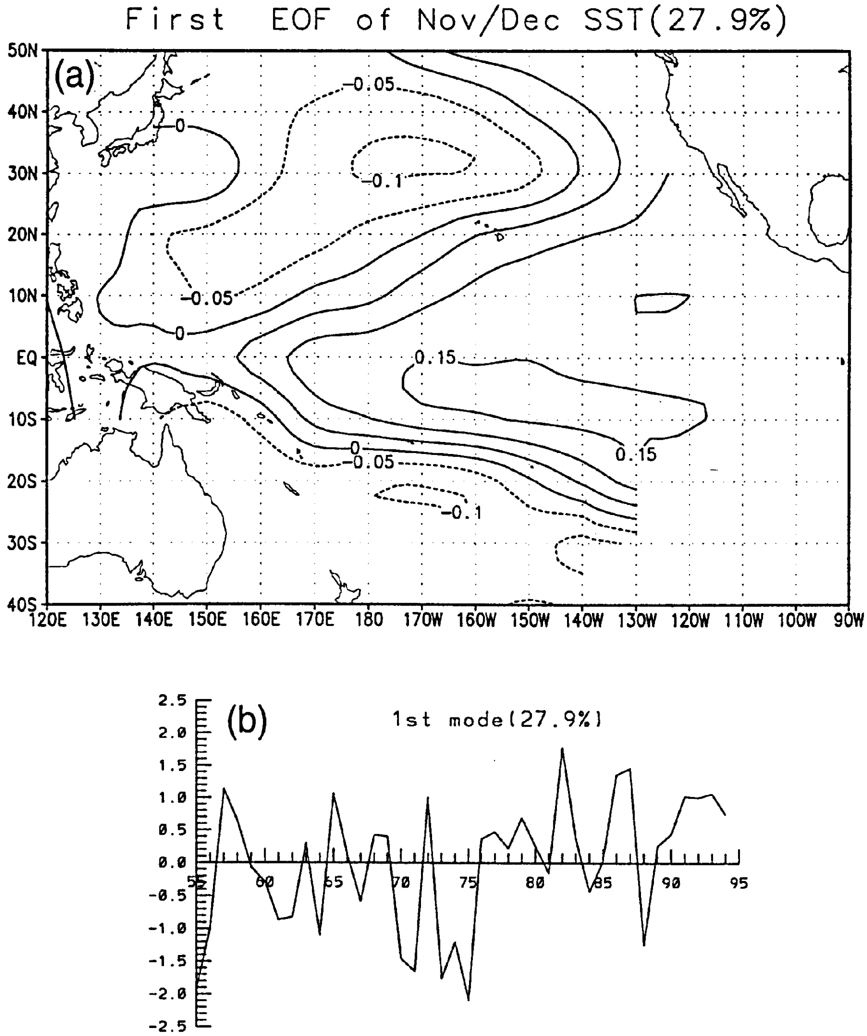

Figure 2. (a) Spatial pattern of the first eigenmode of November–December sea-surface temperatures over the Pacific Ocean, 1955–1994. (b) Time series of the coefficients associated with the first eigenmode of November–December sea-surface temperatures over the Pacific Ocean, 1955–1994

which account for more than 72 per cent of the total variance in the entire domain, are thus selected as the predictor vector.

The SST pattern of the largest canonical correlation is characterized by positive centres in low latitudes of the central–eastern Pacific and negative values in the higher latitudes and in the western Pacific (Figure 3(a)). The CCA rainfall pattern exhibits a mixed structure, with a large positive value in the centre flanked by a pair of smaller regions of opposite sign (Figure 3(b)), suggesting that rainfall response to the variations in the Pacific SST is not uniform across the island. Thus, increased surface heating over the tropical central–eastern Pacific associated with cooling over the western pacific, a feature indicative of El Niño phenomena, tend to be followed by increased rainfall in central and decreased rainfal in northern and southern Taiwan. Conversely, anomalously cold water over the tropical central–eastern Pacific and warm water in the western Pacific are associated with decreased Mei-Yu rainfall in central Taiwan. Time series of the first canonical pair of SST and rainfall records reveals similar fluctuations for both series (Figure 3(c)). Therefore, the high correlation (0·78) between these two series is caused by a coherent variation of both series throughout the entire time period.

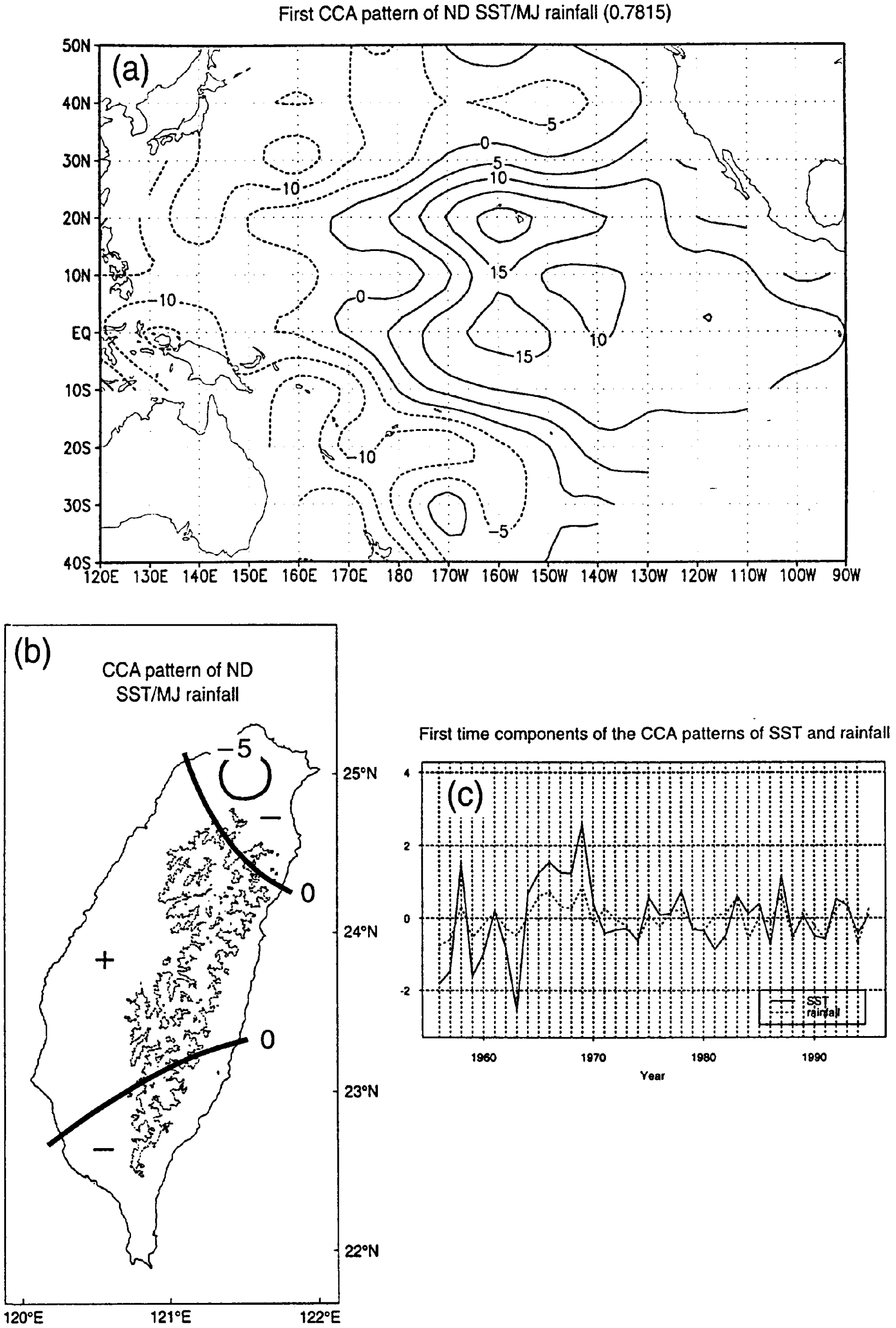

Figure 3. The first spatial canonical pattern pair for the canonical correlation between November–December Pacific SST and the following May–June rainfall in Taiwan. (a) The SST and (b) the rainfall. (c) Normalized time series components of the first CCA patterns of SST (solid) and rainfall (dashed) anomalies. For visualization purposes, the SST series are shifted backward by one year. For example, SST of 1985 in the time series plot is actually the 1984 value

A dependent data set is used for the model development and an independent portion of the record (1986 to 1995) is reserved for prediction experiments. For instance, a forecast for May–June 1986 rainfall is produced based on the CCA model using a data set consisting of the first nine eigenmodes of November–December SST (1955–1984) and the following May–June rainfall (1956–1985) indices. The SST in November–December 1985 is then served as an input to the CCA prediction model to forecast May–June 1986 rainfall. Similarly, a forecast for May–June 1987 rainfall is made using the SST in November–December 1986 as input to the updated CCA prediction model, which uses the nine leading modes of SST (1955–1985) and the eight rainfall series (1956–1986), and so on.

FORECAST AND VERIFICATION

To demonstrate forecast skills, we will first show time series of forecasts and observations for four stations that are representative of four geographical regions in Taiwan. Keelung is selected to represent the northern region and Taichung for central Taiwan (Figure 1). Tainan and Taitung represent southern and eastern Taiwan, respectively. Note that these four stations are commonly used by the Central Weather Bureau in issuing the seasonal weather outlook for the entire island.

For Keelung, forecasts for some years are close to observations, such as 1986 Figure 4(a)). Furthermore, forecasts for the last 3 years, namely from 1993 to 1995, are in reasonable agreement with observations. For Taichung, (Figure 4(b)), there is a larger discrepancy between forecasts and observations as compared with Keelung. Moving southward, forecasts for Tainan are better than Taichung (Figure 4(c)). This is supported by the

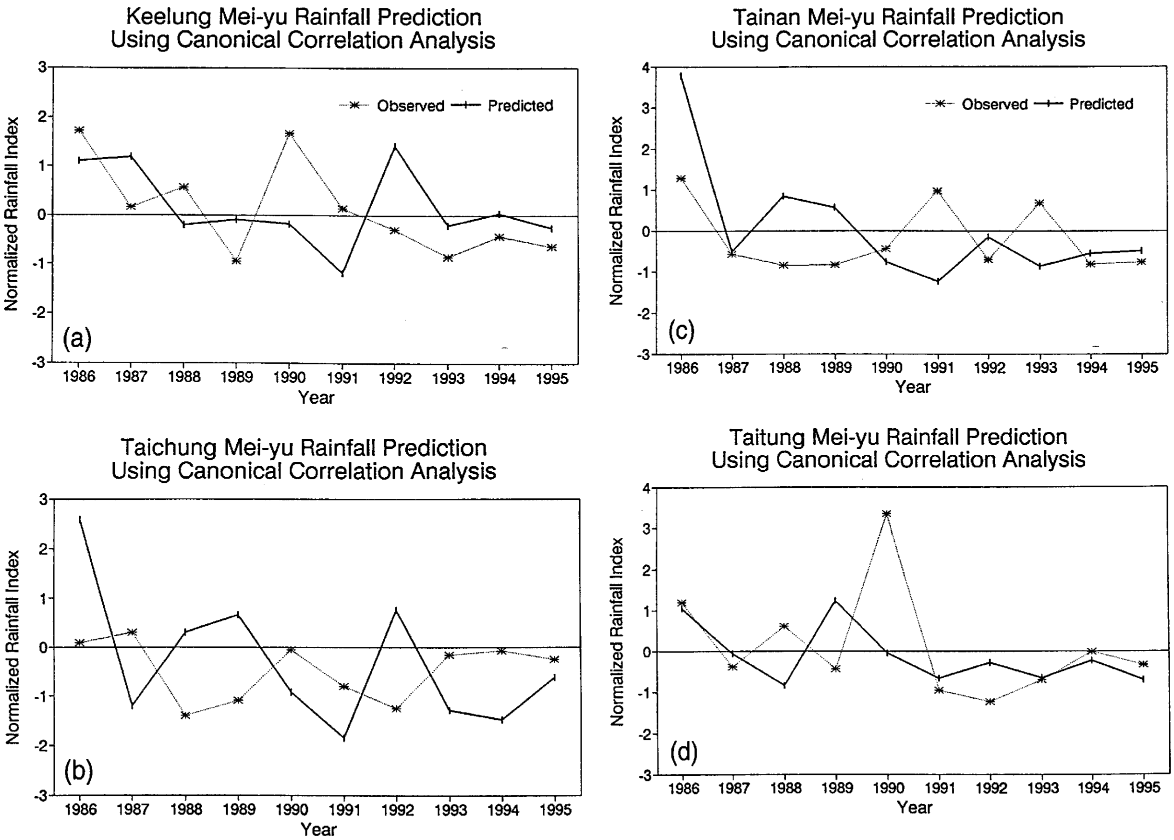

Figure 4. (a) Time series of predicted (solid line) and observed (broken line) Mei-Yu rainfall for Keelung by canonical correlation analysis from 1986 to 1995. (b) Same as (a), except for Taichung. (c) Same as (a), except for Tainan. (d) Same as (a), except for Taitung

successful forecasts for 1987, 1990, 1992, 1994, and 1995. Likewise, there is also some predictability for Mei-Yu rainfall in Tatung, as shown by the good forecasts for 1986, 1987, 1991, and 1993 to 1995 (Figure 4(d)). Therefore, one may say that the state of Mei-Yu rainfall can be reasonably predicted 4 to 6 months in advance for those years. The average Pearson product-moment correlation coefficient between the forecast and observation for 1986–1995 for all eight stations is about 0·375. It should be noted that the forecast skill comes from all nine modes, although the first mode may contribute the most skill as the correlation (0·78) between the first pair of canonical variables is the largest possible value between pairs of linear combinations of the SST and rainfall data sets.

As a further measure of the forecasting accuracy, the root-mean-square error (RMSE) between the forecast and observation is used. Among the eight stations tested, Hualien exhibits the largest error (1·527) and Keelung has the smallest error (1·077) during the last 10 years (Figure 5). The forecast error is relatively small in Tainan and Kaohsiung, where Mei-Yu rainfall constitutes an important component of the annual rainfall.

CROSS-VALIDATION

The aforementioned prediction experiment was applied to 10 Mei-Yu seasons. Given the fact that the sample size for this experiment is small, it is interesting to assess the overall forecasting ability of CCA through the use of a

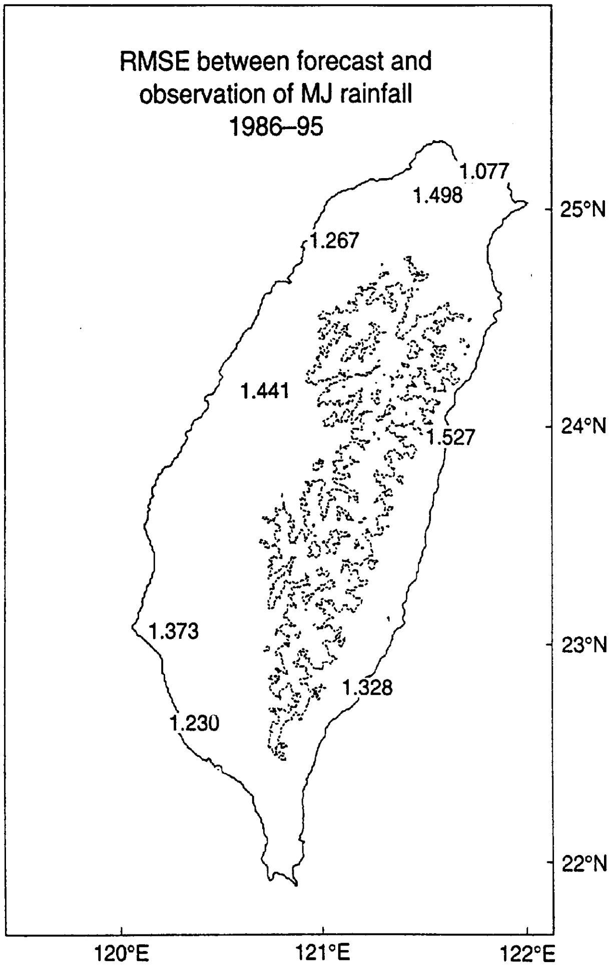

Figure 5. Root-mean-square error of Mei-Yu rainfall between observations and forecasts via CCA for the period 1986–1995

cross-validation technique. Cross-validation is a computer-intensive method of repeatedly omitting a few observations from the data, reconstructing the model, and then making estimates for the cases omitted (Chu and He, 1994).

Specifically, the predictor and predictand time series of N points are divided into L segments. Canonical correlation analysis is used to develop a prediction model using the $L - 1$ segments and is then used to hindcast the predictand in the withdrawn segment. This process is then repeated successively by changing the segment that has been excluded from the model development. By doing so, N hindcasts of Mei-Yu rainfall are obtained, which can be compared directly with N observations (1956–1995), to determine an overall skill of the CCA model. By comparing with the simulated real-time skill, CCA appeared to produce an unbiased estimate using cross-validation (Barnston *et al.*, 1994).

In this study, N equals 40 Mei-Yu seasons, which are divided into eight segments of five seasons each (e.g. 1961–1965). Holding out five seasons at a time would account for the autocorrelative structure that is present in some EOF time coefficients of the SST series (e.g. the third eigenmode has an interannual autocorrelation of about 0·20 up to lag five). Figure 6 lists the RMSE of the CCA hindcast cross-validation skill. The magnitude of error via cross-validation varies from 1·22 to 1·54, in a range comparable with that based on an independent forecasting set (Figure 5). However, when tested on a large data set via cross-validation, the spatial pattern of

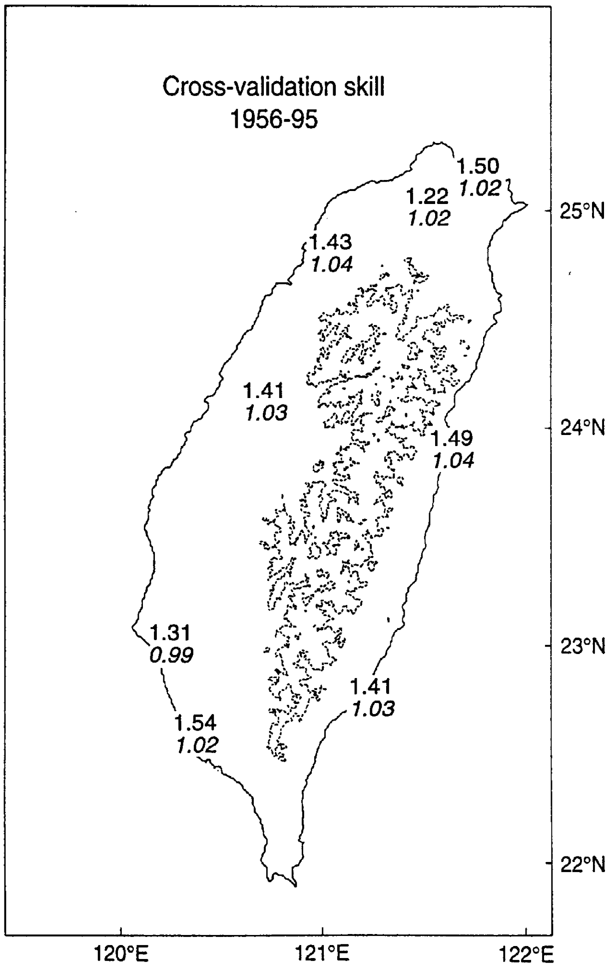

Figure 6. Numbers at the top of each station denote the root-mean-square error (RMSE) of Mei-Yu rainfall between observations and hindcasts, whereas numbers at the bottom denote the RMSE between observations and climatological predictions via CCA. The period is 1956–1995

errors is somewhat different from that based on short but forward-looking forecasts. In Figure 6, the two largest errors occur in the southern and northern tip of the island. Again, Hualien shows a relatively large error. In contrast, Taipei and Tainan exhibit the least error.

Another way to assess the CCA forecast skill is to compute the RMSE derived from the 'climatological prediction' in a cross-validated manner and compare this error with that from the hindcast as presented in the foregoing. Here, the 'climatological prediction' is defined in a way similar to the hindcast, except that the $L-1$ segments are used to calculate the mean standardized rainfall and then this mean value is used to 'forecast' the rainfall in the withdrawn segment (i.e. a constant value is chosen). To be consistent with the previous experiment, a five-season segment is applied.

Figure 6 also portrays the results of the RMSE of the 'climatological prediction'. The climatological prediction appears to have smaller errors than the corresponding one computed from the hindcast experiment. However, the Pearson correlation between the observation and 'climatological prediction' is $-0{\cdot}41$ when the average of eight stations is considered. This is in contrast to the average correlation coefficient of 0·15 between the observations and the hindcasts. Therefore, as judged from the commonly used measure of linear association between two variables, the cross-validated hindcast experiment has a better skill than the climatological prediction.

SUMMARY AND DISCUSSION

In this study, a scheme has been developed to explore the feasibility of short-term climate prediction of Mei-Yu rainfall in Taiwan with a CCA model. The predictor variables include the antecedent November–December SST over a very large domain of the Pacific Ocean. To compress the SST data sets prior to prediction experiments, an EOF analysis is used. A distinction is made to keep a dependent data set used as a training period and an independent data set (1986–1995) reserved for prediction experiments. All forecasts are made 4 to 6 months in advance.

For the last 10 years, a moderate skill is achieved for Mei-Yu rainfall prediction based on antecedent large-scale boundary forcings over the Pacific. The forecasting skills are better for the last 3 years (1993–1995) when Mei-Yu rainfall was low. The spatial pattern of root-mean-square error suggests that the error is smaller in the southern and western parts of Taiwan, implying more predictability in this region. This result is fortunate because Mei-Yu rainfall provides about one-third of annual rainfall in this region. Mei-Yu rainfall is more difficult to predict for Hualien, Taipei and Taichung. For the region as a whole, the average correlation coefficient between the forecasts and observations is about 0·375.

When tested on a large data set ($N=40$) via cross-validation, the magnitude of the error is comparable to that seen in the independent forecasting period. However, the spatial pattern of error is somewhat different from that tested during the recent 10 years. The hindcast error is smallest for Taipei and Tainan but largest for Keelung and Kaohsiung. Hualien, again shows a large error. The average correlation coefficient between the observations and hindcasts is 0·15 when all eight stations are considered. In order to assess the CCA hindcast skill, a comparison is made with the 'climatological prediction', which is also calculated in a cross-validated mode. A mixed result is obtained. That is, the 'climatological prediction' shows a lower error relative to the hindcast experiment; however, the average correlation coefficient between the 'climatological prediction' and the observation is $-0{\cdot}41$, indicating a negative (or no) skill from the 'climatological prediction'.

Recently, YCS studied rainfall predictability for a number of islands in the tropical western Pacific (not including Taiwan) using the Pacific SST as the only predictor variable. They showed that, if the target season is boreal winter (January through to March), CCA cross-validation has a high skill (i.e. average correlation coefficient more than 0·41 for $N=38$) when the lead time is one to two seasons; however, if the target season is boreal spring (April through to June), this skill drops noticeably to 0·24 when the lead time is one to two seasons. This result suggests that predictive skills of rainfall for the tropical western Pacific islands (e.g. Guam, Yap) vary with seasons, being higher in winter and lower in spring. The moderately low skill in spring implies that accurate climate prediction of precipitation in the tropical Pacific (perhaps including Taiwan) for certain seasons remains a challenging task even using a method as sophisticated as CCA.

It should be cautioned that our forecasts are based on the predictor field from the Pacific only. It is recognized that variations in mid-latitude synoptic systems over the East Asian continent (e.g. the position of the tropospheric longwave trough and ridge) and south-west moist monsoonal flows from the Indian Ocean, among others, may also be important in regulating Mei-Yu rainfall. In this regard, our current work should be reviewed as a first step in providing an empirically based approach toward seasonal to interannual climate prediction. In the future, we will continue to explore the predictability of Mei-Yu rainfall for Taiwan by combining the Pacific SST with other relevant general circulation parameters.

ACKNOWLEDGEMENTS

I wish to express my thanks to the staff members of the Central Weather Bureau of the Ministry of Transportation and Communications in Taiwan, Republic of China for their interest and support. Thanks are also extended to Brian Giles and one of the reviewers who provided constructive and insightful comments which improved the presentation of this paper. J. Wang assisted with data processing.

REFERENCES

Barnston, A. G. 1994. 'Linear statistical short-term climate predictive skill in the northern hemisphere', *J. Climate*, **5**, 1514–1564.
Barnston, A. G. *et al.* 1994. 'Long-lead seasonal forecasts—Where do we stand?', *Bull. Am. Meteorol. Soc.*, **75**, 2097–2114.
Chu, P.-S. and He, Y. 1994. 'Long-range prediction of Hawaiian winter rainfall using canonical correlation analysis', *Int. J. Climatol.*, **14**, 659–669.
Ding, Y. 1994. *Monsoon over China*, Kluwer Academic Publishers, Dordrecht, 419 pp.
Graham, N. E. 1995. 'Decadel-scale climate variability in the tropical and North Pacific during the 1970s and 1980s: observations and model results', *Climate Dyn.*, **10**, 135–162.
Hastenrath, S. and Heller, L. 1977. 'Dynamics of climatic hazards in northeast Brazil', *Q. J. R. Meteorol. Soc.*, **103**, 77–92.
He, Y. and Barnston, A. G. 1996. 'Long-lead forecasts of seasonal precipitation in the tropical Pacific islands using CCA', *J. Climate*, **9**, 2020–2035.
Nicholls, N. 1987. 'The use of canonical correlation to study teleconnections', *Mon. Wea. Rev.*, **115**, 393–399.
Tao, S. and Chen, L. 1987. 'A review of recent research on the East Asian summer monsoon in China', in *Monsoon Meteorology*, Chang, C.-P. and Krishnamurti, T. N. (eds), Oxford University Press, Oxford, pp. 60–92.
Tatsuoka, M. M. 1988. *Multivariate Analysis*, Macmillan Publishing Company.
Trenberth, K. E. and Hurrell, J. W. 1994. 'Decadel atmosphere–ocean variations in the Pacific', *Climate Dyn.*, **9**, 303–319.
Wilks, D. S. 1995. *Statistical Methods in the Atmospheric Sciences*, Academic Press, 467 pp.
Yu, Z.-P., Chu, P.-S. and Schroeder, T. A. 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', *J. Climate*, **10**, 2586–2599.

