The Recent Rainfall Climatology of the Mediterranean Californias

EDGAR G. PAVIA AND FEDERICO GRAEF

CICESE, Ensenada, Baja California, Mexico

(Manuscript received 20 August 2001, in final form 14 March 2002)

ABSTRACT

In this work, recent (1948–2001) rainfall data in a southwestern California station (San Diego) and a northwestern Baja California station (Ensenada) within a region called Mediterranean California, around 33°N, 117°W, are studied. Cumulative annual means are used as indicators of climatological variability; but the entire datasets are analyzed by modeling the histogram of each set as a Weibull distribution probability density function, f. The climatology of both stations, defined simply as the arithmetic average, is compared with their theoretical mean; that is, the first moment of f. It is assumed that this comparison would be indicative of the reliability of the available rainfall climatologies.

If these assumptions hold, in particular if the data is indeed Weibull distributed, it can be concluded that the climatological annual mean precipitation in this region is slightly overestimated at this time.

1. Introduction

Northwestern Baja California (Mexico) and southwestern California have a distinctive wintertime precipitation regime. This region, sometimes called Mediterranean California, encloses the cities of San Diego, California; Tijuana, Rosarito, and Ensenada, Baja California, Mexico, for which seasonal rainfall, averaging less than 300 mm yr^{-1}, is of great importance. This region's positive (negative) rainfall anomalies seem to be related to the warm (cold) phases of both the El Niño–Southern Oscillation (ENSO) and the eastern representation of the North Pacific Oscillation (NPO) (see, e.g., Pavia and Badan 1998; Gershunov et al. 1999). For the high-frequency ENSO relationship (Pavia 2000), annual precipitation anomalies are easily calculated by substracting the mean climatological value ($\overline{x}$) from the available records; but for the low-frequency NPO relationship the available records may produce an unreliable $\overline{x}$. In this paper $\overline{x}$ is defined as the arithmetic average that somewhat approximates the unknown, true climatology, $\mu(x)$; for example,

$$\overline{x}(N, x, i, \ldots) \equiv \frac{1}{N}\sum_{i=1}^{N} x_i \approx \mu(x), \tag{1}$$

where x is total annual rainfall,[1] and N is the number of years of data used to calculate $\overline{x}$. Equation (1) implies the assumption that there is a $\mu(x)$, because we are questioning all estimates of $\overline{x}$; in turn, $\mu(x)$ may be regarded as a theoretical climatology, as discussed in appendix A. In other words, we look for $\mu(x)$ being aware that at best we will find only a better approximation to it than $\overline{x}$.

In addition, to see how $\overline{x}$ has changed in time, we also define the cumulative annual means (CAM) time series:

$$y_j \equiv \frac{1}{j}\sum_{i=1}^{j} x_i, \qquad j = 1, 2, \ldots, N. \tag{2}$$

We could, similarly, define a variety of other means, for example the 30-yr mean time series:

$$y_k^{30} \equiv \frac{1}{30}\sum_{i=k}^{k+29} x_i, \qquad k = 1, 2, \ldots, N - 29. \tag{3}$$

The goal of this paper is to verify the reliability of the climatological value $\overline{x}$, and to examine y_j, of course, $y_{j=N} = \overline{x}(N)$. We will attempt to do this by finding an alternate μ, here called μ', and compare it with $\overline{x}$; the difference between μ and μ' is that μ' is an obtainable approximation of μ. It is assumed that this comparison, plus the examination of y_j, will give us insight into the reliability and representativeness of $\overline{x}$ as a climatological value.

The study is complemented with an autoregressive experiment, as described in appendix B.

The data

The climatological rainfall data used are the total annual (July–June) precipitation records from July 1948

[1] Recall that for this wintertime precipitation regime total annual rainfall is measured from July of one year to June of the next year.

Corresponding author address: Dr. Edgar G. Pavia, CICESE, P.O. Box 434844, San Diego, CA 92143-4844.
E-mail: epavia@cicese.mx

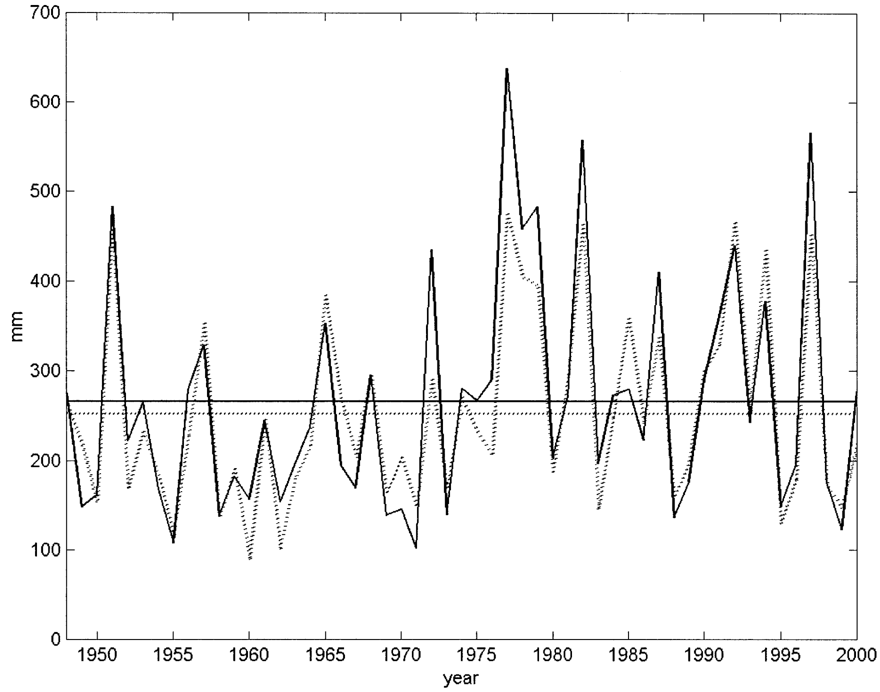

FIG. 1. The 1948–2001 annual precipitations (Jul–Jun) in Ensenada (solid line) and San Diego (dashed line). Horizontal lines indicate entire record averages.

to June 2001 at Ensenada, and San Diego (see Fig. 1). Thus $N = 53$. There are more data available from the National Weather Service and the Mexican water authority (Comisión Nacional del Agua); but this is the longest simultaneous period of reliable instrumental data with no gaps. The two datasets have been recorded independently and may represent differences due to station location and measuring procedures, therefore they are not directly comparable. Nevertheless their high and significant correlation at zero lag (Pavia and Badan 1998) allows us to examine them together.

2. Statistical method

Since obviously as $j \rightarrow N$, $y_j \rightarrow \overline{x}$, we can further assume that for a certain N, here called N', where in general $N' > N$,

$$y_{j=N'} = \overline{x}(N') \equiv \overline{x}' = \mu' \approx \mu.$$

This means that $\overline{x}'$ represents a future value of $\overline{x}$, calculated with more available data and therefore it is assumed to be more reliable than $\overline{x}$. Obviously we cannot directly calculate $\overline{x}'$, but we can estimate μ' if we know the data's probability density function f. For example,

$$\mu'(N, x) = \int x f(N, x)\, dx, \tag{4}$$

where $f(N, x)$ is obtained with the N available data; that is, μ' is the first moment of this particular f. We regard μ' as the mean value of the ''future state'' of the N data histogram as $N \rightarrow N'$, where the residual of f and this histogram is a measure of $\delta N = N' - N$; that is, the number of years needed to reach $\overline{x}'$. However, at this time we do not know, unfortunately, the functional relationship between the residual and δN.

To model the histogram of the positive definite precipitation data we propose a Weibull pdf (Wpdf):

$$f(x; a, b) = \frac{b/a}{(x/a)^{(1-b)}} e^{-(x/a)^b},$$

for which the first moment is

$$\mu' = a\Gamma[1 + (1/b)],$$

where Γ is the gamma function [expressions for higher moments are given in Pavia and O'Brien (1986) and references therein]. Thus the problem of finding a more reliable climatology is reduced to finding the parameters a and b of a Wpdf. We can do this in different ways except the iterative ''method of moments'' because then (1) determines (4). The most common method is a least squares procedure, and for this it is convenient to fit a cumulative distribution function

$$F(x; a, b) \equiv \int f(x; a, b)\, dx = 1 - e^{-(x/a)^b}, \tag{5}$$

to the ogive or cumulative histogram; as, for example, in Pavia and O'Brien (1986).

TABLE 1. The mean climatological value, the mean of the μ's and its standard deviation, and the mean of y_j and its standard deviation (mm).

Station	$\overline{x}$	$\overline{\mu'}$	$\sigma_{\overline{\mu'}}$	$\overline{y}$	$\sigma_{\overline{y}}$
Ensenada	265	249	3	244	3
San Diego	251	244	2	235	2

a. *Estimation of the averaged first moments*

Once it has been decided to use a least squares procedure there are still several choices to be made in order to estimate the Weibull parameters. For example, in this case we choose to linearize (5) and minimize an expression of the form

$$\sum_{k=1}^{n} \epsilon_k^2 = \sum_{k=1}^{n} [(A + BX_k) - Y_k]^2, \qquad (6)$$

where n is the number of classes forming the ogive, $A = \ln(a^{-b})$, $B = b$, $X_k = \ln(x_k)$, and $Y_k = \ln\{-\ln [1 - F(x_k)]\}$. We still have to choose the number of classes n and their form $F(x_k)$. This is an important step because different choices may give different results and still pass any statistical test satisfactorily. Therefore in this work we calculated all μ's for $N/10 < n < N/2$ (the number of μ's is $\nu = 21$, because $N = 53$) and, since the choice of $F(x_k)$ seems to be less critical, we used $F(x_k) = k/n$; $k = 1, 2, \ldots, k = n - 1$. [For other possible choices of $F(x_k)$ see Pavia and O'Brien (1986).] We will average all μ's and obtain the lowest limit of their standard error $\epsilon = \sigma_{\overline{\mu'}}$ to compare $\overline{\mu'} \pm \sigma_{\overline{\mu'}}$ and $\overline{x}$; we call ϵ the lowest limit of the standard error because the μ's are not uncorrelated and $\sigma_{\overline{\mu'}} \equiv \sigma_{\mu'}/(\nu)^{1/2}$, where $\sigma_{\mu'}$ is the standard deviation of the $\nu(=21)$ μ's; ($\sigma_{\overline{\mu'}}$ is also called the "standard deviation of the mean'').

b. *Estimation of the averaged cumulative means*

Similarly, we will average all ys and obtain the lowest limit of their standard error $\varepsilon = \sigma_{\overline{y}}$ to compare $\overline{y} \pm \sigma_{\overline{y}}$ and $\overline{x}$; $\overline{y}$ is, of course, a weighted average with more weight to the earliest values and $\sigma_{\overline{y}} \equiv \sigma_y/(N)^{1/2}$, where σ_y is the standard deviation of the $N(=53)$ ys.

3. Results

The results of the estimations mentioned in the previous section are summarized in Table 1. We can see that, for the two stations considered, both $\overline{\mu'}$ and $\overline{y}$ are statistically lower than $\overline{x}$. These results are somewhat verified by an autoregressive experiment (see appendix B for details), in which the estimated first moments of synthetic series yield mostly lower values than their corresponding global averages.

In Fig. 2 we compare y_j, $j = 1, 2, \ldots, N$, with $\overline{y}$. The most important feature here is the change, from below $\overline{y}$ to above $\overline{y}$, during the second half of the 1970s in both the Ensenada and San Diego stations. This seems to coincide with the climate shift of the Pacific Ocean (Miller et al. 1994), and the beginning of an episode of wetter cool seasons in the southwest United States (Swetnam and Betancourt 1998).

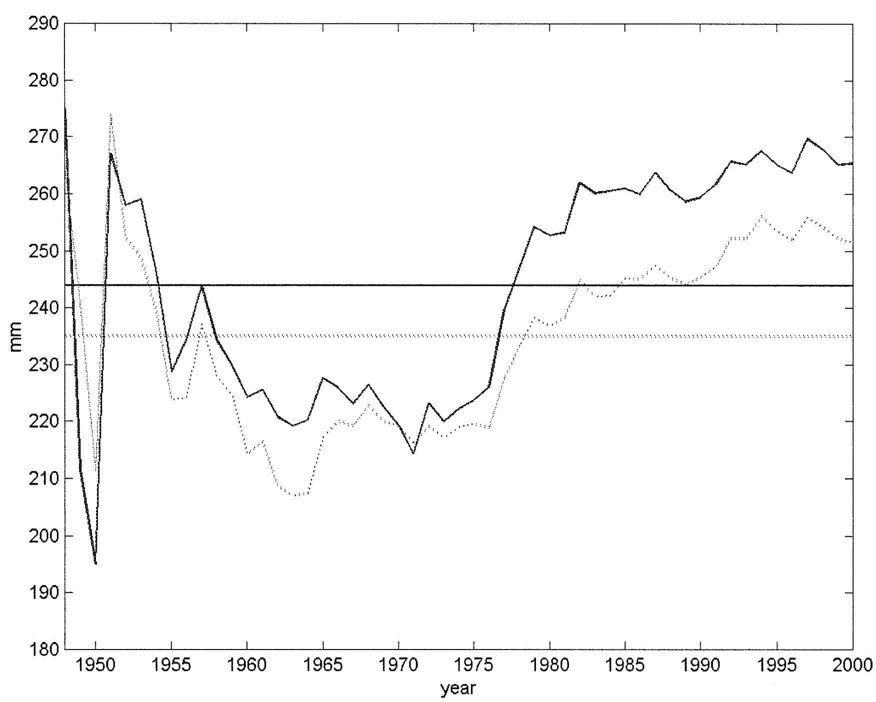

FIG. 2. The cumulative averaged means (CAM) time series and the averaged CAM in Ensenada (solid line) and San Diego (dashed line).

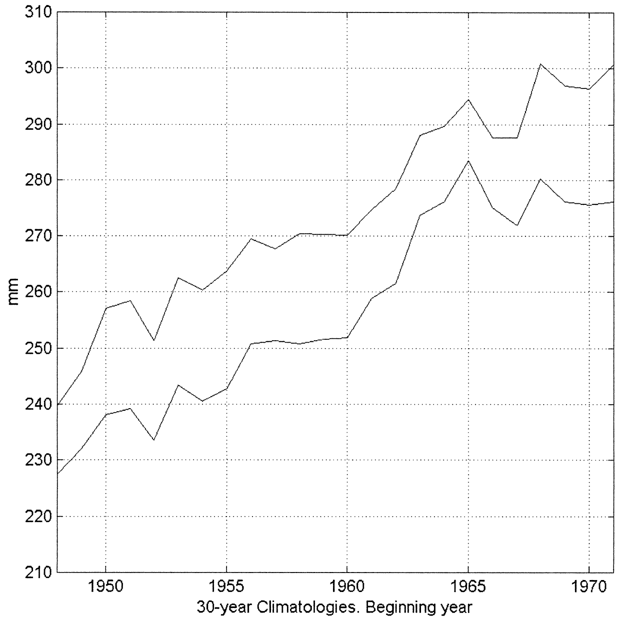

FIG. 3. The 30-yr climatologies time series for both stations Ensenada (higher curve) and San Diego (lower curve).

Time series of 30-yr climatologies from 1948–78 to 1971–2001 calculated with (3) are shown in Fig. 3. We notice that both stations show a clear positive tendency with a maximum value for the 1968–98 Ensenada climatology and for the 1965–95 San Diego climatology. (Recall that here precipitation year goes from July to June, e.g., July 1971–June 2001.)

4. Discussion

The results of the previous section describe what we call the recent Mediterranean California's rainfall climatology. We believe the Weibull analysis suggests that the annual precipitation means (1) are slightly overestimated. In other words, we believe that as N grows ($N \to N'$), $y_j \to \mu'$, because the estimations of μ' are more robust than the y_j time series; that is, if the change of total rainfall from one year to the next is order 1: $\Delta(x_i) \sim O(N^0)$, the change of $\bar{x}$ and y decreases with increasing N: $\Delta(\bar{x}) \approx \Delta(y) \sim O(N^{-1})$; but the change of the estimated alternate ''theoretical mean'' is believed to be smaller: $\Delta(\mu') \sim O(N^{-z})$, where $z > 1$. [The exact value of z is not known, but of course $\Delta(\mu) \equiv 0$, by definition.]

A cursory extrapolation seems to indicate that $y_j \to \overline{\mu'}$ in $\delta N \sim O(10\ \text{yr})$; but regardless of the specific values of $\overline{\mu'}$ and $\overline{y_j}$, between 233 and 252 mm, the important thing here is the suggestion that the present climatological values have a tendency to decrease (265 mm for Ensenada and 251 mm for San Diego). This claim is also supported by the Weibull statistics of first-order autoregressive AR(1) modeled long synthetic series (see appendix B). Incidentally, these analyses seem to suggest that the San Diego climatology might be closer to μ than the Ensenada climatology, and therefore it might be more reliable. We believe that future values of y_j would be smaller than $\bar{x}$, and thus closer to the true climatology μ, which we also believe to be the same for both stations (~246 mm).

Although half a century of annual rainfall data is usually too short a record for most climatological studies, the Weibull analysis presented above seems sound enough to warrant consideration. Indeed, even if the use of Weibull statistics is relatively new for examining climatologies, similar analyses have been used for a long time to characterize datasets (Gumbel 1958) or, more specifically, to redefine probability distributions (Muraleedharan et al. 1999).

This approach has avoided, at the same time, the simplistic problem of linear extrapolation (which would have made us to conclude that the mean value will continue to increase), and the very complicated issue of stationary versus nonstationary processes in climatology (if this climatology is a stationary process its mean value will decrease, if it is a nonstationary process we cannot reach any conclusion). In the end we had to assume that this climatology is ''quasi-stationary,'' in the sense of the dendrochronological data of Swetnam and Betancourt (1998), which suggest that the last quarter of the twentieth century was a one-in-a-thousand-years unusually wet event in the southwest United States.

5. Summary and conclusions

This diagnosis of the climatology of precipitation in the Mediterranean California is part of a first step toward a long-term climatological prediction. In order to proceed we need to find the value of z and the relationship between the least squares residual and δN. Both of these tasks are now in the works and shall be reported in a future contribution.

Finally, the main conclusion of this work seems to indicate that the mean climatological value of precipitation in the Mediterranean California region will decrease over the near future. Since this empirical result is independent of other studies of low-frequency oscillations, such as NPO, which may lead to similar findings, the likelihood of the preliminary scenario suggested by this work may be even greater than it has been mentioned here.

Acknowledgments. We thank Tere Cavazos, Sasha Gershunov, Jim O'Brien, Pepe Ochoa, and Julio Sheinbaum for brief but insightful comments. The opinions and suggestions of two anonymous reviewers were also very helpful and mostly resulted in the two appendixes below. The first author acknowledges the hospitality of Bernard Barnier and Joël Sommeria at the Laboratoire des Ecoulements Géophysiques et Industriels in Grenoble, France, during the reviewing process of this paper. Funding for this work has been received from the Mexican SEP-Conacyt system.

APPENDIX A

The Theoretical Climatology

In general we can define the theoretical climatology of x, $\mu(x)$, as a mean over space and time, for example,

$$\mu(x) \equiv \frac{1}{T}\int_T \frac{1}{A}\int_A x(a, t)\, da\, dt,$$

where $x(a, t)$ represents the space- and time-dependent variable of interest, A an isoclimatological region, and T an isoclimatological period. Of course there are other integral expressions for the mean value, for example, (4), but in practice we cannot know $\mu(x)$ because we do not know $x(a, t)$. Therefore we usually approximate $\mu(x)$ by $\overline{x}$, as in (1), when we know some measurements x_i, $i = 1, 2, \ldots, N$, which we assume to be a representative sample of $x(a, t)$ within the region A and the period T. Obviously if these assumptions are poorly justified $\overline{x}$ will be a poor approximation of $\mu(x)$, and then we can either question the sample x_i as representative of $x(a, t)$ within A, or the sample size N as representative of T. The former would be an unlikely case because usually the spatial distribution of the x_i samples is used to define A. And thus the latter case seems to be our only choice. Nevertheless in this case even T is difficult to define; for example, we cannot just use the sample size because N may overlap two (or more) isoclimatological periods. One possibility is to assume that T is better represented by N' samples, where $N' \gtrsim N$, which has been done in this paper.

APPENDIX B

The Autoregressive Experiment

We modeled the two time series as first-order autoregressive, AR(1), processes from which large sets of long synthetic time series with statistics similar to the original series were randomly generated. Each series of both sets was Weibull fitted and its first moment was estimated. Finally the distributions of the first moments of each set were compared to the global average of their corresponding set.

We run several cases varying the size of the large set ($S = 50, 100, 200$), the length of the synthetic series ($L = 300, 400, 500$), and the number of classes used to fit the Weibull distribution ($n = 10, 15, 20$). With the exception of some cases where $S = 50$ and/or $n = 10$ the results were very similar. The global average for the Ensenada simulation was 278 ± 1 mm, or somewhat higher than the climatological mean of 265 mm; the corresponding value for San Diego was 247 ± 1 mm, or slightly lower than its climatological mean of 251 mm. However the (''quasi Gaussian'') distributions of the estimated first moments were in all cases significantly lower than their global averages. For Ensenada, the distribution was centered around a mean of 253 mm with a standard deviation of 9 mm. For San Diego, the distribution was centered around a mean of 227 mm with a standard deviation of 7 mm.

REFERENCES

Gershunov, A., T. P. Barnett, and D. Cayan, 1999: North Pacific interdecadal oscillation seen as a factor in ENSO-related North American climate anomalies. *Eos, Trans. Amer. Geophys. Union,* **80,** 25–30.

Gumbel, E. J., 1958: *Statistics of Extremes.* Columbia University Press, 302 pp.

Miller, A. J., D. Cayan, T. P. Barnett, and N. E. Graham, 1994: The 1976–77 climate shift of the Pacific Ocean. *Oceanography,* **7,** 21–26.

Muraleedharan, G., N. Unnikrishnan Nair, and P. G. Kurup, 1999: Application of a Weibull model for redefined significant wave height distributions. *Proc. Indian Acad. Sci., Earth, Planet. Sci.,* **108,** 149–154.

Pavia, E. G., 2000: Secondary forecast models—The ENSO example. *J. Appl. Meteor.,* **39,** 1952–1955.

——, and J. J. O'Brien, 1986: Weibull statistics of wind speed over the ocean. *J. Climate Appl. Meteor.,* **25,** 1324–1332.

——, and A. Badan, 1998: ENSO modulation of rainfall in the Mediterranean Californias. *Geophys. Res. Lett.,* **25,** 3855–3858.

Swetnam, T. W., and J. L. Betancourt, 1998: Mesoscale disturbance and ecological response to decadal climatic variability in the American Southwest. *J. Climate,* **11,** 3128–3147.