Specification of Monthly Precipitation in the Western United States from Monthly Mean Circulation

BRYAN C. WEARE AND MARC A. HOESCHELE

Department of Land, Air and Water Resources, University of California, Davis 95616

(Manuscript received 6 December 1982, in final form 2 March 1983)

ABSTRACT

The ability to specify monthly precipitation at 36 stations in Washington, Oregon and California by various indices of monthly mean circulation has been investigated. Most of the circulation indices were derived from eigenvector analyses of 700 mb geopotential height or relative vorticity for 1951–76 or 300 mb relative vorticity for 1963–76. The results indicate that an average up to 60% of the rainfall variance may be specified by same-month average circulation parameters. The 700 mb parameters analyzed within a limited domain in the vicinity of the rainfall network result in the best specification ability. Indices derived from larger domains or from the 300 mb vorticities result in generally poorer specifi cations. Considerable spatial variation in specification ability is apparent with greatest skills evident in the coastal region of northern California and Oregon and lowest skills in the eastern regions of all three states.

1. Introduction

Experimental attempts at forecasting monthly precipitation one or more months in advance have often relied upon forecasting mean circulation patterns and then inferring the associated precipitation from those patterns (Namias, 1980). The circulation forecasts have depended on a combination of methods including analogs, trends, teleconnections and air–sea interactions (Namias, 1978). The inferences of associated rainfall have relied upon (Namias, personal communication, 1982) the earlier statistical studies of Stidd (1954) and Klein (1963). A number of temperature and precipitation forecast schemes have been recently evaluated by Preisendorfer and Molby (1982). They find that the quality of predictors is rather low and differs substantially for different geographic regions and time periods.

Most researchers have attempted to improve the veracity of forecasts by investigating new methods of forecasting the circulation patterns. In a series of papers Namias (1975) has discussed and utilized North Pacific sea surface temperatures in this regard. More recently Horel and Wallace (1981), Hoskins and Karoly (1981) and others have proposed a link between tropical Pacific surface temperatures and western Northern Hemisphere circulation.

Relatively little recent work has been carried out on understanding and improving the quality of the inferences of monthly rainfall from current monthly mean circulation indices (Harnack, 1979). This seems useful at this time because ideas concerning methods of specification have been altered by the work of Davis (1976, 1978), Barnett and Preisendorfer (1978), Barnett and Hasselman (1979), and others. They have emphasized the importance of an a priori reduction of predictor or specifier variables in making good statistical predictions or specifications. The earlier correlation studies of Stidd (1954) and Klein (1963) preceded these developments. Furthermore, reexamination of the specification problem at this time seems worthwhile because circulation data are now available for many more years at a number of levels and for a number of parameters. Finally, the earlier studies had attempted to specify average precipitation over state-size regions, whereas precipitation in smaller regions important to agriculture or water management problems is perhaps more relevant.

Therefore we have carried out a pilot study exploring our ability to specify monthly precipitation for five “winter” months at 36 stations in Washington, Oregon and California using indices of circulation variables for the same months. It should be emphasized that the following comparisons are all of the ability to specify rather than forecast monthly precipitation. The circulation indices used in the specifications are in most cases the time coefficients of the three most important eigenvectors (Kutzbach, 1967) of monthly mean circulation variables. The eigenvector analysis provides an a priori method of reducing the large number of potential predictors (i.e., grid values over the Northern Hemisphere) to a very few predictors to be used in the statistically derived specification equations.

The series of “experiments” to be described compare the ability of various circulation indices to specify precipitation relative to the ability of a reference index. In each case the differences in specification
ability in both space and time are compared to the reference experiment. In this way it is hoped that a better understanding may be gained of the limits of our ability to specify monthly precipitation with monthly mean circulation variables.

2. Data

There are a large number of potential atmospheric circulation variables at many levels that might be of use in specifying rainfall. We have chosen three. The first is 700 mb geopotential height, because this is the variable that has been traditionally used in long-range forecasting and because earlier studies indicate that it is moderately linked to precipitation amounts. As a second variable we have chosen 300 mb relative vorticity. This was chosen mainly because modeling studies (e.g., Hoskins and Karoly, 1981) have used it to diagnose the middle latitude response to tropical forcing. Finally, we chose to use 700 mb relative vorticity in order to at least partially differentiate whether differences in specification ability between the 760 and 300 mb variables are primarily due to the differences in elevation or differences in parameter.

Monthly mean 700 mb geopotential heights ($Z_{700}$) at $5^\circ$ latitude by $10^\circ$ longitude intersections, north of $20^\circ$N, for the period December 1946–March 1976 were kindly supplied by Professor Jerome Namias. Since our purpose is to better understand the factors that affect specification, rather than develop the “best” specification relation, we choose not to update this data set. From this same data we calculated finite difference estimates of the monthly relative vorticity ($\zeta_{700}$) at the same intersection points (excepting the northern and southernmost perimeters). Monthly mean 300 mb relative vorticity ($\zeta_{300}$) were derived at $5^\circ \times 5^\circ$ intersections for April 1963 through December 1979 (excepting January and February 1970) from the National Meteorological Center analyses of mean 300 mb wind components which were obtained from the National Center for Atmospheric Research. For all three data sets long-term monthly means were calculated and removed from the values used for further study.

Rainfall data for the “cooperative” stations in Washington, Oregon and California for 1951–80 were obtained from the National Climatic Center. These data, which are used to derive climatological normals, are serially complete, since the few missing values were estimated from the records of nearby stations. From the stations available we chose a group of 36 stations which had as few estimated values as possible and represented as well as possible most of the climatic and topographic regions in these three states. The locations of these stations are shown in Fig. 1.

The circulation variables were subjected to eigenvector analysis (Kutzbach, 1967) in order to define a few indices which describe most of the variability. It seemed to be an open question as to how large a spatial domain one should use in the eigenvector analysis. If one chose too small a domain, one might exclude a region which had strong associations with the rainfall in question. If, on the other hand, too large a region were chosen, the dominant eigenvectors, explaining less of the overall variance, may largely be describing the variability of a region unimportant to the specification. To explore this problem we chose to look at three successively larger domains. The twelve points of the smallest domain, for which most of the studies were carried out, are indicated by plusses on Fig. 1. The other regions, whose data were subjected to the eigenvector analyses, were simply defined by expanding the longitudinal extent in both directions until 24 or 48 $5^\circ \times 10^\circ$ grid points were included. The time coefficients of the eigenvectors of the data sets will be designated by their type, level and number of grid points so that, for instance, $Z_{700}(12)$ refers to the time coefficients of the eigenvector analysis of the 700 mb height at the 12 locations illustrated in Fig. 1.

Since the significance of the derived specifications relations depends strongly on the number of predictors, most of the following examples use only the time coefficients of the three most important eigenvectors of the circulation variables to specify the rainfall. This
is done despite the fact there are moderate differences in the cumulative variance explained for the different variables over different domains. For instance, the first three eigenvectors of \( Z_{700}(12) \) explain a cumulative percent of the variance at those 12 points of between 94.2 and 92.6 for different calendar months from November through March. The first three eigenvectors of \( Z_{700}(24), Z_{700}(48), Z_{700}(12) \) and \( Z_{700}(12) \) explain together approximately 84, 72, 85 and 86% of the variance, respectively.

3. Results

All of the specification experiments which will be discussed were carried out in a similar manner. In general the first experiment using 700 mb heights was considered the reference to which the other experiments were compared. In this case, as with the other experiments, linear multiple regression equations for each rainfall station for each month from November through March were derived from the available circulation and rainfall data. These months were chosen because they represent the bulk of the “rainy” season over much of the region being considered. The independent variables in this reference experiment are the time coefficients of the three most important eigenvectors in the analysis of the 700 mb heights at the 12 points shown in Fig. 1 for the period 1951–76. Thus for a month \( m \), year \( y \) and station \( i \), precipitation \( P_{ymi} \) may be specified using the 700 mb heights for the same month by

\[
P_{ymi} = \alpha_{0mi} + \alpha_{1mi}Z(12)_{1my} + \alpha_{2mi}Z(12)_{2my} + \alpha_{3mi}Z(12)_{3my} + E_{my},
\]

where the \( \alpha \)'s are regression coefficients derived from the entire 1951–76 data set (Weisberg, 1980), the \( Z(12) \)'s are the time coefficients of the eigenvector analysis, and the \( E \)'s are the errors resulting from imperfect specification.

Fig. 2 shows the fraction of the monthly rainfall variance specified by the three time coefficients of the 700 mb height averaged over the five calendar months for 1951–76. Table 1 shows the fraction of explained variance averaged over all stations for each calendar month and the five-month average. The spatial patterns of the specification maps for the individual months is quite similar to the average map in Fig. 2. This figure suggests the best specification is possible along the coast, especially from central California through Oregon. The poorest specifications are along the ridge of the Sierra Nevadas in California and in the high, relatively dry, plains of Washington and Oregon. This general pattern is probably indicative of the fact that most coastal rainfall is due to frontal systems which are enhanced by orographic uplift, whereas the rainfall in the eastern areas is partially influenced by convective activity. Furthermore, the high mountains of much of the eastern region have considerable amounts of snow, which is probably not measured with the accuracy of rainfall.

This reference experiment was modified using the time coefficients of two or four eigenvectors. The results in both cases were quite similar, except that the two-parameter case resulted in slightly lower explained variance.

The input data of the reference experiment was altered in two other ways. In the first case the time coefficients were derived using the 700 mb data at the 24 points; in the second the coefficients correspond to the 48-point eigenvector analyses. Table 1 also indicates the average results for these two cases. Overall, the use of the 24-point data set results in a slight deterioration of specification. The use of the 48-point set results in a substantial reduction in explained variance for all months but January. Fig. 3 illustrates the average fraction of explained variance for all five months (corresponding to Fig. 2) when the 48-point data set is used. Comparisons of Figs. 2 and 3 suggest that the use of the 48-point set results in a deterioration of specification ability at all stations with the
greatest losses of up to 27% in western Washington. This strongly suggests that the size of the spatial domain of the specifier field is very important.

The reference experiment was again altered. This time the first three time coefficients of the 700 mb relative vorticity at the 12 points shown in Fig. 1 were used to specify the same month's precipitation for each of the five calendar months. Fig. 4 indicates the average fraction of the total precipitation variance at each station for the same 1951–76 period for all five months. Table 1 indicates the average over all stations for each calendar month. As may be seen from Table 1 the overall ability of the relative vorticity to specify the precipitation is quite comparable to the geopotential height. Comparison of Figs. 2 and 4, however, indicate that the pattern of the ability to specify differ somewhat for the two data sets. In western Washington and Oregon the height data explains from 4–14% more of the precipitation variance than do the vorticities. In the remainder of the region, however, differences between the two results are quite small with the vorticities being slightly superior.

An additional modification of the reference experiment was to substitute 300 mb relative vorticity data for the 700 mb height data. Thus the time coefficients of the three most important eigenvectors for the vorticity at the 12 points shown in Fig. 1 for the period 1963–76 were used to specify the precipitation. Fig. 5 illustrates the spatial pattern of the variance explained, averaged over all five months. Table 2 shows...
The possible exceptions are the southeastern deserts of California and parts of eastern Oregon.

A comparison of the results illustrated in Fig. 6 and those of Fig. 2 illustrates some of the difficulties in making comparisons and, indeed, in generating statistically reliable regression equations relating circulation indices to precipitation amounts. These figures suggest that much more effective formulas can be generated with the 1963–76 data to specify the 1963–76 precipitation than can be calculated with the 1951–76 data to specify the 1951–76 precipitation. This implies that a closer linear link existed between the heights and precipitation in the latter period than the former or that the nature of that relation changed sometime during the period 1951–76. The implication is that on new data the regression equations, which have been generated, may explain a far smaller fraction of the variance than expected (Davis, 1976). This seems to be at least partially true. When we calculated the fraction of the variance explained by the 300 mb data through the year 1979 (three years beyond the 700 mb data record), the results were somewhat poorer in all months and much poorer in March.

As a final set of experiments we wished to see how another a priori method of defining the circulation variables to be included in regression equations similar to Eq. (1) might compare to the results shown in Figs. 2–6. Horel and Wallace (1981) have shown that monthly mean 700 mb geopotential heights over various regions of the Northern Hemisphere are moderately correlated with several indices of the Southern Oscillation for one or more preceding months. Given that this implies a potential method of forecasting, it seemed useful to attempt to relate 700 mb height for those regions most closely associated with the Southern Oscillation with the precipitation at the stations shown in Fig. 1. From Horel and Wallace's Fig. 9, we have chosen six regions in the Northern Hemisphere. These are shown in Fig. 7. The 700 mb geopotential height data for these six points for 1951–76 were used to generate multiple linear regression equations to specify the precipitation at the stations in Fig. 1. The fractions of the explained variance at each station averaged over the five months of November through March are shown in Fig. 8. The averages of all the stations for each month are between 0.35 in March and 0.46 in December. In general, the heights at these six locations explain less variance than the

<table>
<thead>
<tr>
<th>Variable</th>
<th>Comment</th>
<th>Nov</th>
<th>Dec</th>
<th>Jan</th>
<th>Feb</th>
<th>Mar</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Z₁₀(12)</td>
<td>As reference except 1963–77 only</td>
<td>0.64</td>
<td>0.53</td>
<td>0.52</td>
<td>0.55</td>
<td>0.55</td>
<td>0.56</td>
</tr>
<tr>
<td>Z₃₀(12)</td>
<td>1963–77 only</td>
<td>0.47</td>
<td>0.50</td>
<td>0.43</td>
<td>0.39</td>
<td>0.46</td>
<td>0.45</td>
</tr>
</tbody>
</table>
First three eigenvectors of 700 mb height at the 12 points of the reference experiment. A comparison of Figs. 2 and 8 suggests that the heights at these locations are less useful except in eastern Oregon and northeastern California. In addition, since these specification models use six independent variables compared to those previously discussed with three independent variables, they should be statistically less reliable when applied to new data (Davis, 1976).

4. Discussion

The purpose of this research was to better understand the factors which control the ability to specify monthly precipitation from monthly mean circulation parameters. The data in Tables 1 and 2 suggest that an average of between ~35 and 60% of the precipitation variance may be specified by three 700 or 300 mb circulation variables. In general, the ability to specify coastal stations is appreciably greater than that for stations in the interior mountains or high plains. This points out the possible danger of specifying precipitation for a course grid or large area averages and assuming that those specifications are equally reliable over the entire region. At 700 mb the abilities of the geopotential heights or relative vorticities seem comparable. On the other hand the 300 mb vorticity seems less useful as a precipitation specifier. Finally, the 700 mb height data at the six points suggested by the work of Horel and Wallace seem to do a poorer job of specifying precipitation than the time coefficients defined by the 12 or 24 point eigenvector analysis.

Since we have chosen to follow a single methodology in most of our experiments (using the time coefficients of the three most important eigenvectors as the dependent data in the multiple regression equations), we have not fully explored the optimization of the rainfall specification even for the chosen variables. For instance, one might circumvent the eigenvector analyses and calculate the regression coefficients for circulation variables chosen at sites dictated by a “screening” analysis (Klein, 1981). We have chosen not to do so because we believe, following the work of Davis (1976) and others, that the present method is likely to give more statistically significant...
results even though the fraction of explained precipitation variance for dependent data may be somewhat higher from the screening analyses. The question of significance is especially important for the proposed use of any derived specification relations, i.e., interpretation of month ahead forecasts of monthly mean circulation parameters. In such cases one would hope for a correct overall pattern of the circulation but expect relatively large errors at individual points. In this case it would seem likely that the use of the eigenvector analysis, which provides coefficients based upon weightings of the entire circulation field, would be more reliable than using "screening" equations which focus on specific locations which may have considerable errors. A reviewer has suggested the possibility of combining the two methodologies by performing a "screening" analysis using a pool of all significant eigenvector time coefficients.

Although we have by no means exhausted the number of possible circulation specifier variables, the comparison of the results of $Z_{700}$ with those of $Z_{700}$ suggest that different variables are unlikely to give substantially improved results. It is possible, however, that apparently improved results might be derived if one chose to specify an alternate precipitation vari-

able, for instance the square root of the monthly precipitation, which is more normally distributed.

Davis (1976) has stressed that "forecast" skills derived from statistical specification equations are likely to be far less than "hindcast" skills. A preliminary analysis using 700 mb height data for 1977-80 has been used to explore this. The average specification skill for all stations and all five months using the regression equations derived from the $Z_{700}(12)$ data is about 0.32 for the new data compared to the hindcast skill of 0.48 shown in Table 1. The average skill for the regression equations for $Z_{700}(48)$ is 0.05 for 1977-80 compared to 0.38 shown in Table 1. These results not only illustrate the decline in "forecast" versus "hindcast" skills but also tend to verify our conclusion that the regression equations based upon the 12-point eigenvector analyses are superior to those based upon the 48-point analyses.

Finally our analysis suggests a result which may be important to water management decisions in the western United States. Many reservoirs in this region derive their water from the mountainous regions of eastern California or central Washington. Fig. 2 suggests that the 700 mb geopotential heights are relatively poor specifiers of the precipitation in these regions compared with the coastal regions. This would imply that even if accurate month ahead forecasts of the 700 mb height were possible, there may be severe limitations in inferring predicted precipitation from those heights. On the other hand, Fig. 4 suggests that in the above mentioned region the 700 mb relative vorticity may be a better specifier. Thus the variable of choice may be a function of the requirements of the forecast.

Acknowledgments. This work was supported by Grant ATM-8241336 from the National Science Foundation and funds from the Department of Agriculture.

REFERENCES


