Composite Predictor Maps of Extraordinary Weather Events in the Sacramento California Region

By

Richard Grotjahn

Dept. of Land, Air, and Water Resources,

University of California, Davis, CA 95616, U.S.A.

Ghislain Faure

Cellule Recherche Cyclone, Météo-France

B.P.4, 97491 Ste Clotilde Cedex, France

Submitted June 2006

Revised 14 June 2007

Corresponding author address: Richard Grotjahn, Dept. of Land, Air, and Water Resources, One Shields Avenue, University of California, Davis CA 95616

Email: grotjahn@ucdavis.edu

Abstract

Extraordinary weather events in the Sacramento California, USA region are examined using a simple compositing technique. The extraordinary events identified are uncommon and the worst of their kind, but not necessarily severe. While our criteria are drawn from Sacramento weather station data, the identified events are extraordinary elsewhere over much, if not all, of California's Central Valley. Several types of extraordinary events are highlighted, including the hardest freezes, heaviest prolonged rain events, longest duration fog, and worst heat waves (onset and end) in a 21 year period. Bootstrap resampling establishes the statistical significance of features on the composite maps. The composite maps with statistically significant features highlighted allow a forecaster to search for key features in forecast maps that co-exist with or that precede an extraordinary weather event. Local and regional-scale extraordinary events have larger scale signatures that can be traced back in time. Many of these features are intuitive and known to local forecasters (and that provides a check upon our methodology). However some features appear to be unexpected. For example, a ridge (in height and thermal fields) over the southeastern USA generally occurs prior to the worst heat waves and hardest freezes. Some features appear to exhibit the theoretical concept of downstream development. Several extraordinary weather types are preceded by a ridge either over Alaska (hardest freezes and heaviest prolonged rain) or just west of Alaska (worst heat waves). While the Alaskan ridge passes a significance test, the presence of other features (such as the southeastern ridge) determine what, if any extraordinary event occurs near Sacramento. However, a feature that passes the significance test for the composite might

not occur in *every* member of a given extraordinary event. The height and thermal patterns over the West Coast and north Pacific are similar for summer's worst heat waves and winter's longest duration fog: both types of events are preceded by a trough in the eastern mid Pacific.

1. Introduction

This study focuses upon diagnosing past extraordinary weather events in the southern Sacramento Valley of the state of California. The article has two purposes. The first purpose is to summarize the techniques used to identify key elements on standard weather maps so that a forecaster can perform equivalent analyses for a different forecast region. The second purpose is to show illustrative examples of key elements found on standard weather maps before and during four types of extraordinary events affecting broader areas that overlap in the southern Sacramento Valley. Such key features are useful and important for forecasters to learn; for example Grumm and Hart (2001) discuss the value to forecasters of recognizing such patterns for unusual events, since they are typically missed by a model or statistical analyses of model output. These large scale patterns include features that are intuitive and well known to local forecasters. (If such features were not recovered by the scheme the approach would be called into question.) However, there can be other significant features that do not appear to be commonly recognized by local forecasters and as such represent value added by the technique.

The region studied straddles 38^{o} N latitude. The inland valley has a Mediterranean climate characterized by hot dry summers and cool damp winters. The climatology of the region is summarized in Bevan and Cline (2005). Though the climate is generally mild, there are several types of weather that may be considered extraordinary. The extraordinary weather events studied here include the hardest freezes (which combine lower temperatures with duration below freezing) that can damage agricultural and ornamental plants as well as public utilities (e.g. freezing pipes). Other extraordinary weather events studied here are: the hottest heat waves, the longest duration fog, and the heaviest prolonged rain. While our focus is on the southern Sacramento Valley, most of these extraordinary events encompass a much larger region (the exception being the ends of the hottest heat waves which are localized to the southern Sacramento and northern San Joaquin valleys. The broader area affected varies with the type of event. For example, the hottest heat waves affect the entire Central Valley and coastal regions of central and northern California. On the other hand, sunny days near the coast occur during long duration fog episodes of the Central Valley.

Before proceeding, we need to explain why we selected the name 'extraordinary' to highlight the uncommon nature of the events we study. Other candidate names might be 'uncommon', 'unusual', 'severe', or 'significant'. Events when temperatures exceed 40 C are not 'uncommon' in the Central Valley. But there are unusual heat waves that have significant economic effects and those unusual events are our focus. While the events we study are 'unusual' for their type, 'extraordinary' conveys the meaning that that *type* of event is not unusual just the *properties* of that event. While two of our extraordinary

event types: hottest heat waves and heaviest prolonged rain events might be considered ‘severe’ for this region, another event we isolate (longest duration fog) would not be considered ‘severe’. Also, some readers may prefer to restrict usage of the term ‘severe’ to officially sanctioned terminology such as: ‘a severe thunderstorm’. Finally, while these events might be labeled ‘significant’ we avoid any confusion caused by multiple uses of the word and reserve the word significant for statistical use, specifically for identifying key regions on the charts that are significant to a certain level of confidence.

A simple technique for forecasting the weather is to search the historical record for patterns on weather maps that correspond with current forecast maps. For unremarkable weather situations, this method may have limited utility. For example, Lorenz (1969) showed that 500 hPa weather patterns rarely repeat except over a limited domain.

However, some extraordinary weather events have well defined weather patterns that can be recognized by a weather forecaster. For example, strong north wind events in the southern Sacramento valley of California are invariably associated with these three factors: 1.) strong mid and upper tropospheric northerlies (i.e. a jet stream and sometimes a jet streak oriented meridionally over the region), 2.) recent passage of an upper level short wave (the negative vorticity advection in the middle troposphere associated with downward motion of the strong winds), and 3.) sea level pressure (SLP) gradient directed northward (higher SLP to the north, lower SLP to the south of the region) that reinforces the surface northerlies. Other factors, principally topography and strength of the

nocturnal boundary layer influence the strength of the winds measured at the surface. Pauley, et al. (1996) describe an example event.

Staudenmaier[1] created composite maps during eight different types of extraordinary Sacramento weather events in an effort to quantify the weather patterns associated with each event type. Staudenmaier averaged geopotential height (850, 700 and 500 hPa), SLP, and a few other fields on the dates of occurrence of each event type. Thus, Staudenmaier obtained a composite map for each event and each variable. The geographical area of each map encompassed much of North America (except Alaska) and a small part of the adjacent oceans. The present study was inspired by Staudenmaier's report but our study is in no way dependant upon that report.

The remainder of the paper is organized around the two stated purposes. The next section describes the statistical methodology and testing of the procedures. For a forecaster to apply this technique elsewhere, additional details and considerations are presented in the appendix. The third section applies the techniques to the four types of extraordinary events listed above.

[1] Staudenmaier, M., 1995: Composite Maps of Meteorologically Significant Events in Sacramento. 22 pp Being an internal publication of the National Weather Service Forecast Office, Sacramento, this unrefereed publication is not widely available nor appropriate for inclusion in the reference list. Its citation here is intended as a courtesy to that author.

2. Methodology

a. Event criteria

This study considers four types of Sacramento weather events: hardest freezes, heaviest prolonged rain, longest duration fog and worst heat waves. These events illustrate different types of challenges as well as sample different seasons. For example, the heavier prolonged rainfall usually occurs over a much shorter period than a notably long fog event making the timing of the event onset relative to upper air observing times more challenging. Each extraordinary event must have a precise or at least quantitative definition, in order to assemble similar situations without bias. Moreover, precise dates of occurrences are needed in case a given variable must be stratified by time of day.

One concern in forming composites is the imprecise starting times of a given event. Consecutive twice-a-day upper air charts can have different lower tropospheric properties due to diurnal variations. To address this concern, additional information was gathered for subjective review of the event identified by the objective criteria. Thus, for each occurrence of a heat wave, the start and the end dates, the day of the maximal temperature, the maximal temperature and the average of the maximal temperature during this period were stored.

The criteria for identifying each type of extraordinary event are summarized in Table 1. These criteria have not been defined arbitrarily, but in accordance with the experience of forecasters at the Sacramento NWSFO and with the climatological records during the 21-

year period of the study. Accordingly, most of our definitions are similar to those in Staudenmaier's study. Selecting the hardest freeze events was difficult to express with a precise definition and our criteria differ from the ones used by Staudenmaier. Our definition includes a limit upon the subsequent maximum temperature. This limit is intended to avoid situations where the freezing temperatures are mainly due to radiational cooling and select for situations where strong cold air advection is taking place ('advective freezes'). Our reasoning is that conditions favoring nocturnal radiational cooling also favor radiational heating during the daylight hours leading to maximum temperatures above 10 ^{o}C. Consequently the duration of very cold temperatures for primarily radiational cooling is short. Duration below a critical temperature threshold could be an important secondary effect besides the peak minimum reached for most cold-sensitive crops like citrus (Attaway, 1997). That threshold temperature varies with genus, species, and variety. Obviously, radiational cooling magnifies the hardest freeze events we select. Another reason to focus upon advective freezes is that many freeze protection measures farmers adopt work much better for primarily radiational than for advective freezes. By excluding events dominated by radiational cooling, we accomplish two things. First, we isolate a more meteorologically homogeneous target sample. Second, we isolate the truly damaging events.

The criteria were applied to data from one or both of two stations: Executive Airport (a station located in suburban south Sacramento) and the downtown Sacramento station. In the case of hardest freezes, only the Executive Airport station data were used. To qualify as a heat wave event both Executive and Downtown stations must pass the criteria. For

heavy prolonged rain cases, dates were selected by a multi-step process. First, candidate periods were identified from highest precipitation days and some consecutive rainy day periods of different lengths and different activity. Heavy prolonged rain occurrences are the only ones not defined as a 'several day' period, but as a 'several hour' period. While for the other types of events all the first day data are gathered, at 0 Z then at 12 Z, in this case it is the data corresponding to the hour of beginning that must be gathered. Thus the exact hour of the start is very useful to minimize the offset between each occurrence. To illustrate: an occurrence that started at 3 p.m. LT was represented by fields at the following 0 Z time, whereas an occurrence that started at 9 a.m. LT was represented by the preceding 12 Z fields. The long duration fog dates were found by a tedious manual search of paper archives. When dense fog is reported in one hour, the date is flagged as a foggy day. Fortunately, not all months need be searched; the inland climate means that long duration fog episodes occur from November through February.

An alternative approach bases the identification of events on how anomalous the large scale fields are from their seasonal means. Hart and Grumm (2001) rank anomalous events (based on a normalized anomaly of four standard deviations or more) but a mixture of events is captured and their analysis focuses on identification, frequency, etc. rather than the conditions of specific event types. Stuart and Grumm (2006) identify major East Coast snow storms using normalized anomalies of winds. Such a technique has advantages (guaranteeing selection of an event affecting a broad region) and disadvantages (false positives and misses of events affecting a region of interest; questionable statistical assumptions, see below). Also, before one can identify what

anomalies in the large scale flow are relevant, one needs to identify the events from local information anyway (if only to verify that the technique is working). For our study, it is simpler and more accurate to derive the key anomalous features from the events and not the other way around. A drawback to our approach is that a significant feature identified in the composite pattern may be present in most, but not all of the events of that type, thereby limiting the predictive power of that feature. This drawback is discussed further in the conclusions.

b. Data

Paper and digital archives of station data are available from the Sacramento NWSFO. To devise the composite maps, NCEP/NCAR Reanalysis 1 (NNR1) data are used. These data are on a 2.5 degree latitude-longitude grid and are described in Kalnay et al. (1996). The domain we choose extends across the north Pacific. The upper air reanalysis data across the north Pacific are strongly influenced by satellite and aircraft measurements. Others (e.g. Kalnay et al., 1996) have shown that the climatology of the NNR1 data differs before and after the onset of routine satellite data in late 1978. So, this study uses the approximately 21 year period from December 1978-1999.

NNR1 data are provided in three classes based on the perceived reliability of the fields (Kalnay et al., 1996). Classification A is assigned to the most reliable variables; these variables are strongly influenced by the observations (e.g., pressure level temperature). Class B designates variables that are still influenced by some observational data, but the model also has a great influence on their value (e.g., pressure level humidity). The C

class is applied to variables (like precipitation rate) that are not directly observed but derived from the model fields; such variables are less reliable than 'B' variables. All the variables used here are drawn from classes A and B, in order to obtain more reliable results. The final choice of the variables was made in cooperation with Sacramento NWSFO staff, in order to target fields commonly used in daily forecasting. Although, some other fields less widely used were included in order to have a more general view of the situation (See Table 2).

A subset of NNR1 data are used. Our domain extends from the equator to 70^{o} N and from 140^{o} E to 290^{o} E. Only four of the 17 NNR1 pressure levels are used in our analysis (850, 700, 500 and 300 hPa) to sample the different troposphere layers. Tropopause and surface data are available separately. The frequency of the data has been reduced from four times a day to twice daily by storing only the 0 Z and 12 Z data. This choice was dictated by the primary radiosonde observing times.

c. General statistical considerations

Composite maps for each type of extraordinary weather event are constructed from a 'target group' formed by the occurrences of the event. The average and the dispersion estimates chosen for this study are respectively the mean and the standard deviation. The overall challenge is to determine objectively how the meteorological patterns are similar through all the occurrences, which leads to an identification of which areas of these composite maps are significant. The statistical procedures to accomplish these tasks are described below.

The null hypothesis to test underlying statistical significance is simply that the observed target group "test" statistic is exactly the same as the population value from which it is drawn. This leads to the construction of confidence intervals. The null distribution, i.e. the sampling distribution of the test statistic under the null hypothesis, is obtained in a way that is specific to the method chosen to assess the significance. While using a parametric test, such as the "Student's t test", the null distribution is a theoretical one, whereas it is an empirical one if using a resampling method. Finally the observed value of the test statistic is compared to the null distribution, to know if the null hypothesis can be rejected or not.

The method used to assess the significance could be the Student's t test. The Student's t test may seem appropriate: it examines the null hypothesis that an observed sample mean has been drawn from a population characterized by some previously specified mean. From this the variance significance can be tested too. This test is a parametric one, with the following basic hypothesis: the sample must be sufficiently large that its sampling distribution is Gaussian. The main disadvantage is that only mean and variance can be used as test statistics, thus the use of another test statistic implies the implementation of a new method. That is why this parametric approach has been rejected in favor of a more general method with better evolution possibilities: the bootstrap.

The bootstrap has been used in several climate studies in the past few years (Gershunov and Barnett,1998a,b; Matthews and Kiladis, 1999). The bootstrap method is a

resampling test, a computer-based method for assigning measures of accuracy to statistical estimates (Efron and Tibshirani, 1993). The idea behind the bootstrap is very simple, and goes back at least two centuries; however it has only been recently developed because it requires modern computer power.

The basic idea is to build up a collection of artificial data groups of the same size as the target group. In this study, the target group is a group of dates corresponding to the occurrences of a type of event, and the artificial data groups are groups of the same number of dates, but these ones are randomly picked in the dataset. Then the test statistic of interest is computed for each artificial group. This procedure results in as many artificial values of the test statistics as there are artificially generated data groups. Taken together, these values constitute an estimated null distribution against which the test statistics computed from the target group are compared.

A non-parametric resampling test has two major advantages. The first is that no assumptions regarding an underlying theoretical distribution for the data are necessary. For example, the method does not assume a Gaussian distribution and such an assumption might be unreasonable. The second is that *any* statistic, so long as it can be computed from the data, can form the basis of the test. Consequently, the use of a new test statistic requires just the implementation of the test algorithm, not of a whole new method. For example, we do not need the members of the target group to have a Gaussian distribution.

d. Basic bootstrap algorithm

The bootstrap method in the significance-testing program is briefly outlined here with further details in the appendix. The initial parameters are discussed first, because they influence many aspects of the program behavior. Then the general algorithm is presented, and its two main parts are described more deeply: the elaboration of the null distribution by resampling and the comparison between the occurrence test value and the null distribution.

A target group of N dates is specified; it is the target group whose significance is assessed. Typically, several groups of N dates are used. The target group marks the beginning of each instance of the extraordinary event. The other groups are defined relative to the target group and track the evolution of the pattern prior to the event (for instance, a group of N dates corresponding to 24h before the event and another N dates 48h before the event). The study of each extraordinary event is performed with seasonal stratification. Each extraordinary event happens only during a three or four month period, and so the use of the other months to assess the significance would bias the results. The total dataset used for the statistical significance has length M and includes all the relevant months (consecutive or not) for all the consecutive years being examined. Also, this dataset must contain all the event-related occurrences dates, even those prior to the event. There are two different ways to perform the bootstrap test: with or without replacement. In the case of bootstrap testing with replacement, the dates among each *randomly* drawn group are not necessarily all different, while without replacement all dates among a randomly drawn sample must be different.

The different steps to create a null distribution using resampling are: 1. generate random numbers; 2. create random samples from them; 3. perform a statistical test on them.

To analyze N occurrences of an extraordinary event, one generates P random groups each of size N. (Typically P=1000 and N is about a dozen.) The procedure to generate the random groups simply calls a random number generator and associates a subrange of the random numbers uniquely with an available date. All subranges have the same size, thereby making the selection of each date theoretically equally-likely. Using the bootstrap with replacement generates P random lists of size N', while using the bootstrap without replacement requires P random lists each of size N, where N'>N and N' is large enough to contain at least N different dates. In our experience, N' = 20 N is sufficient to find N different dates.

Next, a grid point test statistic is performed on each random group and the P results of this test create the null distribution for each grid point. The statistical test performed on the target group is compared to this null distribution. An example is illustrated in Fig. 1. The statistical test is the mean for each group; the plot shows one target group and 1000 (= P) random groups. The data are binned into 0.25 ^{o}C intervals. Significance is determined by where the statistic for a grid point compares to the distribution as follows. Significance at the α % level means that the statistic for the grid point lies either in the highest Pα/200 values (right tail of the distribution) or in the lowest Pα/200 values (left tail). For plotting purposes, significant values in the right tail have positive value, while

those in the left tail are assigned a negative value. In Fig. 1, $\alpha = 1\%$ and dashed lines denote the threshold for the upper and lower ½ % of the values. Between the two dashed lines the null hypothesis is accepted and such values if they occur in the target group would not be considered significant. The figure is for a grid point that one expects to be highly significant by design. The target sample statistic (with heart symbol above it) lies in the right tail and is clearly significant at the 1% level.

The process is repeated at every grid point. A map can thereby be constructed showing any regions of positive and negative significance. Overlaying this significance map onto the original statistic identifies parts of the statistic that are key to identifying the given extraordinary weather event. In addition to assessing the significance of the mean, one can test the standard deviation, forming a standard deviation map identifying areas where the mean pattern has a significant, unusually high or low variability. This section outlines the basics of the bootstrap method sufficient for readers to interpret the selected results. However, readers interested in applying the bootstrap procedure must consider additional enhancements that improve the robustness of the results. Those enhancements are summarized in the appendix.

3. Selected results

Only selected results that highlight some interesting properties or characteristic forecast patterns are shown here. Interested readers can find all the maps having any significant

areas (at $\alpha = 1\%$ level) from all the variables tested (c.f. Table 2) at all levels up to 96 hours before the event at this website: http://atm.ucdavis.edu/~grotjahn/#analogs.

a. Hardest freezes

Eleven hardest freeze events were identified from December 1978 through December 1999; eight start in December, two in January, and one in February. Minimum temperatures range from 18 to 29 ^{o}F (-8 to -2 ^{o}C), with a median of 26 ^{o}F (-3 ^{o}C). The maximum temperatures during each event range from 34 to 49 (1 to 9.5 ^{o}C), with a median of 42 ^{o}F (5.5 ^{o}C). While the events are identified from Sacramento temperature criteria, hardest freeze events in this target sample are often noteworthy for their statewide economic losses. For example, the December 1990 freeze caused statewide economic losses totaling $3.4B. In that event, temperatures in the southern San Joaquin valley citrus growing region were below critical values for 4-12 hours each night for several nights in a row and caused 100% loss of fruit; San Jose recorded its second lowest temperature on record, 19 ^{o}F (-7 ^{o}C), on three nights during this outbreak. Other events in this target sample have similar statewide coverage.

The composite analysis indicates that hardest freezes are preceded by higher temperatures and heights in both the southeastern US and over much of Alaska. Significant areas defining these two regions are consistently found at all tropospheric levels tested in geopotential. As these two ridges occur, a deep sharp trough (in geopotential height and temperature) amplifies in between as cold temperatures are ultimately carried over California. Fig. 2 illustrates the evolution of the geopotential height pattern. Over the 3

days prior to onset, the Alaskan ridge strengthens until -48 hours, and finally the ridge axis rotates from north-northwest to north-northeast during the last two days prior to event onset. This ridge orientation is well known to local forecasters. *Not* noticed by local forecasters is the significance of a geopotential ridge in the southeastern USA that changes little at 300 hPa over the 3 days prior to onset; below (at 700 and 850 hPa) this ridge appears and grows over the final day or so. Significantly warm temperatures prevail in the southeast, especially at upper levels several days prior to the onset and at most levels just before the onset. A representative lower tropospheric temperature evolution is seen in Fig. 3. Geostrophically consistent wind fields also occur, such that the warming over Alaska and the southeastern USA are associated with warm advection by southerly winds just to the west. The zonal wind component at jet stream level is displaced northwards over Alaska and the eastern USA while southward over the western USA. Thus cold air stays colder traveling over land than would occur from crossing over the Pacific Ocean. The meridional component shows unusual southerly winds west of both ridges and unusual northerly winds west of the Canadian trough at all four levels examined (850, 700, 500, 300 hPa). Because the flow is more northerly instead of northwesterly over western Canada, the zonal wind is weak to the north and strong to the south of the Canadian trough at all four levels. Vertical motion is rising near the west side of each ridge with sinking on the west side of the trough in between. The tropopause elevation changes consistently. The moisture fields show highly significant, unusually low humidities associated with that western USA trough. That is to be expected for a polar air mass moving over a continent in winter. However, most of Alaska has unusually high humidities at most levels, even several days before event onset.

The ridge-to-ridge zonal wavelength is 60-70 degrees longitude, depending on the level. The meridional separation between the two ridges (measured from the middle of the area of highest significance) is 15-25 degrees latitude. (The meridional extent varies between different variables and levels.)

The ridge in the southeast may be the most unexpected result. Cold air outbreaks affecting the central and eastern USA are also led by the Alaskan ridge. So, the ridge in the southeast may help differentiate outbreaks affecting California from those only affecting areas to the east by helping direct the cold air further west. Colucci and Davenport (1987) have some discussion of how a ridge over Alaska precedes downstream cold air outbreaks over the eastern half of the USA; they do not seem to find a ridge in the southeast in their cases.

Fig. 4 illustrates the variation between cases by showing 500 hPa geopotential heights *anomaly* field at the onset of each long duration fog event. While there is some variation between the events, nearly all have the three primary features. All events have a trough over or close to California, though that is hardly surprising given the event criteria. Ten of eleven events have a small ridge near the southeast coast of the United States. Ten of the eleven events have a ridge over Alaska. The consistency between events makes this field useful for forecasting future advective hard freezes.

b. Heaviest prolonged rain

Fourteen heavy prolonged rain events occurred during the months of November (3 events) through April (1 case) from 1979 - 1999. The duration ranged from 24 to 58 hours at average rates from 1.2 (for the longest duration event) to 3.7 inches (30 to 94 mm) per day. Total rainfall varied from 2.3 to 6.8 inches (58 to 173 mm).

These heavy prolonged rain events impact a much wider area than just the southern Sacramento Valley. Ely et al. (1994) studied winter flooding in southwestern USA desert watersheds but their studied time period has little overlap with our time period. Nonetheless, the earliest of our heavy rain events subsequently appears (2-4 days later) as the latest date in lists of winter floods of southern California's Mojave river and drainages of central Arizona studied by Ely et al. (1994). The broad extent of the 8-10 January 1995 event in our target sample can be seen in Miller and Kim (1996; their fig 2, whose caption appears to have a misprint.)

In devising appropriate criteria, one must focus on the goal. For example, if the goal is to identify periods of peak river flow and associated flooding, then that is much more complex than measuring precipitation at the bottom of a watershed. Peak river flow involves additional factors such as soil saturation, whether the rain accelerates snowmelt, and river flow rate prior to the event onset. An excellent case in point is the so called New Year's Eve storm (e.g. Galewsky and Sobel, 2005) from 31 December 1996 through 2 January 1997. This storm produced the highest 3-day flow recorded on the American River through Sacramento. Heavy rainfall affected the coastal ranges and northern Sierra Nevada Mountains of northern California; the regional flooding during the month of

January 1997 caused more than $2B in damages. The New Year's Eve event just missed passing our criteria based upon Sacramento weather stations (the high river flows were partly due to warm rain melting snow in the Sierra Nevada Mountains and above normal precipitation throughout the month of December 1996). An event 3 weeks later that month *did* meet our criteria based on Sacramento weather stations. Some readers might view the miss of the New Year's Eve event as a problem with our criteria, however other nearby Central Valley stations are consistent with the Sacramento ranking of these two events in January 1997. An alternative set of criteria might average a wider array of stations, including foothill and mountain stations but that would not be consistent with other event types discussed in this report. (Such a scheme would certainly fail for long duration fog events.) Interestingly, many of the features of the New Year's 1997 storm are consistent with the composite: upper and lower level flows from the southwest drew upon moist subtropical air, a cloudband extended from near Hawaii to northern California (these features can be seen, for example in, Galewsky and Sobel, 2005). One inconsistency with our composite is the midtropospheric ridge over Alaska in our composite (fig. 5) is located west over the Bering Sea on New Year's day 1997(not shown). While the New Year's Eve event is not selected, two other prominent flooding events from the period of study are flagged by our criteria. The first event, in April 1982, caused damage across much of the state, including a huge landslide that closed for 2 years California highway 1 along the Big Sur coast. The second event, February 16-17, 1986, caused the second highest recorded 3-day flow on the American River. So if flooding is the issue of interest, a different set of criteria should be adopted.

This extraordinary event type is the only one we delimit with one-hour accuracy because a heavy rain period can start at any time of day including just prior to the start of a new day. Hence the time period used in the criteria starts with the hour the rain event commences, not 0 Z or 12 Z. (In contrast, longest duration fog events have longer periods and require just one report in a day; hottest heat waves and hardest freezes are based on daily maximum and minimum reports which are clearly linked to specific dates.) However, we only use 0 Z and 12 Z observing times of upper air data because radiosonde data are incorporated into the data at those times. (The climatology of some variables, notably moisture, can be different at 6 and 18 Z.) Twelve of the fourteen heavy rain occurrences start at times closest to 12Z.

Dates used for testing were chosen to avoid differences created by seasonal and diurnal cycles. The statistical testing done depends upon when the event occurred in the rainy season; hence, the two March and April events are excluded for tests of *total* fields. Anomaly fields have the long term daily mean removed and so remove the seasonal cycle. For some variables that have a strong diurnal variation, such as low level temperature, the statistics are separately applied to those events closest to a particular upper level observing time. Two null distributions are created for variables with a notable diurnal variation: one distribution drawn solely from 0 Z data, the other solely from 12 Z data. In this case, only the ten heaviest prolonged rain events from November through February that commenced closer to 12Z than to 0 Z are compared with the 12Z null distribution. For other variables (mid and upper troposphere) a single null distribution is sufficient for the 12 events that occurred from November through February. For the

statistical assessment using the *anomaly* fields, all fourteen dates are taken for the variables without daily oscillation, and twelve for those other variables with strong diurnal variation. In spite of all these differences, a great seasonal homogeneity between all the fields is reached.

The upper air composite pattern (see Fig. 5) has a ridge in geopotential height and temperature over Alaska, somewhat similar to the hardest freezes target group. However, there is no characteristic ridge over the southeastern US. The Alaskan ridge[2] is less prominent for these heaviest prolonged rainfall events (see below). The Alaskan ridge is present even 3 days prior to heavy rain onset in the composite. There is a deep main trough which forms and amplifies over the eastern north Pacific during the 3 days prior to the event. Upstream of this main trough weak (but not significant) troughs are seen that appear to merge with the main low, possibly slowing the eastward motion of the main trough. The rainy period over Sacramento is thereby prolonged as one or several short waves migrate around the main trough. Ahead of this main trough significant southwesterly winds at mid and lower tropospheric levels (not shown) advect moist tropical and subtropical air over California providing the moisture needed for heavy precipitation. Lower tropospheric relative humidity is significantly high (not shown) along and southwest of the California coast. Relative humidity at 850 hPa exceeds 90% over northern California at the start of the heavy precipitation. Not unexpectedly, upward

[2] A reviewer points out that Sacramento forecasters often find that when a ridge over the Pacific migrates northward, a trough is likely to 'undercut' that ridge and bring heavy rainfall to Northern California.

motion (not shown) in the reanalysis data is significant and unusually strong over much of California. These significant features are familiar to most forecasters in northern California. As such, the identification of these features with the bootstrap analysis becomes a check upon the method. Other features that are less well known to forecasters are features that precede the event or are located further away. The anomalous flow over California is linked to contrary changes over the Gulf of Alaska and adjacent North America. Normally strong westerly winds over the Gulf of Alaska are significantly reduced for the obvious reason that the Aleutian low is south of its common location. Vertical motion over the Alaskan 'panhandle' is significantly downward.

One note of caution: at the beginning of the event, i.e. at the relative hour 0, the offset between two occurrences can be as large as 11 hours because the data are available only twice a day. That is already important, but the further back in time one goes, the greater may become the offset in the location of the Pacific trough between the members of this ensemble, due to differences in speed or pattern. That explanation has an important consequence for the composite maps: the mean trough *may* appear to deepen because the individual troughs for each event become progressively more in phase.

Another caution is that the ridge over Alaska is not seen in many individual events of the heaviest prolonged rain target group. Fig. 6 illustrates the variation between cases by showing 500 hPa geopotential heights at the onset of each heaviest prolonged rain event. Our subjective impression is that a ridge over Alaska is present only in about half the events. The trough off the west coast is present in all cases, but its strength and shape

vary between cases as well. Because of the wide variation between individual events, this field is probably not useful as an independent indicator for heaviest prolonged rain, despite the statistical significance of various parts of the field. A forecaster may find it more useful to consider this field in conjunction with several other fields, most obviously a forecaster would place greater emphasis upon moisture content indicator fields of the upstream air. All 14 cases and of course the composite (not shown) have significantly high lower tropospheric specific humidity (and relative humidity) plus southwesterly winds to the southwest of Sacramento.

Heavy precipitation events in the western USA have been studied by many researchers. Ely et al. (1994) show composite maps (with 10% confidence regions shaded) of 700 hPa geopotential height anomaly; their map for the Mojave river drainage has large negative anomaly centered off the California coast and extending into the Pacific Northwest with a positive anomaly over northern Alaska, similar to Fig. 5d and the ensemble mean in Fig. 6. (700 hPa geopotential height, not shown here, matches the Ely et al. pattern and is posted at http://atm.ucdavis.edu/~grotjahn/Analogs/heavy_rain/heavy_rain.htm.) Lackmann and Gyakum (1999) show 500 hPa geopotential height composite maps for 46 heavy rain events in the Pacific Northwest, also shading regions of 95% and 99% confidence; they also track these fields back in time. Lackmann and Gyakum find a significant ridge over eastern Siberia leading an amplified trough in the Gulf of Alaska that in turn leads development of a significant ridge over the western states. Their pattern differs from the one shown here: we find heavy rain in central California preceded by a ridge over northern Alaska and then a significant trough *due south* of it in the northern

Pacific (centered near 30-35 N) and we do *not* find a significant downstream ridge. Higgins et al. (2000) include lagged composites for 25 heaviest 3-day rainfalls in groups of grid points along the USA West Coast. They show similar features as we find: a ridge at 200 hPa, southwesterly flow with a long fetch extending back into the tropics for events near southwest Oregon. As noted above, the Alaskan ridge feature has limited forecast use due to wide variation in our study; it is unclear how much variation occurred between the 46 events studied by Lackmann and Gyakum or the 25 studied by Higgins et al. (2000). Lackmann and Gyakum also show a somewhat opposite pattern of 850 hPa temperature than we find. We find a warm anomaly over Alaska and cold anomaly over and near the Pacific Northwest. In our composites, the Alaskan thermal ridge exceeds the 1% confidence criteria for >60 hours prior to the event suggesting it is often a persistent feature.

c. Longest duration fog

Radiation fog formation in the central valley of California is a common wintertime event. Underwood et al. (2004) and references therein discuss the conditions during which fog occurs. Occasionally, the fog episodes have unusual duration. Twenty five events meet our criteria for longest duration fog events during the months of November through February from 1978 – 1999. The longest duration event was 17 days, the next longest was 11. Fourteen of the twenty-five events lasted the minimum duration of 5 days. Holets and Swanson (1981) studied 7 fog episodes in the Central Valley selected from the period 1954 through early 1980. The episodes they studied range from 4 to 13 days, but they do not include all episodes of significance. (They included a 4 day event in January 1980.

They did not include a 6-day episode we study in the same month.) Holets and Swanson did not composite results and focused on the local meteorological conditions so it is not possible to assess what broader meteorological conditions are key from the few specific episodes that they described. Underwood et al. (2004) discuss 20 of 31 long duration Central Valley fog events identified by their criteria during the 'fog seasons' of late 1997 through early 2002. Their focus was on identifying *spatially* extensive fog episodes, based upon visibility below 1 mile simultaneously at any 3 of these stations: Bakersfield, Fresno, Merced, Sacramento, and Chico. Though there is little overlap in the time periods studied by them and here, 3 of their events overlap 2 of our events (a single event lasting 7 days is counted twice by Underwood et al.). Several of their events occur primarily in the San Joaquin Valley, where winter fog is more common and persistent.

The most obvious feature found during the longest duration fog events is a broad blocking ridge (Fig. 7a) that establishes itself over the west coast of the United States during winter, after sufficient rainfall to saturate the upper soil levels over the preceding days or weeks. A similar ridge (geopotential height at 500 hPa) is located by Underwood et al. (2004). The pattern of 850, 700, 500, and 300 hPa heights is similar to a heat wave condition during summer (see below) in having a ridge over the coast with a trough to the west in the middle North Pacific. The distance from ridge to trough is about 40 degrees longitude at 40N. Closer inspection finds the 700 and 500 hPa ridge for longest duration fog events centered right at the coast, but for a heat wave the ridge at those levels is centered over Nevada. At 850 hPa, the ridge for longest duration fog events is centered over Nevada. Of course heat waves occur in summer while long duration fog events

occur in winter. But long duration fog events for Sacramento often bring unusually warm temperatures to the adjacent Sierra Nevada Mountains and San Francisco Bay area.

This stable pattern has sinking motion over the region (Fig. 7b) which is not particularly significant *statistically*, but clearly the sinking plays a role in maintaining the 'high inversion' typically found (e.g. Holets and Swanson, 1981) above the fog. Weak, often offshore flow occurs in the lower troposphere, above the fog layer (e.g. 850 hPa winds shown in Fig. 7d). The ridge is characterized by warm lower tropospheric temperatures that are significant offshore (Fig. 7c).

The composite maps have limited forecast utility because few areas are significant at map times prior to the onset of the event. Hence, one is limited to matching the target group at the onset of the event with model-generated forecast maps. Larger and longer significant areas at times preceding event onset show up in the standard deviation fields, indicating that a stable pattern needs to set up first prior to the development of the fog. The preconditions described here are consistent with forecaster experience. The lead time is smaller for long duration fog onset than for a hard freeze. A day before the onset, the geopotential height ridge over the West Coast is present but not strong enough to surpass the significance threshold. It varies with level, but appears in 15 to 17 of the events comprising the composite. What is significant 24 hours before onset are two troughs over the Pacific. Both troughs are initially moving eastward, then slow down. The leading of the two troughs helps to build the west coast ridge, consistent with anomalous significant

southerlies ahead of the leading trough. The leading trough is present in 15 to 17 of the events comprising the composite (again depending upon the level).

d. Hottest heat waves (onset and end)

A general criterion for identifying a heat wave might be checking whether the apparent temperature (or the related quantity: heat index) stays above some threshold over the course of one or more days (Robinson, 2001). However, the highest Central Valley temperatures generally occur with low relative humidity, hence the difference between maximum apparent and actual temperature is small. So, our criteria are reasonably based on actual temperature. Robinson (2001) requires that temperatures remain elevated at night (heat index staying above 81 ^{o}F, 27 ^{o}C) to distinguish heat waves from 'hot spells'. The heat index in Sacramento rarely remains above 27 ^{o}C overnight and few events in Sacramento appear to meet Robinson's criteria. However, the hottest heat waves identified here are notable in having very high or record setting electric power demand. The heat waves so identified also lead to increased heat-related fatalities. Fifteen heat wave events met our criteria during the months of July (6 events), August (8 events) and September (1 event) during the 1979 – 1999 years. The longest duration was 9 days; three of the events only lasted 3 days. The maximum temperature in a given event ranged from 112 to 106 ^{o}F (44.4 to 41.1 ^{o}C) and occurred 1 to 4 days after the onset of the event. Both the beginning and the end of the heat wave period are studied. The heat wave typically extends over much of California.

The usual way a heat wave is ended *in the Sacramento region* is by a reversal of the low level winds to become a sea breeze sufficiently strong to penetrate well into California's Central Valley. The sea breeze is known locally as the Delta Breeze and is so named because the sea breeze enters the Central Valley mainly at the delta formed by the Sacramento and San Joaquin rivers. The Delta Breeze is challenging to study, because contrary to the start, the end of a heat wave often happens for a small portion of the Central Valley. Forecasters see only slight differences of geopotential height and sea level pressure fields between times when the Delta Breeze is or is not present. Furthermore, the Delta Breeze may be strong enough to end the Sacramento area heat wave, but too weak to moderate maximum temperatures at the northern and southern ends of the Central Valley.

The evolution of the composite pattern (Fig. 8) for the hottest heat waves target group begins, ironically, superficially similar to the composite for a hardest freezes: it is preceded by warmth near Alaska and in the southeast US (and significant ridging in geopotential at both locations, Fig. 8a). However, important details differ: the ridge over Alaska is shifted westward by half a wavelength compared to the hardest freezes pattern. Also, the ridge in the southeast does not linger and a huge ridge over the western USA quickly grows to prominence. Another difference is the ridge-trough-ridge locations hardly change between the map times shown. Subjectively, the height contours in Fig. 8 appear to illustrate downstream development: first the ridge over the Aleutians, then a mid-Pacific trough, then the ridge over the western USA. However, it is outside the scope of this study to investigate whether downstream development (e.g. Chang, 1993) actually

explains this evolution. However, anomaly fields (not shown) have a clear west to east progression: 60 hours before heat wave onset the Aleutian ridge is stronger than the trough and the West Coast ridge is nonexistent. By 24 hours prior to onset the Aleutian ridge has been diminishing for more than 12 hours, the West Coast ridge has begun to amplify for at least 12 hours, and the Pacific trough in between is now the strongest feature. By the time of onset, the West Coast ridge has reached comparable strength to the Pacific trough while the Aleutian ridge has become the weakest of the three anomaly features.

The combination of a broad ridge over the western USA plus trough in the mid Pacific is similar to the longest duration fog pattern. The geopotential trough and ridge are further east for the hottest heat wave composite than for the longest duration fog. The distance from the western USA ridge to the Pacific trough is about 40 degrees longitude at 40N. The superficial similarities of parts of the hardest freezes, hottest heat waves, and longest duration fog patterns suggest that deflections in the generally westerly flow can lead to an extraordinary event on the downstream side of significant areas in the pattern. Unlike the hardest freezes, the lead time is smaller for the hottest heat waves onset (as judged by how long before onset there are maps with sizable significance areas).

Generally, the hottest heat waves are associated with a deep upper level trough somewhat east of the middle of the north Pacific (near 160W). On the east side of that upper level trough are significant middle and upper tropospheric southerlies (not shown). Not surprisingly, temperature fields at all levels tested are significantly warm over the much

of the west coast. However, these significantly warm temperatures (e.g. Fig. 9d) are well east of the significant southerly winds, and are located where lower tropospheric winds are northerly (though not significantly so). Somewhat contrary to the geopotential height fields, the Aleutian thermal ridge keeps on strengthening, whereas the associated geopotential ridge dimishes during the 2 days before onset. Curiously, 850 hPa temperatures (Fig. 9) and surface temperatures (not shown) are elevated in a large area of the Pacific (west and northwest of Hawaii) well before the onset of the heat wave event. The lower tropospheric temperature evolution over the Pacific is consistent with the geopotential height changes above that may be indicative of downstream development.

The presence of a large ridge in geopotential height where the heat wave occurs is hardly surprising, given the elevated temperatures through a large depth of the troposphere. However, such a ridge is commonplace during summer and the ridge being significantly higher during the hottest Sacramento heat waves is notable. What is less well recognized by forecasters may be the significant trough further west. This trough is present in all 14 events. Even less well recognized are the significant ridge and elevated temperatures, through the depth of the troposphere (near and south of the Aleutians) that precede the onset by more than a day. This near-Aleutians ridge is strongest in the composite at 36 hours and is clearly present in 11 of the 14 events. Also unrecognized is the presence of a significant ridge over the eastern and southeastern USA prior to onset. This eastern and southeastern ridge is very weak or missing in 4 of the cases and present in the remaining 10 at 36 hours prior to onset of the hottest heat waves.

One might expect a heat wave to be led by rising temperatures locally. While the 850 hPa temperatures on the preceding afternoon are not significantly warm locally (Fig. 9b), the heat wave onset is led by unusually warm temperatures the night before (Fig. 9c).

Target groups were also formed from the dates on which a heat wave ends. The heat wave end initiated by the onset of a strong Delta Breeze is a mesoscale phenomenon. Over much of the western USA high temperatures may continue even as the Delta Breeze cools Sacramento maximum temperatures by >15 ^{o}C. Hence, much of the significant areas found in geopotential and thermal patterns look very similar to the situation when a heat wave begins. An exception is the sea level pressure pattern, shown in Fig. 10. During a heat wave, sinking and offshore flow prevail, suppressing the sea breeze. During the worst of a heat wave, the SLP gradient is directed northward or northeastward. The sea level pressure (SLP) gradient undergoes a subtle shift when the Delta Breeze is established: the SLP gradient becomes northwestward. However, the end of the hottest heat waves composite may have limited utility for forecasting because the change in the SLP gradient is subtle. Local forecaster experience also includes an onshore movement of a weak middle tropospheric trough. The positive vorticity advection is believed to facilitate deepening of the moist cool marine boundary layer offshore while fostering an onshore movement of that cooler air. Perhaps the features are too weak or their timing too variable from case to case, but upper air charts for the target group do not pick up such a feature.

5. Summary and Conclusions

This study applies statistical tests to composite maps before and during extraordinary weather events affecting the southern Sacramento Valley in California. The extraordinary weather events have large scale signatures that can be tracked backwards in time for several days. Four weather events are described here: hardest freezes, heaviest prolonged rain, longest duration fog, and hottest heat waves.

The hardest freeze episodes are more homogeneous than the other cases considered here. Fig. 4 illustrates the homogeneity by showing the 500 hPa geopotential height anomaly field for each of the 11 events in this target ensemble. All but one hardest freeze events have a ridge over Alaska and the southeastern USA; all have a trough over California. Heaviest prolonged rain events seem the least homogeneous and Fig. 6 illustrates how a significant feature of the composite field (a ridge over Alaska) is present in only about half of the individual members. One reason for the heterogeneity of the heaviest prolonged rain events is that phase offsets remain between members of the target group due to variation in the timing of the event relative to upper air observing times. The formation of dense fog during only one hour five days in a row is met 25 times, but using six days in a row finds only 11 events. Those features that are significant in the composites of longest duration fog events tend to be significant over small areas and not long before the event onset. Accordingly, significant features (a trough between the Aleutians and Hawaii and a weak ridge along the West Coast) in the geopotential height field composite are found in only about two-thirds of the events. The hottest heat waves

occurrences appear very homogeneous and allow confidence in the accuracy of those results. Significant features of the composite fields tend to be found in nearly all of the 15 members of the ensemble (West Coast ridge in 14; mid north Pacific trough in 14; trough southwest of Aleutians in 13).

Not surprisingly, the large scale patterns for these events have either a trough or ridge in geopotential height centered over or near California. What may be less obvious to a forecaster is the significant role prior to and during the onset of some events (hardest freezes and hottest heat waves) played by a ridge (in geopotential and temperature) over the southeastern USA. Hardest freezes and heaviest prolonged rain events are preceded by thermal and geopotential ridges over Alaska. Hottest heat waves are preceded by a deep trough to the west in the mid Pacific that in turn was preceded by a ridge even further west (and north) near the Aleutian Islands. For heaviest prolonged rain, geopotential heights are sometimes unusually high over Alaska; meanwhile a large trough, normally near the Gulf of Alaska is typically unusually far south, to the west of California (near 140 ^{o}W). Hottest heat waves and longest duration fog occur in opposite seasons but share the occurrence of a ridge over the western US or along the West Coast (respectively) and a trough further west (near 160 ^{o}W).

As noted, an Alaskan upper level ridge occurs with both hardest freezes and heaviest prolonged rain in the Sacramento region. In a study of Alaskan blocking ridges, only 4 of our 14 heaviest prolonged rain events correspond with the 37 blocking episodes identified by Carrera et al. (2004) over nearly the same time period. We also find 5 of our 9 hardest

freezes episodes (occurring in the period considered by Carrera et al.) match periods in their 37 blocking events. No hardest freeze periods match the heaviest prolonged rain periods, nor should they. Closer inspection of Carrera et al. results (e.g., their figs. 3g-3j and 4a-f) do show a small trough off the California coast (similar to our heaviest prolonged rain results) *and* a stronger one over Canada with a portion extending into the northern Rockies (similar to our hardest freeze events). Hence, a mixture of events follows those Alaskan blocks: cold air outbreaks over California or heavy prolonged rain in California or *neither*. Hence, that Alaskan ridge is not sufficient to lead to heavy Sacramento area rainfall or the quite different Central Valley hardest freeze events. Reliance solely upon just the Alaskan block would lead to 'false positives': prediction of an event that does not happen. For heaviest prolonged rain one must consult other fields than geopotential and temperature; elevated moisture to the southwest of Sacramento (along with the southwesterly winds that are consistent with the mass and thermal trough offshore) occurs for every heaviest prolonged rainfall event. Thus, a forecaster cannot rely on one anomalous feature, but is led to note the combination of significant events: ridges and troughs elsewhere and patterns in several variables when assessing the likelihood of an extraordinary event.

Acknowledgements

The authors thank Mr. Scott Cunningham (former SOO, Sacramento NWSFO) for giving useful advice regarding the extraordinary events in Sacramento area and remarks on the results. The authors thank Prof. Richard A. Levine (SDSU) for his valuable statistical advice. The authors acknowledge partial support provided by the National Science Foundation through grants ATM-9615316 and ATM-0354545. G. Faure received additional support from Ecole Nationale de la Météorologie (Météo-France). The authors further appreciate the many helpful comments by anonymous reviewers.

Appendix :

Statistical enhancements

The algorithm presented in section 2 is able to produce good results, but several enhancements were tested. The aim of the first of them is to increase the quality of the composite maps, while the other enhancements seek a more powerful statistical analysis.

a. Anomaly fields

The composite maps and the statistical treatment described in section 2 can be performed on the 'total' fields (i.e. the fields as supplied by the NNRA1 archives). The results shown here make this choice to suit the forecast aim of the project, by working on fields available and meaningful to forecasters.

The problem with total fields is that their values are not comparable all year long. So the statistical assessment was performed using only the three or four months where the great majority of the occurrences of a given type of event take place. In addition, anomaly fields were formed by removing the (30-year) long term daily mean (for that observing time) from each data value. Using such anomaly fields avoids biases created by seasonal trends in the data. Composite maps from the anomaly fields are less useful for direct forecasting than are the maps from the total fields. However, anomaly composite maps can improve confidence in the results for total fields when both fields have similar significance patterns. All results shown here are strongly similar to corresponding analyses of anomaly fields.

b. Double null distributions

Some variables, especially lower tropospheric and surface variables, have a strong diurnal variation. The standard reporting times of 0 Z and 12 Z correspond to late afternoon and early morning local times at the Sacramento region. Hence, target or random groups that mix the two reporting times are not appropriate for such variables. The problem is illustrated in Fig. A1a. A target sample mean containing only 12 Z or 0 Z data would be nearly always significant if compared only to random groups formed with a mixture of day and night values. Except for the heaviest prolonged rain case, the target groups found here are composed solely of either 0 Z or 12 Z data.

Care must be taken when forming the random samples so that those do not mix the reporting times. Since maps at times prior to the event commencement are found at 12 hour intervals, a different null distribution must be used for the odd multiples of 12 hours prior to the event than is used for the other null distributions. Hence, there are two null distributions: one based on randomly chosen 0 Z data the other on randomly chosen 12 Z data.

The case of heaviest prolonged rain is a bit more complicated because in every group of occurrence dates 12 Z and 0 Z data are mixed. If the ratio between 12 Z and 0 Z is not near 1 over 2, the results could be biased the same way as shown before. In fact the case of heaviest prolonged rain in Sacramento area was very similar to the other events, because all but two dates (out of fourteen) are in phase, i.e. round off to 12 Z.

c. Statistical assessment of global significance

All the tests of significance described up to this point are performed grid point by grid point. That means that the significance of each map by itself is unknown. One can think that as long as each test has the confidence level α, so has the whole map. However this problem, known as the multiplicity problem, is not that simple.

An analogy with rolling a die is easier to understand: a map with one hundred independent grid points, each of them assessed at 5%, can be associated with one hundred rolls of a 20-sided die. If the map were significant at 5%, it would mean that the probability for a random map to have five or more significant grid points is 5%. This probability is the same as the one to obtain five or more times the face 1 while rolling one hundred times a 20-sided die. This probability calculation, well known to statistics students, has an actual value of 38%! In fact the probability to obtain **n** times the face 1 goes under 5% only for **n** greater or equal to ***nine***. It means that, under these conditions, a map having one hundred grid points is significant at 5% only when its number of significant grid points is superior or equal to ***nine***.

The maps could be assessed this way, but in fact the grid point values in a meteorological field are not independent, so significance tests applied to every grid point are not independent. Thus a larger number of significant points can be expected by chance for the same confidence level, but its precise value is unknown.

To assess what will be called the global significance of the maps, the bootstrap method has been used again (Matthews and Kiladis, 1999). The test statistic is the number of grid points significant at the **α** level. The corresponding null distribution(s) is (are) calculated from the random composite maps created previously. The number of significant grid points in the target composite map(s) is then compared to that (those) distribution(s) with a one-tailed test. Fig. A1b illustrates the test applied to the 850 hPa geopotential height map at the start of the hottest heat waves.

This significance test identifies the most significant maps, where the meteorological signal (in terms of number of grid points passing a significance test) is bigger than would be expected by chance. *However even if a map as a whole does not have more significant points than might occur by chance, it does not mean that the significant areas on that map are not reliable. Instead, it means that the number of significant points is not larger than the number likely for a random composite map.*

d. Improvement of standard deviation

Even if a target group is as homogeneous as possible, there is not necessarily a reason for the meteorological variables to describe exactly the same pattern from one occurrence to another of an event. For instance, a trough in the Pacific could be further south for one occurrence, further west for another, etc. Therefore, it might be easier to find statistical significance among a region instead of on a single grid point.

A simple but efficient local significance test was used mainly to improve the standard deviation maps. The test replaces the value at each grid point by the mean of the grid point and the eight surrounding values. In practice, the 9-point averaging produced similar maps as using data that are not 9-point averaged. This happens because the fields used are already very smooth, but the significant areas with the latter are sometimes bigger than the ones with the former.

REFERENCES

Attaway, J., 1997: *A History of Florida Citrus Freezes*. Fl. Sci. Source, Inc., 368 pp.

Bevan, L., and G. Cline, 2005: Climate of Sacramento, California. NOAA Technical Mem. NWS WR-272. 87 pp. (Available online at *http://www.wrh.noaa.gov/wrh/techMemos/272.pdf*)

Carrera, M. L., R. W. Higgins, and V. E. Kousky, 2004: Downstream weather impacts associated with atmospheric blocking over the northeast Pacific. *J. Climate,* **17,** 4823-4839.

Chang, E., 1993: Downstream development of baroclinic waves as inferred from regression analysis. *J. Atmos. Sci.*, **50**, 2038-2053.

Colucci, S. J., and J. C. Davenport, 1987: Rapid surface anticyclogenesis: Synoptic climatology and attendant large-scale circulation changes. *Mon. Wea. Rev.,* **115,** 822–836.

Efron, B., and R. J. Tibshirani, 1993: *An Introduction to the Bootstrap*. Chapman and Hall, 456 pp.

Ely, L. L., Y. Enzel, and D. R. Cayan, 1994: Anomalous North Pacific atmospheric circulation and large winter floods in the southwestern United States. *J. Climate,* **7,** 977–987.

Galewsky, J., and A. Sobel, 2005: Moist dynamics and orographic precipitation in northern and central California during the New Year's Flood of 1997. *Mon. Wea. Rev.,* **133**, 1594-1612.

Gershunov, A., and T.P. Barnett, 1998a: ENSO influence on intraseasonal extreme rainfall and temperature frequencies in the contiguous United States: Observations and model results. *J. Climate*, **11**, 1575-1586.

Gershunov, A., and T.P. Barnett, 1998b: Interdecadal modulation of ENSO teleconnections. *Bull. Amer. Meteor. Soc.*, **79**, 2715-2731.

Grumm, R., and R. Hart, 2001: Standardized anomalies applied to significant cold season weather events: Preliminary findings. *Wea. Forecasting*, **16**, 736-754.

Hart, R., and R. Grumm, 2001: Using normalized climatological anomalies to rank synoptic-scale events objectively. *Mon. Wea. Rev.* **129**, 2426-2442.

Higgins, J.-K. E. Schemm, W. Shi, and A. Leetmaa, 2000: Extreme precipitation events in the western United States related to tropical forcing. *J. Climate,* **13,** 793–820.

Holets, S., and R. Swanson, 1981: High-inversion fog episodes in central California. *J. Appl. Meteorol.,* **20**, 890-899.

Kalnay, E., M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, Y. Zhu, M. Chelliah, W. Ebisuzaki, W. Higgins, J. Janowiak, K. C. Mo, C. Ropelewski, J. Wang, A. Leetmaa, R. Reynolds, R. Jenne, and D. Joseph, 1996: The NCEP/NCAR 40-year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**,437-471.

Lackmann, G. M., and J. R. Gyakum, 1999: Heavy cold-season precipitation in the northwestern United States: Synoptic climatology and an analysis of the flood of 17–18 January 1986. *Wea. Forecasting,* **14,** 687–700.

Lorenz, E. N., 1969: Atmospheric predictability as revealed by naturally occurring analogues, *J. Atmos. Sci.*, **26**, 636-646;

Matthews, A. J., and G.N. Kiladis, 1999: Interactions between ENSO, transient circulation, and tropical convection over the Pacific. *J. Climate*, **12**, 3062-3086.

Miller, N., and J. Kim, 1996: Numerical prediction of precipitation and river flow over the Russian river watershed during the January 1995 California storms. *Bull. Amer. Meteor. Soc.* **77**, 101–105.

Pauley, P. M., N. L. Baker, and E. H. Barker, 1996: An observational study of the "Interstate 5" dust storm case. *Bull. Amer. Meteor. Soc.* **77**, 693–720. DOI: 10.1175/1520-0477(1996)077

Robinson, P., 2001: On the definition of a heat wave. *J. Appl. Meteor.,* **40**, 762-775.

Stuart, N., and R. Grumm, 2006: Using wind anomalies to forecast east coast winter storms. *Wea. Forecasting*, **21**, 952-968.

Underwood, S. J., G. P. Ellrod, A. L. Kuhnert, 2004: A multiple-case analysis of nocturnal radiation-fog development in the Central Valley of California utilizing the GOES nighttime fog product. *J. Appl. Meteor.,* **43**, 297-311.

Figure Captions

Fig. 1. Null distribution generated while assessing the significance of the 850 hPa mean temperature for the start of hottest heat wave events. This histogram refers to the grid point closest to Sacramento, and compares the means of 1000 random groups. The target group mean has been added and is shown by a heart. 99% of the values stand between the two dashed lines.

Fig. 2 Geopotential height pattern typical of the hardest freezes target group; shown is the 500 hPa level. Contours use 60 m interval for the target group ensemble mean. Shaded areas denote areas passing a 1% significance test. Light shading with a dark edge denotes positive significant anomalies (unusually high heights); dark shading for significantly low heights. Composite maps shown are: (a) 72 hours prior to the onset of the hardest freezes, (b) 48 hours prior, (c) 24 hours prior, and (d) at the time of hardest freezes onset.

Fig. 3. Similar to Fig. 2 except showing temperature pattern typical of the hardest freezes target group; shown is 700 hPa level. Contours use 2 K interval for the target group ensemble mean. Composite maps shown are: (a) 72 hours prior to the onset of the hardest freezes, (b) 48 hours prior, (c) 24 hours prior, and (d) at the time of hardest freezes onset.

Fig. 4. 500 hPa geopotential height anomaly patterns of each member of the hardest freezes target group at the time of hard freeze onset. Contours use 60 m interval; dashed contours are below zero; solid thin contours with darker shading are above zero; thick solid contour for zero value. The larger panel is the target group ensemble mean. The upper left map is the longest event, the event length decreases across then down the page. This field is a good indicator for the hardest freezes because significant features of the composite are found in all or almost all members of the ensemble.

Fig. 5. Similar to Fig. 2 except showing 500 hPa geopotential pattern typical of the heaviest prolonged rain target group. Contours use 60 m interval for the target group ensemble mean. Composite maps shown are: (a) 72 hours prior to the onset of the heavy rain, (b) 48 hours prior, (c) 24 hours prior, and (d) at the time of event onset.

Fig. 6. Similar to Fig. 4 except showing 500 hPa geopotential height anomaly patterns of each member of the heaviest prolonged rain target group at the time of heavy rain onset. There is more variation between these plots than is apparent for the hardest freezes events shown in Fig. 4. The trough off the West Coast is present in all cases but the significant ridge near Alaska in the composite mean is present in only about half the individual events. Consequently, not all significant features of the composite field are useful to forecast the heaviest prolonged rain. A forecaster would need to combine this field with other information, most notably moisture fields.

Fig. 7. Similar to prior diagrams except: showing various composite maps of the target group only at the onset of longest duration fog events. (a) 700 hPa geopotential height with 30 m contour interval , (b) pressure (vertical) velocity at 850 hPa with 4 Pa/s contour interval. Negative values (rising motion) have dashed contours while sinking motion (positive values) use solid contours. (c) 850 hPa temperature with 2 K interval, (d) 850 hPa horizontal wind vectors with scale in m/s below the plot and shading denoting significant areas of zonal wind component.

Fig. 8. Similar to Fig. 2 except showing 700 hPa geopotential height pattern typical of the hottest heat waves target group. Contours use 30 m interval for the target group ensemble mean. Composite maps shown are: (a) 36 hours prior to the onset of the heat wave, (b) 24 hours prior, (c) 12 hours prior, and (d) at the time of heat wave onset.

Fig. 9. Similar to Fig. 3 except showing temperature pattern at 850 hPa typical of the hottest heat waves target group. Contours use 2 K interval for the target group ensemble mean. Composite maps shown are: (a) 36 hours prior to the onset of the heat wave, (b) 24 hours prior, (c) 12 hours prior, and (d) at the time of heat wave onset.

Fig. 10. Similar to Fig. 2 except showing sea level pressure (SLP) pattern typical of the *end* of hottest heat waves target group. Contours use 2 hPa interval for the target group ensemble mean. Composite maps shown are: (a) 60 hours prior to the end of the heat wave, (b) 48 hours prior, (c) 36 hours prior, (d) 24 hours prior, (e) 12 hours prior, and (f) at the end of heat wave.

Fig. A1. (a) Three null distributions generated while assessing the significance of the surface mean temperature for hottest heat wave events. These histograms use the grid point closest to Sacramento and incorporate 1000 random samples. The background one is made from both 12 Z and 0 Z values. The two others contain only 12 Z (foreground one) or 0 Z (middle one) values. (b) Null distribution generated while assessing the global significance of an 850 hPa mean geopotential height map for hottest heat waves. This histogram counts the number of gridpoints significant at the 5% level in each of 1000 random maps. The value for the target group map has been added and is shown by a heart. 99% of the values stand to the left of the dashed line.

Table 1 : Extraordinary weather events and their criteria

Events	Criteria
Hardest freezes	≥ 2 days with minimum temperature < -1 C (30 F) and subsequent maximum temperature < 10C (50 F).
Heaviest prolonged rain	Rainfall lasting ≥ 24 hours in a row with either: a. > 2.5 in (63.5 mm) total for the event or b. > 2.0 in (50.8 mm) per day during the event.
Longest duration fog	≥5 days in a row when visibility becomes < ¼ mile (~400m) at least once each day.
Hottest heat waves	≥ 3 days in a row of maximum temperature > 38 C (100 F) with at least 1 day >40.5 C (105 F)

Table 2 : Summary of the studied fields

Pressure Level Data		
Common features: 17 original pressure levels (except humidity), 4 were consulted (300, 500, 700 and 850 hPa).		
Variable name (NNR1 abbreviation)	Units	Class
Air temperature (air)	degrees K	A
Geopotential height (hgt)	m	A
Relative humidity (rhum)	%	B
Specific humidity (shum)	kg/kg	B
Omega (omega)	Pascal/s	B
Zonal component of wind (uwnd)	m/s	A
Meridional component of wind (vwnd)	m/s	A
Surface Data		
Common features : one level, surface or near the surface (0.995 sigma level)		
Variable name (NNR1 abbreviation)	Units	Class
Air temperature (surf_Tair)	degrees K	B
Omega (surf_omega)	Pascal/s	B
Potential temperature (surf_pot_temp)	degrees K	B
Relative humidity (surf_rhum)	%	B
Sea level pressure (slp)	Pascal	A
Zonal component of wind (surf_uwnd)	m/s	B

Meridional component of wind (surf_vwnd)	m/s	B
Tropopause Data		
Common features : one level : tropopause		
Variable name (NNR1 abbreviation)	Units	Class
Air temperature (trop_Tair)	degrees K	A
Pressure (trop_pres)	Pascal	A
Common features of the dataset used (*original values*) :		
Grid resolution and gridpoint number: 2.5-degree longitude x 2.5-degree latitude grid with 61x29 grid points subset of original (*144x73*)		
Area covered: 70°N-0°N and 140°E-290°E (*90°N-90°S and 0°E-357.5°E*)		
Time length: 1/12/1978 to 31/12/1999 (*1/1/1958 to present*)		
Data frequency: Output every 6 hours (*12*)		
Type of field: Instantaneous values		

(Note to editor: using double spacing spreads table 2 over two pages!)

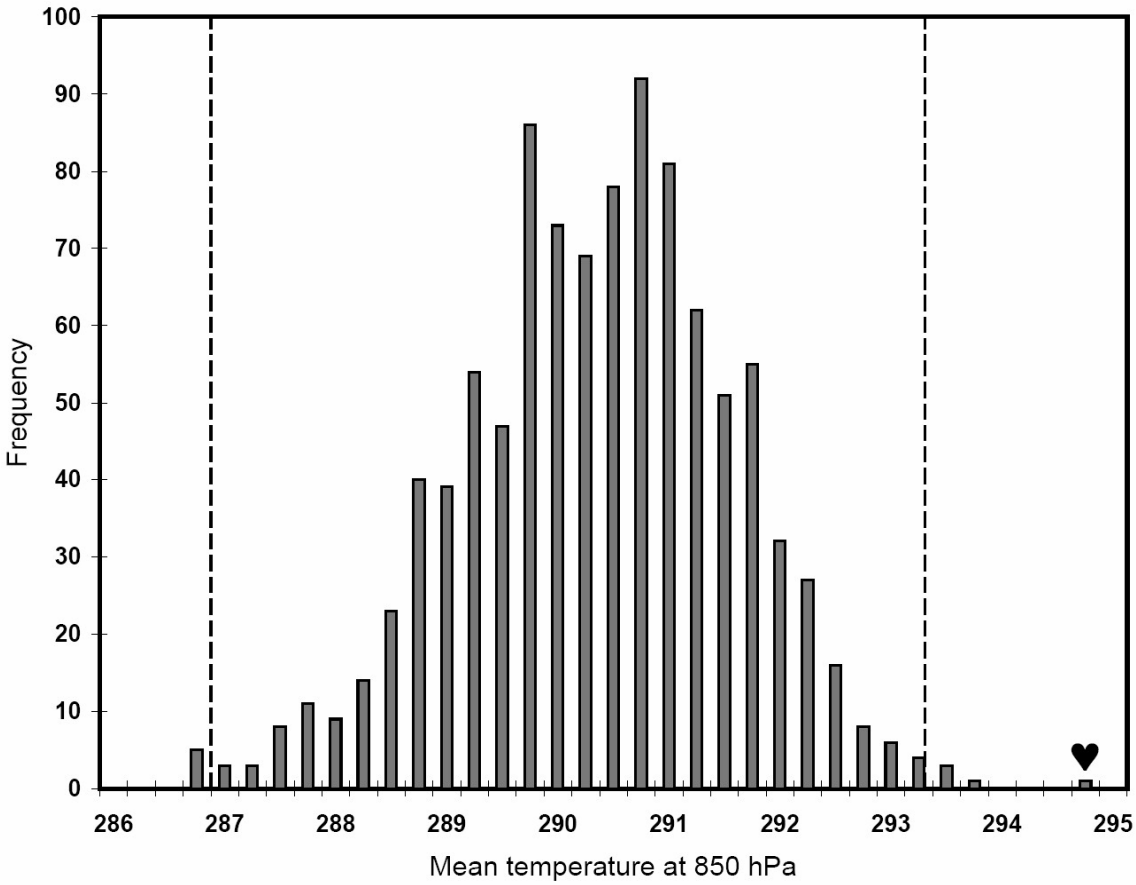

Fig. 1. Null distribution generated while assessing the significance of the 850 hPa mean temperature for the start of hottest heat wave events. This histogram refers to the grid point closest to Sacramento, and compares the means of 1000 random groups. The target group mean has been added and is shown by a heart. 99% of the values stand between the two dashed lines.

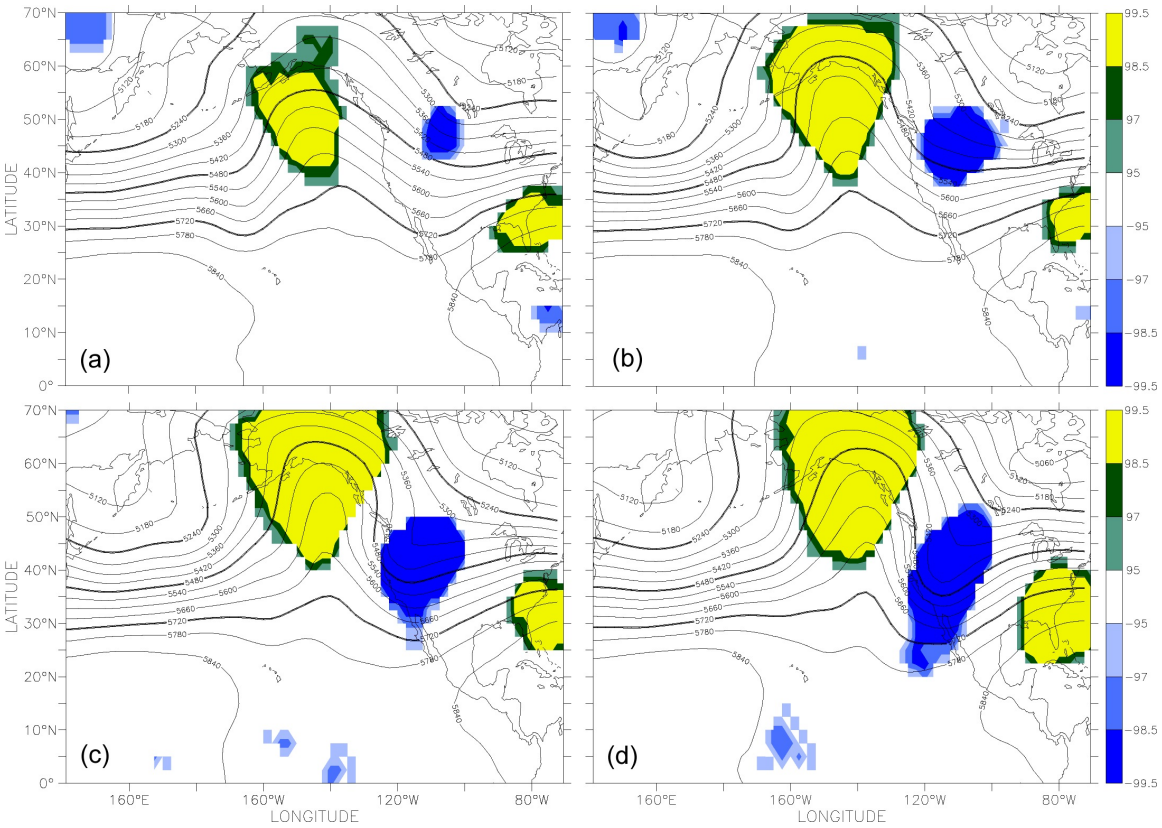

Fig. 2 Geopotential height pattern typical of the hardest freezes target group; shown is the 500 hPa level. Contours use 60 m interval for the target group ensemble mean. Shaded areas denote areas passing a 1% significance test. Light shading with a dark edge denotes positive significant anomalies (unusually high heights); dark shading for significantly low heights. Composite maps shown are: (a) 72 hours prior to the onset of the hardest freezes, (b) 48 hours prior, (c) 24 hours prior, and (d) at the time of hardest freezes onset.

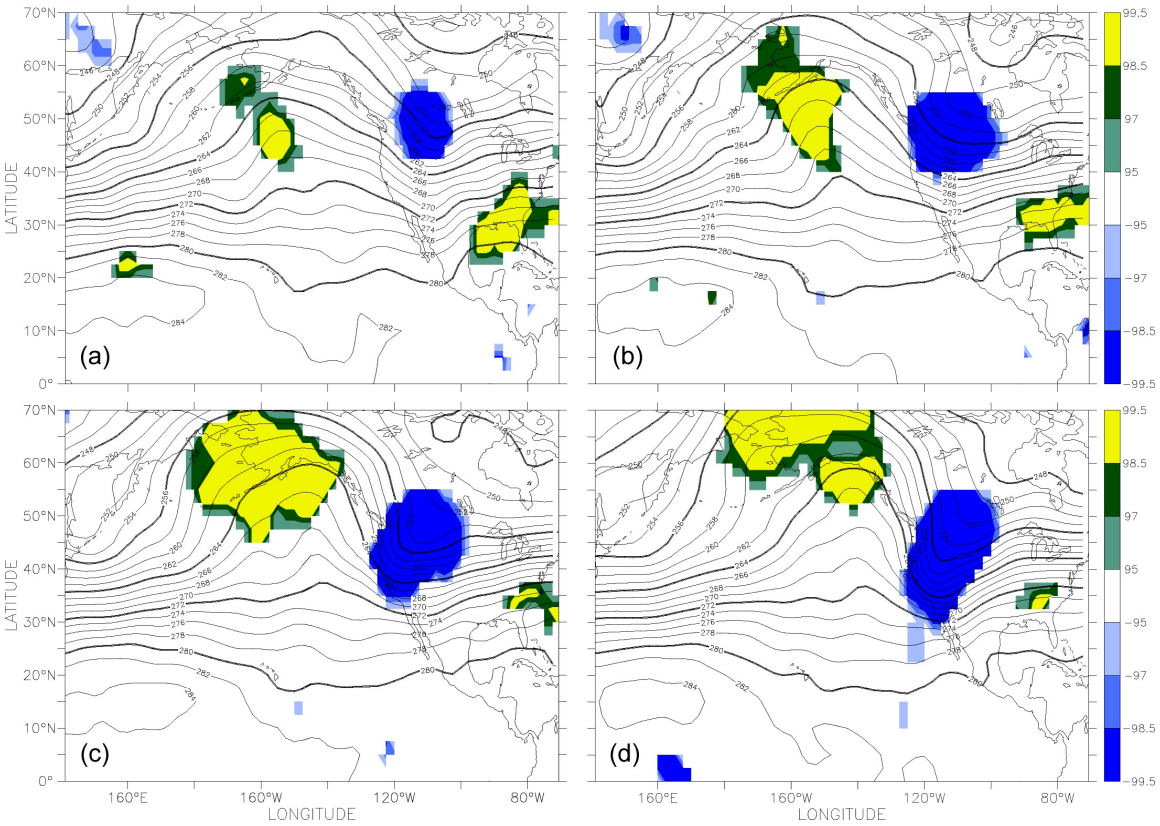

Fig. 3. Similar to Fig. 2 except showing temperature pattern typical of the hardest freezes target group; shown is 700 hPa level. Contours use 2 K interval for the target group ensemble mean. Composite maps shown are: (a) 72 hours prior to the onset of the hardest freezes, (b) 48 hours prior, (c) 24 hours prior, and (d) at the time of hardest freezes onset.

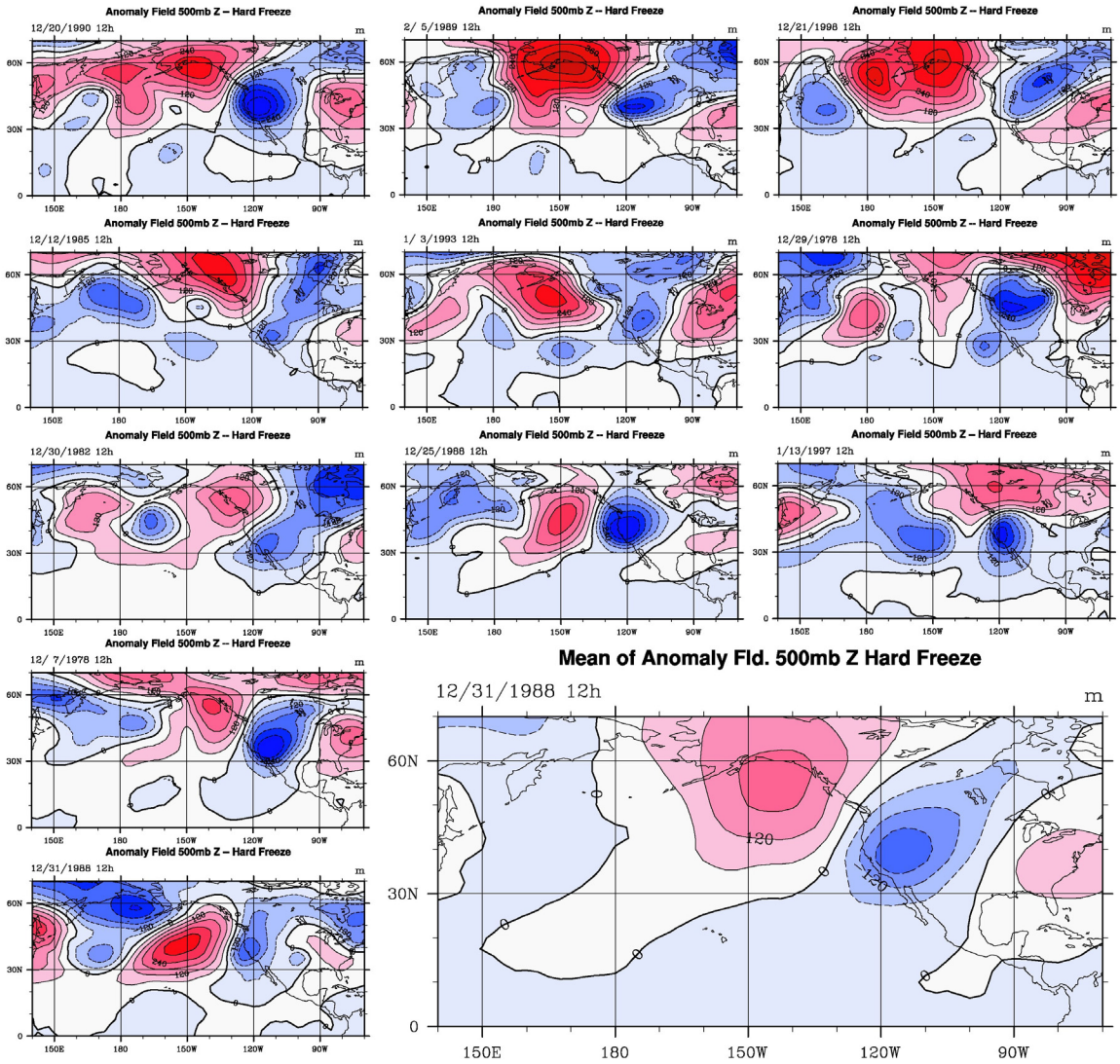

Fig. 4. 500 hPa geopotential height anomaly patterns of each member of the hardest freezes target group at the time of hard freeze onset. Contours use 60 m interval; dashed contours are below zero; solid thin contours with darker shading are above zero; thick solid contour for zero value. The larger panel is the target group ensemble mean. The upper left map is the longest event, the event length decreases across then down the page. This field is a good indicator for the hardest freezes because significant features of the composite are found in all or almost all members of the ensemble.

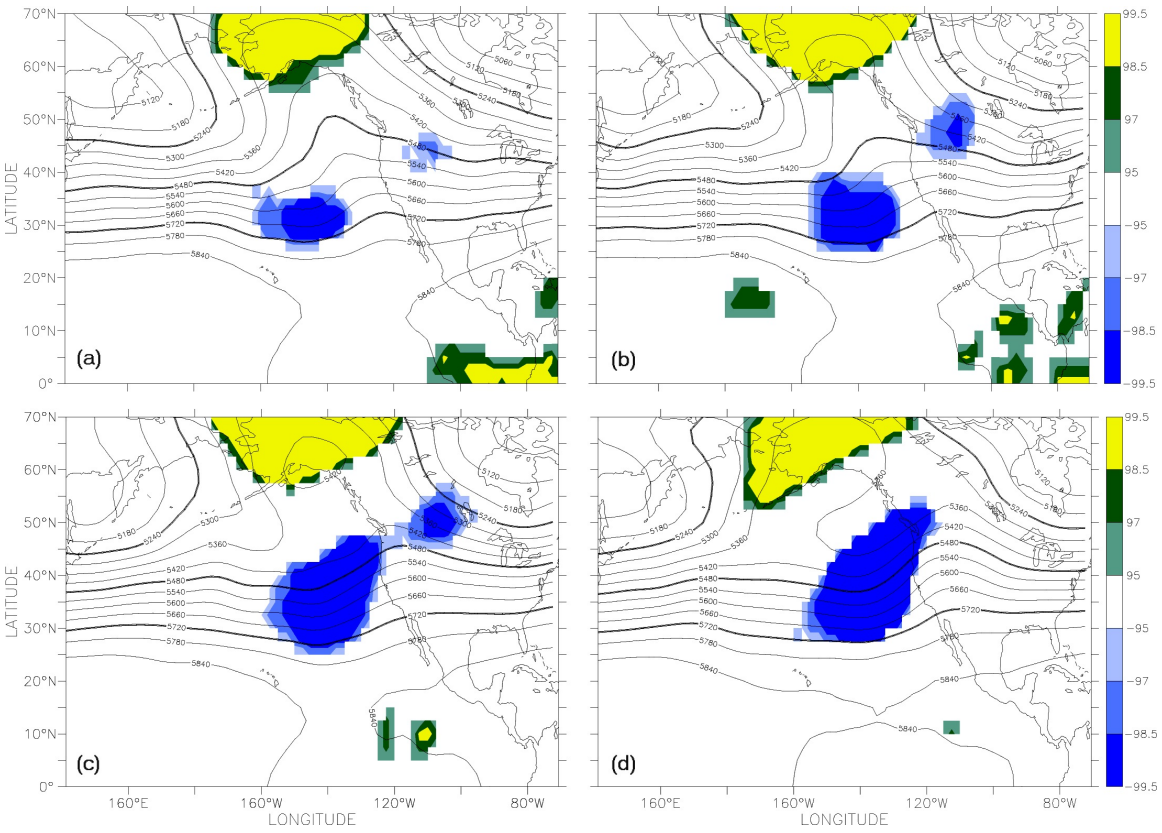

Fig. 5. Similar to Fig. 2 except showing 500 hPa geopotential pattern typical of the heaviest prolonged rain target group. Contours use 60 m interval for the target group ensemble mean. Composite maps shown are: (a) 72 hours prior to the onset of the heavy rain, (b) 48 hours prior, (c) 24 hours prior, and (d) at the time of event onset.

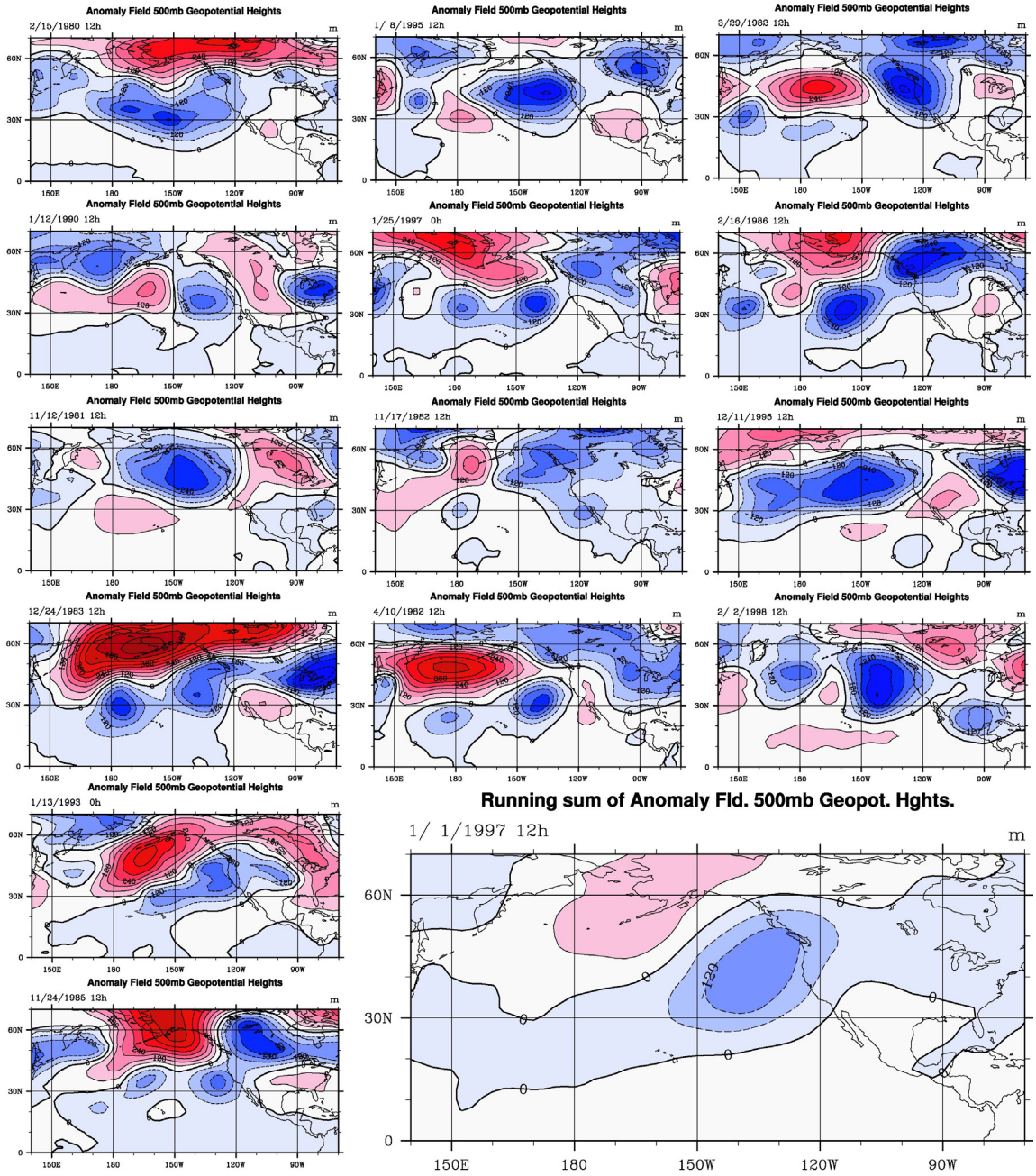

Fig. 6. Similar to Fig. 4 except showing 500 hPa geopotential height anomaly patterns of each member of the heaviest prolonged rain target group at the time of heavy rain onset. This field is not a good indicator for the heaviest prolonged rain because there is much variation between these plots. A forecaster would need to combine this field with other information.

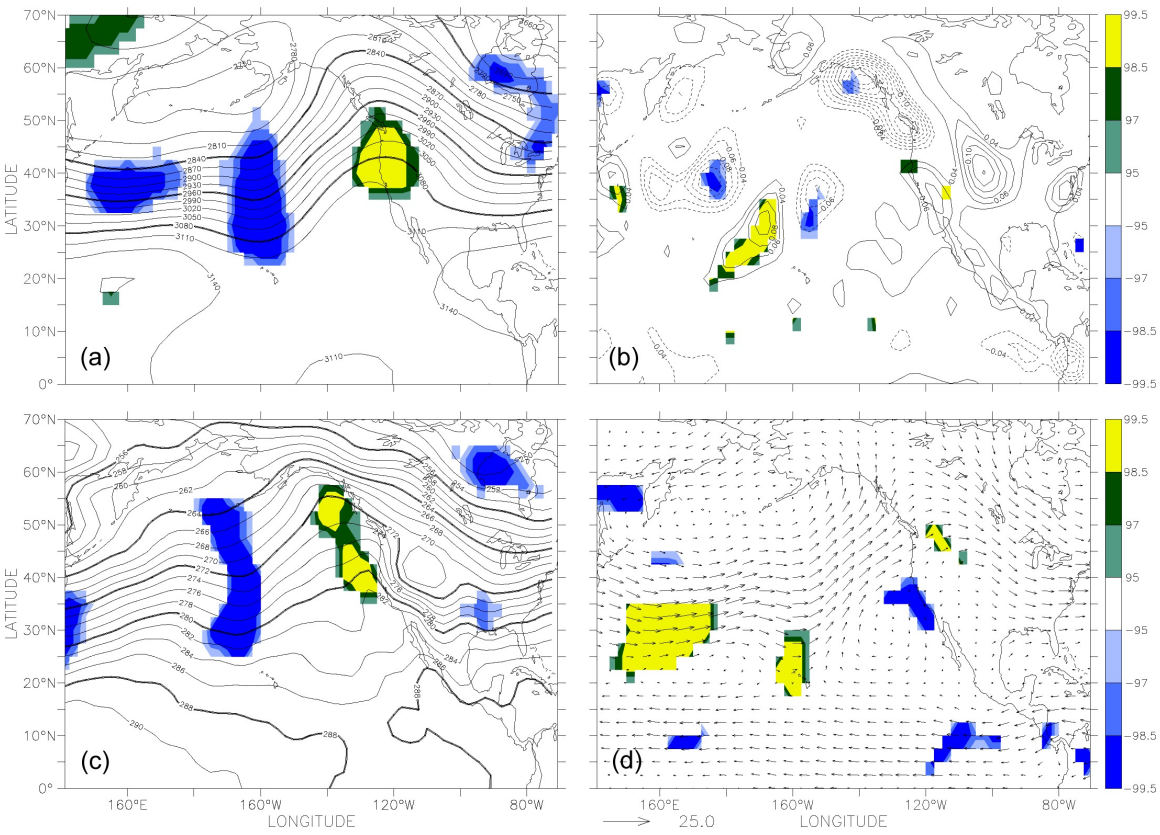

Fig. 7. Similar to prior diagrams except: showing various composite maps of the target group only at the onset of longest duration fog events. (a) 700 hPa geopotential height with 30 m contour interval , (b) pressure (vertical) velocity at 850 hPa with 4 Pa/s contour interval. Negative values (rising motion) have dashed contours while sinking motion (positive values) use solid contours. (c) 850 hPa temperature with 2 K interval, (d) 850 hPa horizontal wind vectors with scale in m/s below the plot and shading denoting significant areas of zonal wind component.

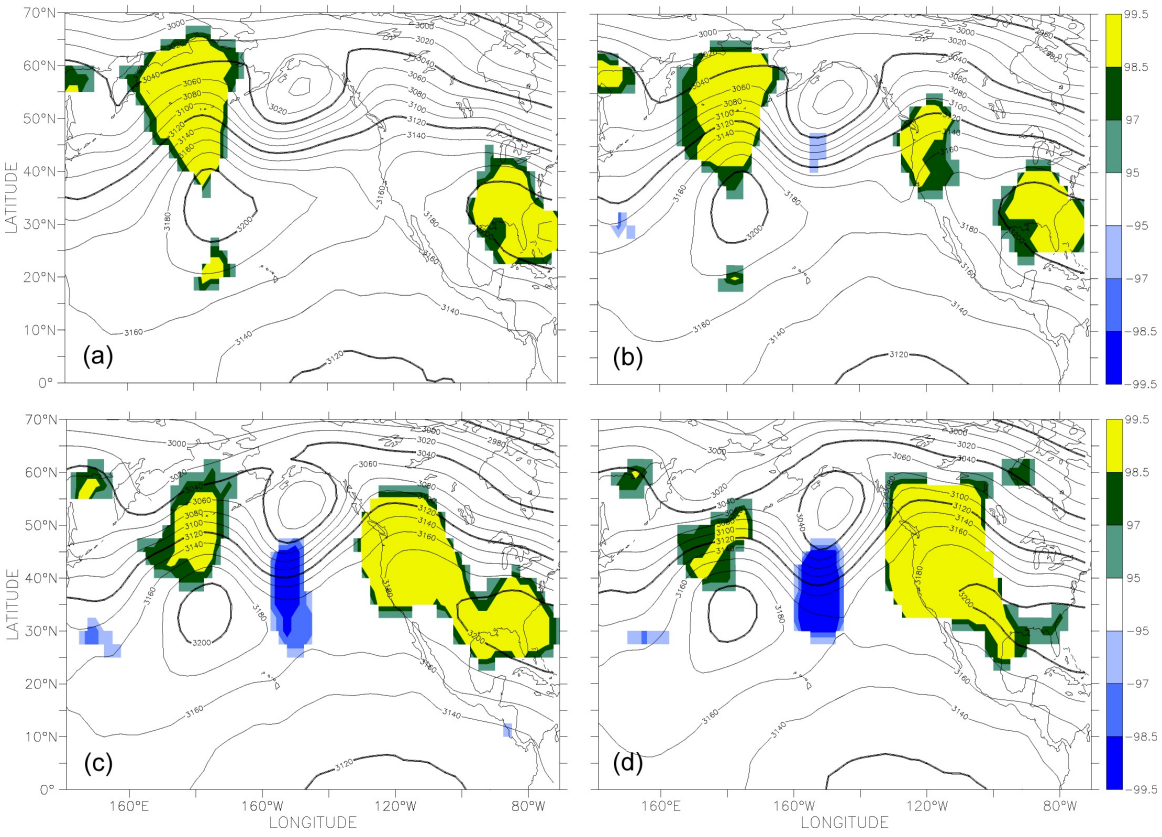

Fig. 8. Similar to Fig. 2 except showing 700 hPa geopotential height pattern typical of the hottest heat waves target group. Contours use 30 m interval for the target group ensemble mean. Composite maps shown are: (a) 36 hours prior to the onset of the heat wave, (b) 24 hours prior, (c) 12 hours prior, and (d) at the time of heat wave onset.

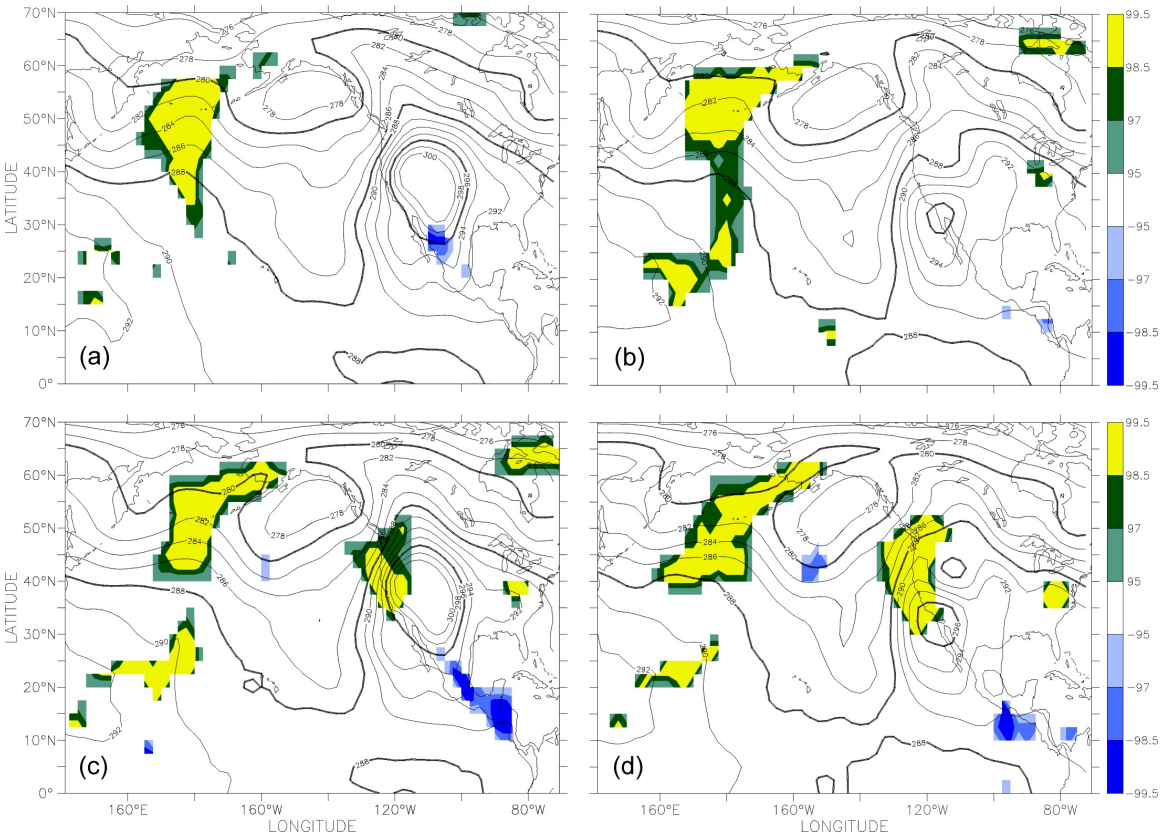

Fig. 9. Similar to Fig. 3 except showing temperature pattern at 850 hPa typical of the hottest heat waves target group. Contours use 2 K interval for the target group ensemble mean. Composite maps shown are: (a) 36 hours prior to the onset of the heat wave, (b) 24 hours prior, (c) 12 hours prior, and (d) at the time of heat wave onset.

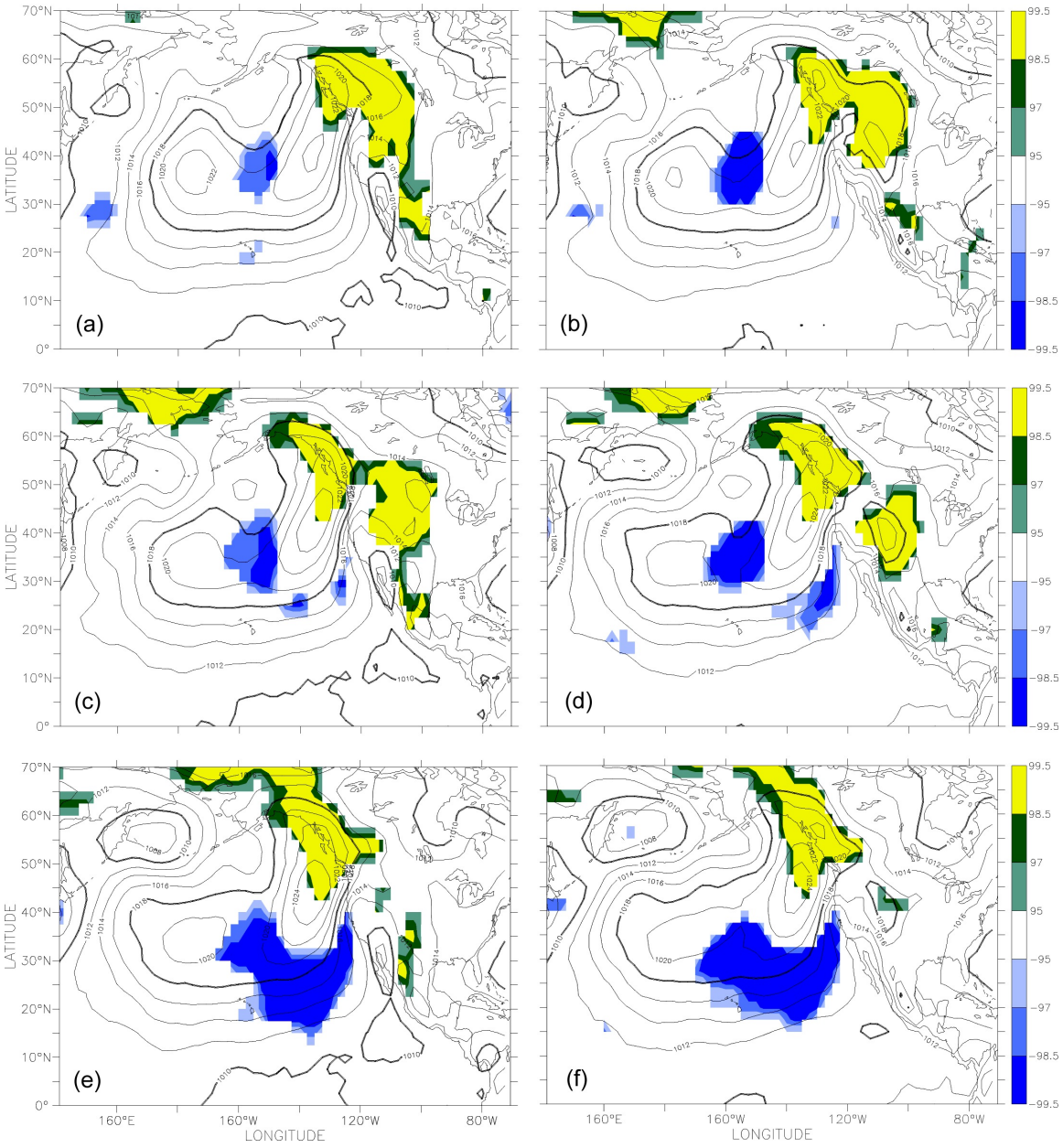

Fig. 10. Similar to Fig. 2 except showing sea level pressure (SLP) pattern typical of the *end* of hottest heat waves target group. Contours use 2 hPa interval for the target group ensemble mean. Composite maps shown are: (a) 60 hours prior to the end of the heat wave, (b) 48 hours prior, (c) 36 hours prior, (d) 24 hours prior, (e) 12 hours prior, and (f) at the end of heat wave.

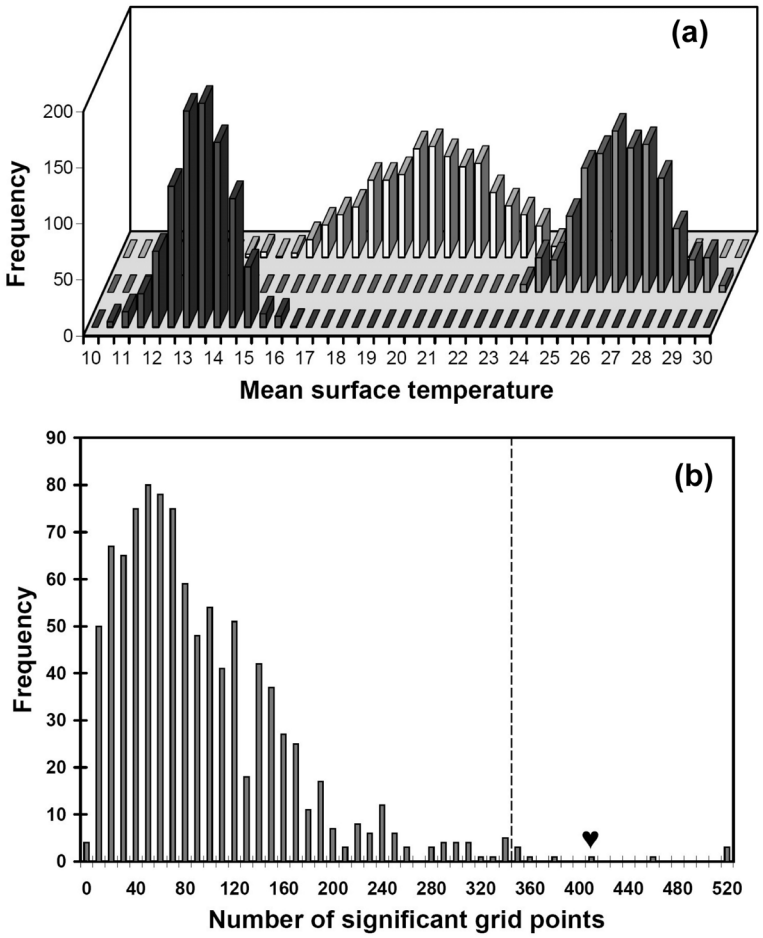

Fig. A1. (a) Three null distributions generated while assessing the significance of the surface mean temperature for hottest heat wave events. These histograms use the grid point closest to Sacramento and incorporate 1000 random samples. The background one is made from both 12 Z and 0 Z values. The two others contain only 12 Z (foreground one) or 0 Z (middle one) values. (b) Null distribution generated while assessing the global significance of an 850 hPa mean geopotential height map for hottest heat waves. This histogram counts the number of gridpoints significant at the 5% level in each of 1000 random maps. The value for the target group map has been added and is shown by a heart. 99% of the values stand to the left of the dashed line.