Ability of CCSM4 to Simulate California Extreme Heat Conditions from Evaluating Simulations of the Associated Large Scale Upper Air Pattern

Richard Grotjahn

Department of Land, Air and Water Resources, University of California, Davis, CA, USA

Submitted to Climate Dynamics 16 May 2012,

Revised 19 September 2012

Re-revised 31 December 2012

Accepted 7 January 2013

Corresponding author address: Richard Grotjahn, Department of Land, Air and Water Resources, University of California, Davis, 95616 USA. Tel: +1-530-7522246, Fax: +1-530-7521793, Email: grotjahn@ucdavis.edu

DOI: 10.1007/s00382-013-1668-1

Abstract

This study assesses how well the Community Climate System Model version 4 (CCSM4) simulates the large scale conditions needed for extreme hot surface temperatures in the California Central Valley (CV). Extreme hot summer days in the CV are associated with a large scale meteorological pattern (LSMP) described in Grotjahn and Faure (2008). The strength and sign of that pattern are assessed using a circulation index developed in Grotjahn (2011). The circulation index is strongly linked to daily maximum surface temperature normalized anomalies at stations spanning the CV. (Extreme heat events in the CV also affect a wider area of California and United States west coast.)

This study makes two primary points. First, the approach used in Grotjahn (2011) can be applied as a novel tool to evaluate how skillfully a model simulates conditions present during CV hot spells by evaluating how well the LSMP is simulated. The circulation index is calculated from historical simulations by CCSM4 and its distribution compared with that of the observed circulation index. Second, values of the CCSM4 based circulation index have smaller standard deviation than observed in reanalysis data. CCSM4 can generate a few comparably large circulation index values (implying high surface temperature anomalies) but not as often as in reanalysis data. Correct simulation of this large scale pattern is a necessary condition for successful simulation of California extreme hot days by a regional climate model. Also, the CCSM4 topography does not have a CV, but a broad topographic slope instead. Various choices of CCSM4 grid points were tested and none satisfactorily represented the CV maximum temperatures. These results should discourage use of CCSM4 surface data directly but encourage use of a regional climate model driven by the CCSM4 to capture hot spells in the CV.

Keywords Extreme hot days simulation, Downscaling hot spells, Surface maximum temperature, CCSM4 global climate model, California Central Valley hot spells weather pattern

1 Introduction

How well does a global climate model reproduce the large scale meteorological pattern associated with extreme hot days of the California Central Valley (CV)? The CV is the most agriculturally productive region in the world and home to 5 million people, hence extreme heat is an important concern. Climate models have been used to estimate how maximum near surface temperatures are likely to change in the future (e.g. Meehl and Tebaldi, 2004) sometimes using indices (e.g. Sillman and Roeckner, 2008). However, this report will show that in the case of surface temperatures over California, using the global model's surface temperatures directly is problematic in part because the model cannot resolve the complex topography of the region. Complex topography can be resolved by dynamically downscaling (e.g. Leung et al., 2004; Kanamitsu and Kanamaru, 2007; Zhao et al., 2011) with a high resolution regional climate model, or RCM. However, the ability of the RCM to simulate extreme events is limited in part by the large scale environment provided by the global model through the RCM boundary conditions. Extreme hot spells in California are associated with highly significant, large amplitude, large scale meteorological patterns (LSMPs) in several variables (Grotjahn and Faure, 2008). These large scale patterns are highly correlated with the extreme hot spells (Grotjahn, 2011) and therefore the LSMP structure resolved by the global model can be a proxy for hot spells in California. This paper applies the scheme developed by Grotjahn (2011) as a tool to interrogate the output from and evaluate hot spells in a particular climate model.

Synoptic patterns are included in some studies of heat waves. Cassou et al. (2005) associate two 'regimes' of summer circulations associated with heat waves over Europe. Grotjahn and Faure (2008) discuss the large scale pattern in several variables and show how it evolves over a few days prior to onset of CV heat waves. Pezza et al. (2012) also look at the synoptic evolution over a few days prior to heat waves affecting southern Australia. Garcia-Herrera et al. (2010; their figure 4) show that within the extraordinarily hot European summer of 2003, the most extreme event (early August) was characterized by a northeastward progression of anomalous heat over a shorter time scale. Stefanon et al. (2012) use a cluster analysis to identify six patterns of 500 hPa geopotential height over the European region associated with heat waves. In all cases the circulation has a mid-tropospheric ridge collocated with the region of extreme heat (an obvious association from hypsometric reasoning). The details of the larger pattern are more interesting, such as the breadth of the ridge, how the pattern evolves over time, and adjacent troughs; the latter can indicate a shift of the midlatitude storm track. Properties such as enhanced sinking (Black et al. 2004, Grotjahn 2011) contribute to amplified surface temperatures. Vautard et al. (2007) associate European heat waves with episodes of large scale southerly surface winds as do Vavrus and Van Dorn for Chicago high temperatures. Grotjahn (2011) associates CV hot spells with an offshore flow that opposes cooling sea breezes. A few studies consider the large scale flow simulated by a global model during heat waves. Gershunov and Guirguis (2012) examine heat waves over California by comparing the ability of four climate models to simulate the large scale sea level pressure (SLP) over western North America and adjacent Pacific Ocean.

The LSMPs focused on here are quite different from well-known low frequency patterns such as the North Atlantic Oscillation (NAO) and Madden-Julian Oscillation (MJO). These low frequency phenomena have some connection to heat waves (e.g. Black and Sutton 2007; Cassou 2008 both studies are for European heat waves). This study does not consider such low frequency phenomena. Such low frequency phenomena may create an environment or ‘envelope’ that may reinforce the short time scale LSMPs emphasized here.

Soil moisture affects heat wave intensity (e.g. Zampieri et al., 2009). Koster et al. (2004) identify regions where soil moisture has high variability coupled to precipitation; in such regions the potential exists for drying of the soil by a preceding drought that can in turn, strongly amplify a heat wave (e.g.

Stefanon et al. 2012; and references therein) by reducing evaporation in the surface energy balance. The CV is not such a region because there is an annual drought occurring during most of the warm season and all of the summer months. Interestingly, Stephanon et al. find that some of their heat wave clusters are not preceded by drought.

The ability of the model surface temperatures to reproduce hot spells is highly limited by the coarse topography present in the model (which cannot capture the CV, note Figure 1). This topographic difference will be explored by using several combinations of points to represent the observed CV, no combination being fully successful. The wind pattern affecting the surface temperature is complex in the CV and those winds are quite different over the valley than over the adjacent western slope of the Sierra Nevada mountains. Over the CV some flows: are channeled and even organized by the topography into low level jets, have a significant meridional component, and have strong sinking nearly to the surface. Over the adjacent mountains the winds: are strongly upslope during the day within a ~1km thick boundary layer (with downslope above) and have a stronger diurnal variation. Such differences can be seen in Zhao et al. (2011) and since the low level patterns over Sierra slopes and CV are different, a bias is introduced by the model’s topography.

Other factors influence the ability of the model to simulate hot spells. For example, the blocking ridge associated with the extreme heat may be influenced by remote, low frequency processes such as ENSO and MJO. The climate model considered has some skill in reproducing the MJO (Subramanian et al., 2011) and ENSO (Deser et al., 2011). However, the events emphasized here have a much shorter time scale.

One might speculate about indirect impacts of biases. For example, a model having too little simulated cloud cover may develop clear skies even during weak subsidence and possibly have reduced soil moisture (than observed) causing a feedback towards positive surface temperature bias. The model studied here tends to under estimate total cloud cover (mainly as low clouds) over the western US during JJA when compared with observational analyses: International Satellite Cloud Climatology Project (ISCCP) D2 and Warren et al. (1986) data (-15 to -30% model bias). However, all those datasets mix together mountainous regions with the CV. The CV has very little cloud cover during summer and the model has little cloud cover over most of the western US. Accordingly, 2m 'surface' temperatures over the CV compare well with the Legates and Willmott (1990) JJA climatology (model bias between -1K to +2K).

2 Data and Procedures

2.1 Data

Circulation indices (as defined in Grotjahn, 2011) are calculated for output from a global climate model simulation of historical climate and compared with corresponding observation-based reanalysis data. The circulation index uses upper air data, specifically temperature at 850 hPa (T850) and meridional wind component at 700 hPa (v700). These upper air data are from the NCEP/DOE AMIP-II (Kanamitsu et al., 2002) gridded reanalyses of daily data interpolated to 2.5 by 2.5 (latitude by longitude) resolution (hereafter NDRA2 data). The global climate model was output by the Community Climate System Model version 4 (CCSM4) described in Gent et al. (2011). The CCSM4 data are from a historical simulation using 1.1 degree finite volume resolution; the upper air data are also interpolated to the same coarser grid. As in Grotjahn (2011) all the upper air data are at 12 GMT, approximately 12 hours before the maximum temperature is typically reached in the CV. The period is all June-September days during 1979-98. Daily maximum surface data (at 2 m above ground) are also used. Surface observations were studied from 3 CV stations: Red Bluff (KRBL), Fresno (KFAT), and Bakersfield (KBFL). Daily maximum 2m temperature values from CCSM4 grid points were also used at full model resolution.

2.2 Procedures

A brief summary of the calculation of the circulation index follows; the interested reader can find details of the procedures in Grotjahn (2011). Anomalies are formed of all data by subtracting the corresponding long term daily mean values. For upper air data, the mean and anomaly are formed from the instantaneous data; long term daily means are calculated separately for time of day, day of the year, and each grid point. The daily maximum temperature surface data are further normalized by the corresponding standard deviation, the result is referred to as 'maxTa'. The normalization makes the maxTa values of different stations (or model grid points) intercomparable and facilitates averaging station data with different variability. Such normalization is used by others, for example, Cattiaux et al. (2012). For each day an unnormalized projection of a portion of each upper air anomaly field onto the corresponding portion of each target ensemble anomaly field is calculated. In Grotjahn (2011) the circulation index defined as a 71% and 29% combination of the projections onto the T850 and v700 ensemble mean fields, respectively worked best. The same ratio is used here. The target ensemble for a given field is the average of that field in NDRA2 data during the 1979-88 training period using the 16 dates when the 3 CV stations have their highest combined maxTa values. This target ensemble is used for both NDRA2 and CCSM4 daily data because the portions of patterns used are similar to blocking ridges. A blocking index based on the simulation of such ridges is notably improved by removing the mean model bias first (Scaife et al., 2010; 500 hPa geopotential heights during winter). However, the grid points used in the projection are mainly off shore and located where the model bias is quite small; hence these summer CCSM4 biases are neglected when calculating the circulation indices shown. Circulation indices in the aggregate are compared between NDRA2 and CCSM4 data and with observed surface maximum temperatures. Extreme value statistics and other tests are used to focus upon properties of the high tail of the distribution of temperature and circulation indices.

The normalized anomalies of CCSM4 surface daily maximum temperatures (maxTa-CCSM) are calculated to see what large scale upper air patterns in the model occur for high values of maxTa-CCSM. Various combinations of CCSM4 surface grid points are explored.

3 Results

3.1 Simulated maximum surface temperatures

The global model cannot resolve the CV making any use of the model's surface temperatures as well as some lower atmosphere temperatures problematic. Though geographically important, Figure 1 shows that the CV is not present with the approximate 1.1 degree model resolution. Where the CV should be, CCSM4 has a broad topographic slope. The CV station elevations (106m at KRBL; 101m at KFAT; 149m at KBFL) are much lower than the CCSM4 elevations at the corresponding locations; at the 3 'CCSM4-CV-locs' locations in Figure 1, the elevations range from about 740m to 860m. The higher elevations make the model surface much closer to the elevations of the LSMPs used and thus the relation between surface and circulation index is artificially improved. The model's topographic slope is also problematic since afternoon heating causes near-surface vertical motions to have opposite sign and reduced potential temperatures (e.g. figure 14a in Zhao et al., 2011) over the topographic slope of the Sierra Nevada mountains compared to the relatively flat CV. Grotjahn (2011) discusses how a low, strong subsidence inversion is associated with extreme hot days. Various lower elevation grid points were tested, indicated by numbers in Figure 1. The idea is that grid points away from the coast but over a similar range of elevation and latitude as the CV might behave more similarly to the temperatures observed within the CV than would points on the broad topographic slope. But, even these-mock-CV grid points are not particularly similar to CV observations.

Maximum temperature values at grid points are compared with CV observations. The median, standard deviation, and range of maximum temperature values from the averaged three CV station observations are 308.0K, 4.38C, and 25.0C, respectively. Daily maximum temperatures at grid point 0 in Figure 1 have too low of: median, standard deviation, and range (288.5K, 1.3C, 10.6C) as do values at grid point 4 (291.3K, 1.4C, 10.2C). In contrast, values at grid point 2 have the best median, standard deviation, and range (299.7K, 3.8C, 23.0C). Grid point 1 has the highest correlation with grid point 2 and third highest standard deviation and range (293.4K, 2.4C, 16.1C). While grid point 3 values have second highest correlation to grid point 2 values, the range is small (13.4C) and it was judged less representative of the CV latitude range than using either points 5 or 6 which are further south. Grid point 6 values are less correlated (0.57) with grid point 2 values than are grid point 5 values(0.72); grid point 5 values also have higher median, standard deviation, and range (295.6K, 3.0C, 19.4C). With this reasoning, grid points 1, 2, and 5 in the CCSM4 data were selected as roughly representative of the CV in the model. The average of values for these grid points will be used and referred to as the 'CCSM4-mock-CV '.

Figure 2 shows three histograms of non-normalized temperatures averaged for various groups of points over the 2440 days of June-September, 1979-98. The observed values are the average of KRBL, KFAT, and KBFL daily maximum temperatures. The middle curve averages the surface maximum temperatures for 3 grid points close to the geographic locations of the 3 CV stations. The remaining curve averages the daily maximum temperatures for the CCSM4-mock-CV. It is obvious that the CCSM4-mock-CV values are systematically low in this region and time of year. The median and mean of the CCSM4-mock-CV are about 11-12C cooler than the observed values. The points further inland, roughly on the middle of CCSM4's topographic slope do much better, only about 3C too cool. More intriguing are the standard deviation, skew, and kurtosis. The observed CV stations have negative skew (-0.51; broader tail below the median than above the median). The CCSM4 values at grid points on the topographic slope (referred to as 'CCSM4-CV-locs' in Figure 1) also have negative skew (-0.47). However, the CCSM4-mock-CV values have positive skew (0.22). The kurtosis of the non-normalized data are: 0.032, -0.069, -0.160 for the observed CV stations, CCSM4-CV-locs, and CCSM4-mock-CV averages, respectively. However, averaging non-normalized data can be misleading. The normalized anomaly averages used later (e.g. Figure 4) have skew (kurtosis) for the observed three stations, CCSM4-CV-locs, and CCSM4-mock-CV of: -0.31 (-0.25), -0.27 (0.61), 0.49 (0.05), respectively. So depending on whether normalized anomalies are used or not, the locations chosen have skew or kurtosis that are more or less similar to the observed, but generally the coastal stations of CCSM4-mock-CV perform less well than grid points on the topographic slope. The standard deviations for the 3 combinations compares similarly. CV observations, CCSM4 grid points on the slope, and CCSM4-mock-CV have standard deviations: 4.38C, 4.70C, and 2.72C for non-normalized data and: 0.94, 0.70, and 0.81 for normalized anomaly averages. Why the grid points behave so differently on the topographic slope than on the narrow coastal plain is outside the scope of this study. It is speculated that the high elevations of the CCSM4 grid points on the topographic slope are more responsive to the mid troposphere than are the CCSM4-mock-CV points, as judged from scatter plots of these values against CCSM4 circulation index.

The robustness of the link between upper air fields and CCSM4-mock-CV extreme values is examined even though the CCSM4-mock-CV points have less association with the upper air pattern than using the CCSM4 grid points on the slope. Figure 3 compares the input NDRA2 fields used for the circulation indices against the corresponding average anomaly fields on the dates of highest 1% of maxTa-CCSM4 at the CCSM4-mock-CV grid points. The general patterns are similar but have three notable differences. First, the amplitude is considerably stronger in the NDRA2 input fields. Second, while both patterns have a ridge over the west coast and troughs upstream and downstream, the zonal wavelength is shorter. Third, the peak anomaly values of T850 are centered onshore using the CCSM4-mock-CV dates but offshore in observations. Grotjahn (2011) emphasizes that the location of T850 anomaly offshore is crucial for creating a SLP gradient that suppresses a cooling sea breeze.

3.2 Circulation index scatter plots

Each panel in Figure 4 compares a circulation index with a corresponding surface maximum temperature anomaly. Figure 4a uses data described in Grotjahn (2011): circulation index defined from NDRA2 data and observed maxTa averaged for the three CV stations, but only for 20 years (1979-98). While there are ‘busts’ in the sense that the observed stations can have a large positive anomaly when the circulation index does not (and vice versa) about half the extreme values (highest 1%) are also the highest 1% of the index values. The scatter using observations can be improved by altering the circulation index calculated in Grotjahn (2011) in several ways. A simple improvement is to include T850 values at 0GMT for grid points over the CV in observations (not shown). Making that improvement to the CCSM4 circulation index narrows the scatter dramatically, especially on the high end; but such is not shown because the relationship is artificial since the grid point elevations are much closer (about half the distance) to the 850 hPa elevation than occurs in the CV; so the independence between circulation index component and maxTa-CCSM is greatly reduced.

Figure 4b compares the circulation index calculated using the NDRA2 target anomaly fields projected onto the daily CCSM4 data (the ‘CCSM4 circulation index’) versus the maxTa-CCSM4 averaged for the 3 grid points near the three CV stations. While the relationship is not as good as in observations (Figure 4a) there is still a tendency for many of the extreme maxTa-CCSM4 values (on the broad topographic slope) to be coincident with extreme circulation index values. The relationship is noticeably reduced (i.e. the scatter is broader) when the CCSM4 circulation index is matched against the maxTa from the CCSM4-mock-CV grid points (Figure 4c).

An appropriate measure of the scatter for points above a threshold is the non-parametric Kendall tau rank correlation coefficient (τ). Values were calculated following Wessa (2012). For the 385 values above 1.0 of 3-station CV surface data, τ=0.33. For the 312 values above 0.7 of CCSM4 surface data at grid points near the CV stations, τ=0.15. For the 368 values above 0.8 of CCSM4-mock-CV surface data, τ=0.10.

3.3 Circulation indices distributions

The distributions of the observed maxTa and the NDRA2 and CCSM4 circulation indices on all 2440 days are shown in Figure 5a. While the data at individual grid points and individual CV stations have been normalized, daily averages formed from such data do not necessarily have standard deviation of 1.0. Also, the intent of the circulation index is to capture properties of the high tail, discussed more fully in connection with Figure 5b. Nonetheless, the observed circulation index (dashed line with diamonds) has similar standard deviation, skew, and other properties as the observed surface maxTa for the three CV stations (solid line with squares). The circulation index performs well for the low tail as well as the high because unusually cool days for the CV are associated with a trough centered near the coast which creates an opposite sign circulation index relative to unusually hot days. Hence, the circulation index captures the whole range of surface variation (further discussion in Grotjahn, 2011). The standard deviations for maxTa and NDRA2 circulation index match well (0.94, 0.91); the skew is negative for the maxTa surface values (-0.26) but near zero for the NDRA2 circulation index (0.02). Somewhat in contrast, the CCSM4 circulation index distribution (dotted line with triangles) has much smaller standard deviation (0.75) and positive skew (0.15). The main implication from Figure 5a is that the CCSM4 historical simulation does not have as much variability in the circulation index as found in the reanalysis data.

The smaller standard deviation in the CCSM4 circulation indices results in CCSM4 counts in bins between 1.0 and 2.0 being roughly 2/3 the counts for the NDRA2 indices. The extreme high end of the distribution of the circulation index is emphasized in Figure 5b. The circulation index is intended to capture the top 1% of the events. So, using a cutoff corresponding to the top 1% of the NDRA2 circulation indices (24 dates) results in about half as many (9 occurrences) of the CCSM4 indices. While there are fewer of the CCSM4 indices on the high end of the tail, it is noteworthy that the CCSM4 does produce as large of values of the circulation index as occurred in the NDRA2 data.

Generalized Pareto distribution (GPD) fits were made to the high tail of the two sets of circulation indices. The purpose in using this well established tool is to have a quantitative means to compare the high tail of the distribution of maximum temperature and circulation indices in the model versus observations. Two tests were made to identify a reasonable threshold for the fit. Mean residual life plots of the maxTa, NDRA2, and CCSM4 circulation indices were ‘linear’ in different ranges but those ranges overlap from about 0.5 to 1.4 for the three distributions. Testing GPD properties for different thresholds finds consistent estimates of scale and shape parameters between 0.5 to about 1.1 for all three distributions. On these bases, GPD fits used 1.0 as the threshold. The shape, scale, and 100 year return estimates based on these GPD fits are given in Table 1. The scale parameter is inversely related to the amplitude. The shape parameter is an indicator of how long the tail is and Table 1 is consistent with comments made above about Figure 5. Negative values of the shape parameter tend to ‘straighten out’ the distribution resulting in a ‘zero crossing’ (which implies an upper bound to the circulation index); the longer tail of the CCSM4 index shows up as the smaller (though still negative) magnitude of the shape parameter. These results show that the general properties of the distribution discussed above (smaller standard deviation coupled with a long tail in the CCSM4 circulation indices) are less prominent at the high tail. Though the tails of the distributions shown in Figure 5 visually look different and despite notable differences in standard deviation, skew, and kurtosis, the longer tail in the CCSM4 indices makes the 100 year return period value quite similar in the distributions of the two circulation indices. However, in short, the CCSM4 tends to have too few of the larger circulation indices.

Heat waves definitions (see Grotjahn, 2011) often include a minimum duration of consecutive extremely hot days. Figure 6 organizes the dates above the 1.0 threshold used for the GPD fits into bins based on the number of consecutive days above that threshold. The longest duration is 15 days which occurred once each for the surface maxTa observations and for the CCSM4 circulation index. The longest duration (10 days) for the NDRA2 data occurred twice. Obviously, longer durations are less common than shorter durations above the threshold. The decrease in the number of events as duration increases looks similar for all three distributions.

There is no single theoretical statistical distribution for the durations above a threshold. However, the geometric distribution (i.e., the discrete analog of the exponential distribution) is appropriate (Katz, personal communication). The key parameter for a geometric distribution fit to durations above a threshold (Furrer et al, 2010) is (the inverse of) the average duration length. The average duration lengths for the three circulation indices are included on figure 6. The average durations compare favorably even though the same threshold is used for the station data (maximum surface temperature anomaly) as for the gridded data (upper air circulation indices). Also, while the CCSM4 data have fewer exceedances of the threshold than the reanalysis data, the average duration is remarkably similar for these two gridded datasets.

4 Conclusions

This project considers how well extreme hot spells over the Central Valley (CV) of California are simulated by the CCSM4 global climate model.

The model cannot resolve the CV but instead has a broad topographic slope where the low-lying and nearly flat CV should be located. Several choices of surface grid points were considered in an effort to capture the maximum temperature behavior observed in the CV. CCSM4 grid points located near the CV stations were found to have a distribution of daily maximum temperature that is a few degrees too cool, but otherwise the distribution agrees well in standard deviation and skew with the observed daily maximum temperatures (Figure 2). One concern with choosing these model points to represent the CV is that the behavior of winds and thermodynamic properties will likely be quite different for a topographic slope open to the Pacific Ocean than for a CV largely ringed by mountains. Significant differences between the near-surface properties over the CV and the western slope of the Sierra Nevada mountains have been found in observations and model studies (e.g. Zhao et al., 2011). In addition, the high altitude of these model grid points artificially boosts the link between surface values and lower tropospheric variables (e.g. at the 850 hPa level). Attempts to use lower elevation model grid points inland from the coast of California proved to be disappointing. The best lower elevation grid points were much cooler, had notably smaller standard deviation, and opposite sign for skew, compared to CV observations. Though they are mediocre choices for representing the CV with CCSM4, the highest 1% of the surface maximum temperature anomalies at these points defined the days used to construct ensemble means of temperature at 850 hPa and meridional wind at 700 hPa; these CCSM4 ensemble means were similar to those based on CV observations and reanalysis data. It was judged that the CCSM4 develops a generally similar pattern because the thermal maximum and accompanying ridge are consistent with simple thermodynamic arguments. The primary differences in the ensemble means

(Figure 3) based on observed versus coastal CCSM4 grid points are that the CCSM4 ensemble mean had weaker amplitude, smaller zonal scale, and the maximum thermal anomaly was onshore instead of offshore. The location offshore of the 850 hPa level thermal anomaly is an important detail for extreme heat in the CV (see Grotjahn, 2011). The weaker CCSM4 LSMP amplitude is consistent with calculations that found fewer larger values of the CCSM4 circulation indices.

An earlier study (Grotjahn, 2011) used a projection of key parts of the ensemble mean fields upon corresponding daily weather maps to calculate a circulation index. The circulation index is intended to measure how strongly a given day resembles a day of observed extremely hot maximum temperature anomaly in the CV. This study applies that projection upon daily data from CCSM4 historical simulations. The CCSM4 circulation index has notably smaller variation and different tail properties compared with the observed (NCEP/DOE AMIP-II reanalysis) data (Figure 5). For the extreme positive values of the index (associated with the hottest days) CCSM4 can create large values, but they don't occur often enough (generally only two thirds as often). The large scale pattern for cool days has locally the opposite anomaly pattern, hence the circulation index matches well the cool anomaly maximum temperatures even though that was not an intended use. The CCSM4 is less adept at generating the negative extreme values of the circulation index associated with unusually cool anomalies than at generating positive extreme index values. While CCSM4 generates fewer positive extremes above a threshold of one standard deviation than are found in observations, those in the model occur for varying durations that do match well the observed durations (Figure 6).

This study does not assess how general such climate model errors are, though a few other studies do consider aspects of the larger meteorological pattern. A general impression is that atmosphere-ocean general circulation models do a fairly good job in simulating the statistics of extreme temperature events (Randall et al., 2007; and references therein). For example, Kharin et al. (2005)

compare maximum temperature extremes for a dozen models with various reanalyses, including an earlier version of the NCAR model (the latter simulates the statistics well). If that simulation skill presumes adequate simulation of the associated LSMPs, then this work would rightly place emphasis upon the topography and other phenomena not resolved by the global model. Meehl and Tebaldi (2004) find similar 500 hPa geopoential height patterns from a global model (the PCM) ensemble mean during the worst 3-day (nighttime minimum temperature) during 1961-90, and from observations during a severe heat wave near Chicago, USA. Vavrus and Van Dorn (2010) compare SLP patterns from two models and observations for the hottest 95th percentile days at Chicago. Simulations and observations find strong southerly advection towards Chicago with a strong 500 hPa ridge aloft. Gershunov and Guirguis (2012) compare the ability of four combinations of global and regional climate models to simulate the large scale SLP over western North America and adjacent Pacific Ocean during heat waves. They emphasize a particular model, but even that model has a tendency for weaker and fewer heat waves (using a 95th percentile threshold). Diffenbaugh and Ashfaq (2010) compare observations with regional climate model simulations forced by boundary conditions from CCSM3 (1951-1999) finding a similar pattern for the hottest summer, though the model had less variance than observed over the western US. Ashfaq et al. (2010) compare 95th percentile temperature thresholds from two models (using a regional model) over the continental United States. They use a regional model driven by a NASA global model and find the patterns compared well with NARR (North American Regional Reanalysis) values. Mastrandrea et al. (2011) compare some surface-based extreme indices over California in observations and 6 global climate models (GCMs) downscaled two ways; since the downscaling includes bias correction, analogs, and model averaging, it is unclear how well the individual GCMs perform, though the downscaled results for mean heat wave duration compare favorably with observations Finally, some studies focus on a much longer time scale than considered here. Seasonal hot spells do not consider the LSMPs emphasized here but instead lead to associations with low frequency phenomena like NAO, ENSO, or MJO. For example, Trenberth and Fasullo (2012) discuss several extreme events of 2010, comment on the models ability to simulate possibly-related remote phenomena, and employ CCSM4 in their analysis. Interested readers are directed to that paper and its references for further discussion of links to low frequency phenomena.

The importance of the circulation index goes beyond being a proxy for extreme hot values in a CV that is not present in the model. These results show that no combination of global model (CCSM4) grid points adequately describes the CV surface temperatures or their association with the upper air LSMP associated with CV hot spells. So, a regional climate model (RCM) with sufficient topographic resolution to have a CV is needed but the RCM needs the right large scale circulation pattern (as boundary conditions) to have a chance of generating extreme CV surface temperature anomalies (hot or cool). An RCM is not likely to overcome boundary conditions from a global model that lack the right large scale circulation patterns. CCSM4 can generate strong enough large scale patterns for CV hot spells and the patterns can last as long as observed thereby encouraging RCM simulations of extreme heat in the CV, but the caution is that CCSM4 does not create the patterns as often as does the atmosphere.

Acknowledgments

The author thanks Gary Strand for identifying the appropriate climate model historical data (b40.20th.track1.1deg.012) and for providing access to that CCSM4 data. These data were accessed from the National Center for Atmospheric Research (NCAR) using the Earth System Grid. The NCEP/DOE AMIP-II Reanalysis (NDRa2) data used in this study were obtained from the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their Web site at http://www.cdc.noaa.gov/. The surface observational data were provided by the University of California (UC) statewide Integrated Pest Management (IPM) program of the UC Division of Agriculture and Natural Resources. Extreme statistics were calculated using the extRemes toolkit. Specifically: Extreme value toolkit for R package version 1.62. http://CRAN.Rproject.org/package=extRemes was developed by Eric Gilleland, Rick Katz and Greg Young (2010). The author is grateful for assistance by Drs. Richard Katz (regarding the Kendall tau and duration analyses) and Gerald Meehl (for comments about the CCSM).

References

Ashfaq M, Bowling L, Cherkauer K, Pal J, Diffenbaugh N (2010) Influence of climate model biases and daily-scale temperature and precipitation events on hydrological impacts assessment: A case study of the United States J. Geophys. Res. 115: D14116 DOI 10.1029/2009JD012965

Black E, Blackburn M, Harrison G, Hoskins B, Methven J (2004) Factors contributing to the summer 2003 European heatwave Weather 59:218–223 doi: 10.1256/wea.74.04

Black E, Sutton R (2007) The influence of oceanic conditions on the hot European summer of 2003 Clim. Dyn. 28: 53–66

Cassou C (2008) Intraseasonal interaction between the Madden–Julian Oscillation and the North Atlantic Oscillation. Nature 455: 523–527

Cassou C, Terray L, Phillips A S (2005) Tropical Atlantic influence on European heatwaves J. Climate 18: 2805–2811

Cattiaux J, Yiou P, Vautard R (2012) Dynamics of future seasonal temperature trends and extremes in Europe: a multi-model analysis from CMIP3. Clim. Dyn. 38: 1949–1964 DOI 10.1007/s00382-011-1211-1

Deser C, Phillips A, Tomas R, Okumura Y, Alexander M, Capotondi A, Scott J, Kwon Y-O, Ohba M (2011) ENSO and Pacific decadal variability in the Community Climate System Model Version 4 J. Climate 25: 2622-2651 DOI: 10.1175/JCLI-D-11-00301.1

Diffenbaugh N, Ashfaq M (2010) Intensification of hot extremes in the United States, Geophys. Res. Lett., 37: L15701, doi:10.1029/2010GL043888.

Furrer E, Katz R, Walter M, Furrer R (2010) Statistical modeling of hot spells and heat waves, Climate Res., 43: 191-205, doi: 10.3354/cr00924.

Garcia-Herrera R, Diaz J, Trigo R, Luterbacher J, Fischer E (2010) A review of the European summer heat wave of 2003, *Crit. Rev. Environ. Sci. and Tech.*, 40: 267-306.

Gent P, and Coauthors (2011) The Community Climate System Model Version 4. *J. Climate* 24: 4973–4991. doi: http://dx.doi.org/10.1175/2011JCLI4083.1

Gershunov A, Guirguis K (2012) California heat waves in the present and future. Geophys. Res. Lett. 39: L18710, doi:10.1029/2012GL052979.

Grotjahn R (2011) Identifying Extreme Hottest Days from Large Scale Upper Air Data: a Pilot Scheme to find California Central Valley Summertime Maximum Surface Temperatures. Clim Dyn 37: 587-604. DOI : 10.1007/s00382-011-0999-z

Grotjahn R, Faure G (2008) Composite predictor maps of extraordinary weather events in the Sacramento California region. Weather Forecast 23: 313–335. doi: http://dx.doi.org/10.1175/2007WAF2006055.1

Kanamitsu M, Ebisuzaki W, Woollen J, Yang SK, Hnilo J, Fiorino M, Potter G (2002) NCEP-DOE AMIP-II Reanalysis (R-2). Bull Am Meteorol Soc 83:1631–1643Koster R D, and Coauthors (2004) Regions of strong coupling between soil moisture and precipitation Science 305: 1138–1140. DOI: 10.1126/science.1100217

Kanamitsu M, Kanamaru H (2007) Fifty-seven-year California reanalysis downscaling at 10 km (CaRD10). Part I: system detail and validation with observations. J Clim 20: 5553–5571. DOI: 10.1175/2007JCLI1482.1

Kharin, V.V., F.W. Zwiers, and X. Zhang, (2005) Intercomparison of near surface temperature and precipitation extremes in AMIP-2 simulations, reanalyses and observations. J. Clim., 18: 5201–5223. doi: http://dx.doi.org/10.1175/JCLI3597.1

Koster RD et al. (2004) Regions of strong coupling between soil moisture and precipitation. Science 305:1138–1140. doi:10.1126/science.1100217

Legates DR, Willmott CJ, (1990) Mean seasonal and spatial variability in global surface air temperature. Theoret. Appl. Clim. 41: 11-21.

Leung LR, Qian Y, Bian X, Washington W, Han J, Roads J (2004) Mid-century ensemble regional climate change scenarios for the western United States Climatic Chg 62: 75-113.

Mastrandrea M, Tebaldi C, Snyder C, Schneider S (2011) Current and future impacts of extreme events in California Climatic Change 109: S43-S70. DOI 10.1007/s10584-011-0311-6

Meehl G, Tebaldi C (2004) More Intense, More Frequent, and Longer Lasting Heat Waves in the 21st Century. *Science* 305: 994-997.

Pezza A, van Rensch P, Cai W (2012) Severe heat waves in Southern Australia: synoptic climatology and large scale connections. Clim Dyn 38: 209-224. DOI 10.1007/s00382-011-1016-2

Randall DA, Wood RA, Bony S, Colman R, Fichefet T, Fyfe J, Kattsov V, Pitman A, Shukla J, Srinivasan J, Stouffer RJ, Sumi A, Taylor KE (2007) Climate models and their evaluation. In: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment

Report of the Intergovernmental Panel on Climate Change [Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL (eds.)]. Cambridge University Press, Cambridge, UK, and New York, NY, pp. 589-662.

Scaife A, Woollings T, Knight J, Martin G, Hinton T (2010) Atmospheric blocking and mean biases in climate models. J. Climate, 23: 6143–6152. doi: http://dx.doi.org/10.1175/2010JCLI3728.1

Sillmann J, Roeckner E (2008) Indices for extreme events in projections of anthropogenic climate change. Climatic Chg 86: 83–104. DOI 10.1007/s10584-007-9308-6

Stefanon M, D'Andrea F, Drobinski P (2012) Heatwave classification over Europe and the Mediterranean region. Environ. Res. Lett. 7: 014023 (9pp) doi:10.1088/1748-9326/7/1/014023

Subramanian A, Jochum M, Miller A, Murtugudde R, Neale R, Waliser D (2011) The Madden-Julian Oscillation in CCSM4 J. Climate 24: 6261-6282. DOI: 10.1175/JCLI-D-11-00031.1

Trenberth K, Fasullo J (2012) Climate extremes and climate change: The Russian heat wave and other climate extremes of 2010. J. Geophys. Res. 117: D17103 DOI: 10.1029/2012JD018020

Vautard R, and Coauthors, (2007) Summertime European heat and drought waves induced by wintertime Mediterranean rainfall deficit. Geophys. Res. Lett., 34: L07711, doi:10.1029/2006GL028001.

Vavrus S, Van Dorn J (2010). Projected future temperature and precipitation extremes in Chicago. J. Great Lakes Res., 36: 22-32 (Supplement 2).

Warren SG, Hahn CJ, London J, Chervin RM, Jenne RL (1986) Global distribution of total cloud cover and cloud type amounts over land. NCAR Tech. Note TN-273 + STR, Boulder, CO, 29 pp. and 200 maps.

Wessa, (2012), Kendall tau Rank Correlation (v1.0.11) in Free Statistics Software (v1.1.23-r7), Office for Research Development and Education, URL http://www.wessa.net/rwasp_kendall.wasp/

Zampieri M, D'Andrea F, Vautard R, Ciais P, de Noblet-Ducoudré N, Yiou P (2009) Hot European Summers and the Role of Soil Moisture in the Propagation of Mediterranean Drought. J. Climate, 22:4747–4758. doi: http://dx.doi.org/10.1175/2009JCLI2568.1

Zhao Z, Chen S-H, Kleeman M, Tyree M, Cayan D (2011) The impact of climate change on air quality–related meteorological conditions in California. Part I: Present time simulation analysis. J. Climate, 24: 3344–3361. doi: http://dx.doi.org/10.1175/2011JCLI3849.1

Figure Captions.

Fig. 1 Surface elevations for a) CCSM4 model and b) actual topography are shown (in meters, starting at 20m, using 60m interval). a) Grid point locations of CCSM4 model data discussed in the text. Geographic locations of California Central Valley (CV) stations used by Grotjahn (2011) in defining the downscaling scheme for extreme hottest days: R (Red Bluff, KRBL), F (Fresno, KFAT), and B (Bakersfield, KBFL). The model grid points used to represent R, F, and B are indicated by ‘+’ symbols . The CV seen in b) is not resolved by the model and the model has notably higher topography at the actual location of the CV. Two grid point combinations: i) numbered 1, 2, and 5 and ii) indicated by + symbols were used (see text) as approximations to R, F, and B. The former are identified as the ‘CCSM-mock-CV points while the latter are ‘CCSM4-CV-locs’ points. Shaded area in a) indicates region above 20m elevation that should be ocean. (The model defines grid points at and near the ocean with both land and ocean fractions.) The actual topography has far steeper slope for the various mountain ranges in the region. Hence, contours range from 20-1640m in a) and 20-3620m in b).

Fig. 2 Daily surface (2 m elevation above ground) maximum temperature comparison histograms; bins are 1.59C wide. ‘Observed’ values (shorter dashed line) are the average values from the three stations (R, F, B) in Fig. 1. ‘CCSM4-CV-locs’ are daily maximum 2 m temperatures (solid line) at three CCSM4 grid points (‘+’ symbols in Fig. 1a) close to the locations of R, F, and B. CCSM4-CV-locs values are generally similar (in values, standard deviation, and skew) to the observations (but ~3C too cool). ‘CCSM4-mock-CV’ values are averages of 2 m daily maximums at grid points 1, 2, and 5 on the coastal plain; those values (dotted curve) are much too cool, have too small standard deviation, and have wrong sign of skew. The averaged distribution of observations located near model grid points 2 and 5 (Sonoma and King City respectively) are shown with a long dashed line.

Fig. 3 Comparison of observed and simulated ensemble means for dates of extreme hot surface temperatures. a) 850 hPa temperature anomaly ensemble mean reanalysis data during dates of the top 1% of normalized anomaly surface (2 m) daily maximum temperatures observed in the California Central Valley from 1979-1988. b) 700 hPa meridional wind component anomaly ensemble mean reanalysis data during the same dates used in panel a). The fields in a) and b) are the same as those used by Grotjahn (2011) to calculate the daily circulation index. c) Corresponding to (a) but using CCSM4 data during dates of the top 1% of normalized anomaly surface maximum temperatures at grid points 1, 2, and 5 (see Fig. 1) on the coastal plain. d) Corresponding to (b) but also using CCSM4 data at points 1, 2, and 5. The two temperature patterns appear to be generally similar but weaker in CCSM4 data: a positive anomaly near and over the CV with negative extrema upstream and downstream. However, the strong temperature anomaly is centered at or just off the west coast in the observation-based reanalysis (a), but in the CCSM4 data (c) the maximum is located onshore generally where the topographic slope is large in Figure 1. Centering the positive anomaly offshore amplifies surface heating inland by creating a near surface pressure pattern opposing a local sea breeze. It is unclear if the same mechanism can operate in CCSM4. Another difference is the upstream negative anomaly is further east than in reanalyses. The meridional wind patterns (b) and (d) have similar comparison.

Fig. 4 Scatterplots of positive values of the circulation index (see Grotjahn, 2011) versus near surface (2m) normalized daily maximum temperature anomaly. a) NDRA2 based circulation index and observed

CV data. b) and c) compare CCSM4 circulation index (projecting CCSM4 daily data onto parts of Fig. 3a,b) against CCSM4-CV-locs and CCSM4-mock-CV values, respectively. Ordinates are: a) observed normalized anomaly mean daily maximum temperature of three stations (R, F, B; see Fig. 1) spanning the CV; b) corresponding normalized anomaly mean daily maximum 2m temperatures at CCSM4-CV-locs grid points whose longitudes and latitudes lie within the actual CV and near R, F, and B; c) corresponding values for 3 CCSM4 grid points (1, 2, and 5) that are on the coastal plain. Generally, the stronger the circulation index the hotter the surface maximum tends to be (a). The strong link between circulation index and surface maximum temperatures seen in observations and reanalyses (a) is not as strong for the CCSM4.

Fig. 5 a) Full range histograms of normalized maximum surface temperature anomaly data and two circulation indices. '3-stn. Obs' are averages of the observed normalized anomaly data at R, F, and B. 'NDRA2' are circulation index values calculated from the NCEP/DOE AMIP-II reanalysis data, while 'CCSM4' uses the same procedure but projecting CCSM4 daily data onto the reanalysis ensemble mean to calculate CCSM4 based circulation index values. NDRA2 values have a similar distribution as surface observations, even for negative values. CCSM4 values should be directly compared with NDRA2 values. CCSM4 values have too little variation and positive skew. b) Histogram bar chart of the high tail of the distribution seen in (a). The top 1% of NDRA2 (solid bars) are those in the columns with circulation index 2.0 and higher. CCSM4 data are cross hatched columns. The NDRA2 data have more of the largest circulation index values than do the CCSM4 data (24 versus 9 above 2.0). However, comparable peak values do occur in the CCSM4 data.

Fig. 6 Duration (in days) above threshold 1.0 for the reanalysis data based circulation index (‘NDRA2’), the CCSM4 based circulation index (‘CCSM4’) and for the observed average normalized maximum temperature anomalies at R, F, and B (‘3-stn Obs). The duration periods above this threshold are similar for both circulation indices and for the observed surface maximum temperature (3-stn Obs).

Table 1 GPD fits of the data using 1.0 threshold

	Observed maxTa average of 3 CV stations (385 values)	NDRA2 analysis data circulation index (313 values)	CCSM4 historical simulation circulation index (239 values)
Scale parameter	0.49 (+/- 0.03)	0.59 (+/- 0.04)	0.39 (+/- 0.03)
Shape parameter	-0.30 (+/- 0.03)	-0.29 (+/- 0.04)	-0.10 (+/- 0.06)
100 year return period value (95% confidence range: low, high)	2.48 (2.41, 2.65)	2.80 (2.69, 3.08)	2.96 (2.63, 3.5)

Figures

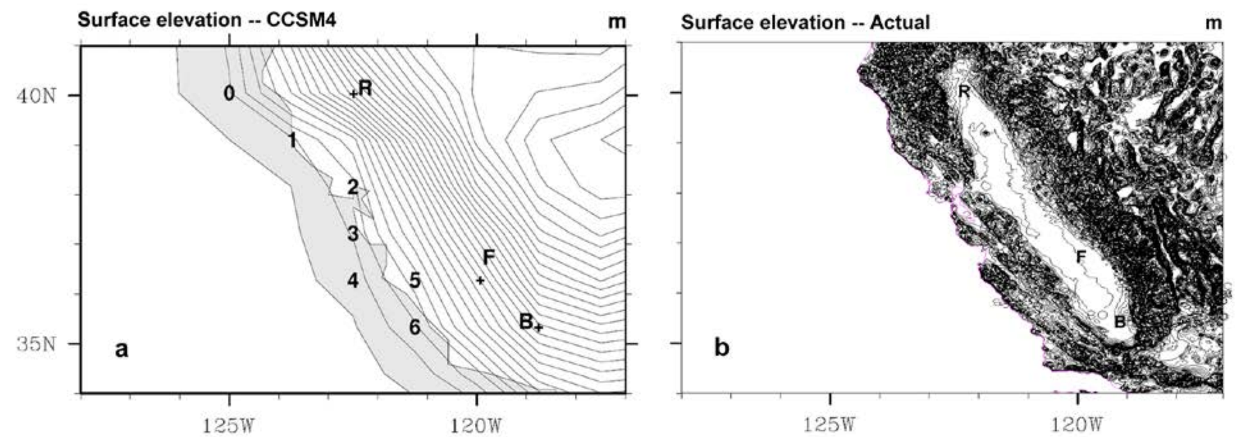

Fig. 1 Surface elevations for a) CCSM4 model and b) actual topography are shown (in meters, starting at 20m, using 60m interval). a) Grid point locations of CCSM4 model data discussed in the text. Geographic locations of California Central Valley (CV) stations used by Grotjahn (2011) in defining the downscaling scheme for extreme hottest days: R (Red Bluff, KRBL), F (Fresno, KFAT), and B (Bakersfield, KBFL). The model grid points used to represent R, F, and B are indicated by '+' symbols . The CV seen in b) is not resolved by the model and the model has notably higher topography at the actual location of the CV. Two grid point combinations: i) numbered 1, 2, and 5 and ii) indicated by + symbols were used (see text) as approximations to R, F, and B. The former are identified as the 'CCSM-mock-CV points while the latter are 'CCSM4-locs' points. Shaded area in a) indicates region above 20m elevation that should be ocean. (The model defines grid points at and near the ocean with both land and ocean fractions.) The actual topography has far steeper slope for the various mountain ranges in the region. Hence, contours range from 20-1640m in a) and 20-3620m in b).

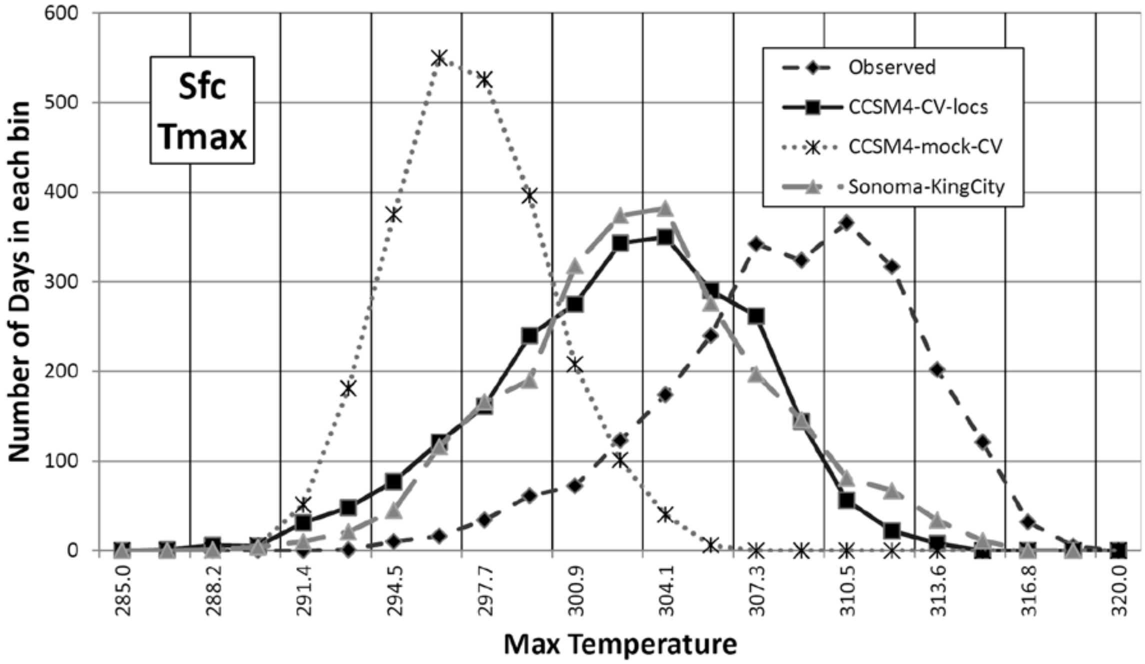

Fig. 2 Daily surface (2 m elevation above ground) maximum temperature comparison histograms; bins are 1.59C wide. 'Observed' values (shorter dashed line) are the average values from the three stations (R, F, B) in Fig. 1. 'CCSM4-CV-locs' are daily maximum 2 m temperatures (solid line) at three CCSM4 grid points ('+' symbols in Fig. 1a) close to the locations of R, F, and B. CCSM4-CV-locs values are generally similar (in values, standard deviation, and skew) to the observations (but ~3C too cool). 'CCSM4-mock-CV' values are averages of 2 m daily maximums at grid points 1, 2, and 5 on the coastal plain; those values (dotted curve) are much too cool, have too small standard deviation, and have wrong sign of skew. The averaged distribution of observations located near model grid points 2 and 5 (Sonoma and King City respectively) are shown with a long dashed line.

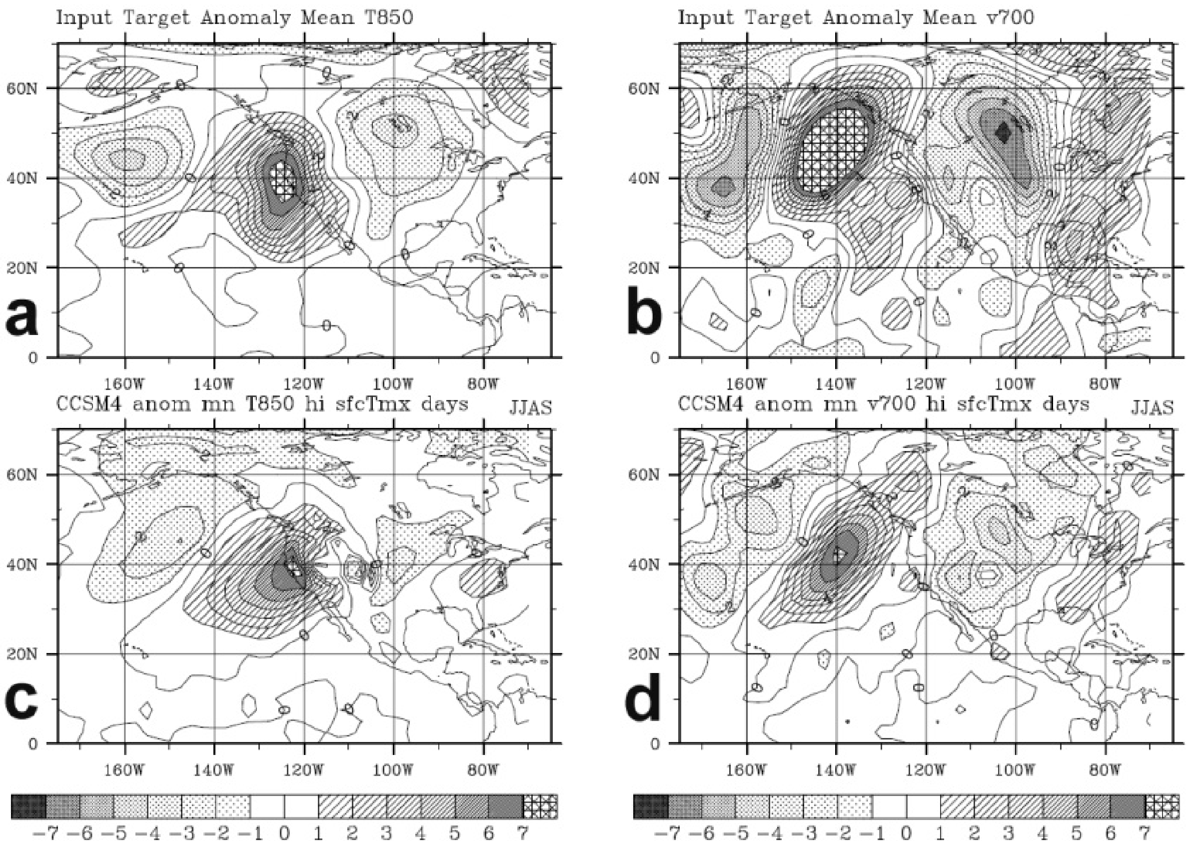

Fig. 3 Comparison of observed and simulated ensemble means for dates of extreme hot surface temperatures. a) 850 hPa temperature anomaly ensemble mean reanalysis data during dates of the top 1% of normalized anomaly surface (2 m) daily maximum temperatures observed in the California Central Valley from 1979-1988. b) 700 hPa meridional wind component anomaly ensemble mean reanalysis data during the same dates used in panel a). The fields in a) and b) are the same as those used by Grotjahn (2011) to calculate the daily circulation index. c) Corresponding to (a) but using CCSM4 data during dates of the top 1% of normalized anomaly surface maximum temperatures at grid points 1, 2, and 5 (see Fig. 1) on the coastal plain. d) Corresponding to (b) but also using CCSM4 data at points 1, 2, and 5. The two temperature patterns appear to be generally similar but weaker in CCSM4 data: a positive anomaly near and over the CV with negative extrema upstream and downstream. However, the strong temperature anomaly is centered at or just off the west coast in the observation-based reanalysis (a), but in the CCSM4 data (c) the maximum is located onshore generally where the topographic slope is large in Figure 1. Centering the positive anomaly offshore amplifies surface heating inland by creating a near surface pressure pattern opposing a local sea breeze. It is unclear if the same mechanism can operate in CCSM4. Another difference is the upstream negative anomaly is further east than in reanalyses. The meridional wind patterns (b) and (d) have similar comparison.

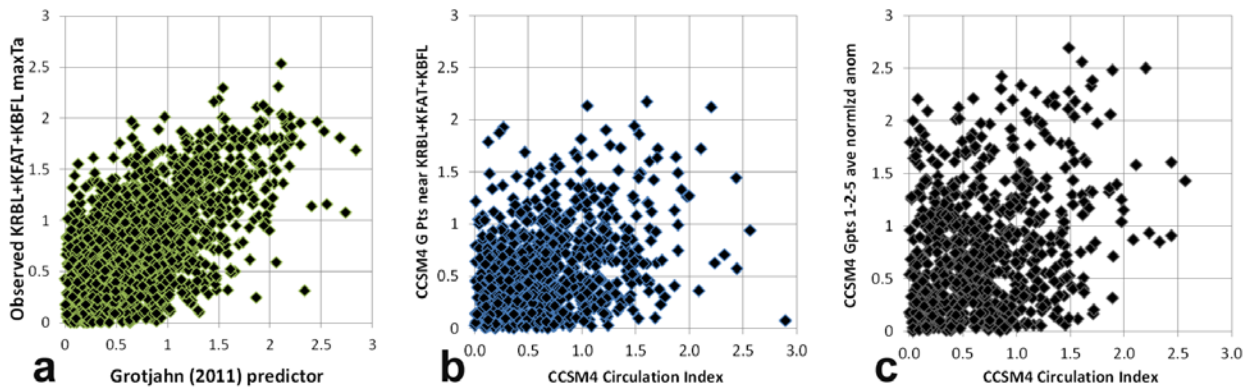

Fig. 4 Scatterplots of positive values of the circulation index (see Grotjahn, 2011) versus near surface (2m) normalized daily maximum temperature anomaly. a) NDRA2 based circulation index and observed CV data. b) and c) compare CCSM4 circulation index (projecting CCSM4 daily data onto parts of Fig. 3a,b) against CCSM4-CV-locs and CCSM4-mock-CV values, respectively. Ordinates are: a) observed normalized anomaly mean daily maximum temperature of three stations (R, F, B; see Fig. 1) spanning the CV; b) corresponding normalized anomaly mean daily maximum 2m temperatures at CCSM4-CV-locs grid points whose longitudes and latitudes lie within the actual CV and near R, F, and B; c) corresponding values for 3 CCSM4 grid points (1, 2, and 5) that are on the coastal plain. Generally, the stronger the circulation index the hotter the surface maximum tends to be (a). The strong link between circulation index and surface maximum temperatures seen in observations and reanalyses (a) is not as strong for the CCSM4.

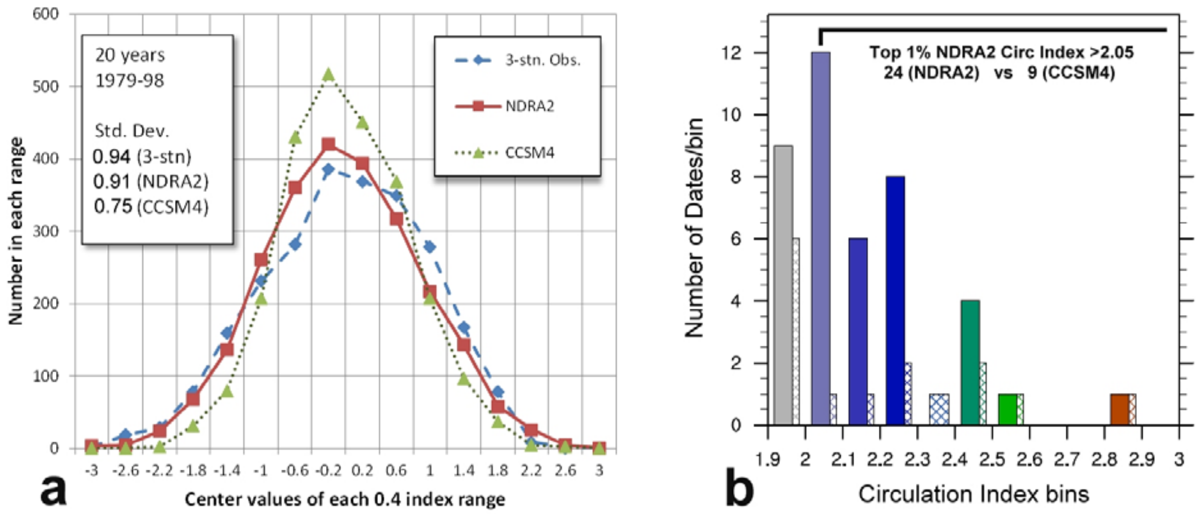

Fig. 5 a) Full range histograms of normalized maximum surface temperature anomaly data and two circulation indices. '3-stn. Obs' are averages of the observed normalized anomaly data at R, F, and B. 'NDRA2' are circulation index values calculated from the NCEP/DOE AMIP-II reanalysis data, while 'CCSM4' uses the same procedure but projecting CCSM4 daily data onto the reanalysis ensemble mean to calculate CCSM4 based circulation index values. NDRA2 values have a similar distribution as surface observations, even for negative values. CCSM4 values should be directly compared with NDRA2 values. CCSM4 values have too little variation and positive skew. b) Histogram bar chart of the high tail of the distribution seen in (a). The top 1% of NDRA2 (solid bars) are those in the columns with circulation index 2.0 and higher. CCSM4 data are cross hatched columns. The NDRA2 data have more of the largest circulation index values than do the CCSM4 data (24 versus 9 above 2.0). However, comparable peak values do occur in the CCSM4 data.

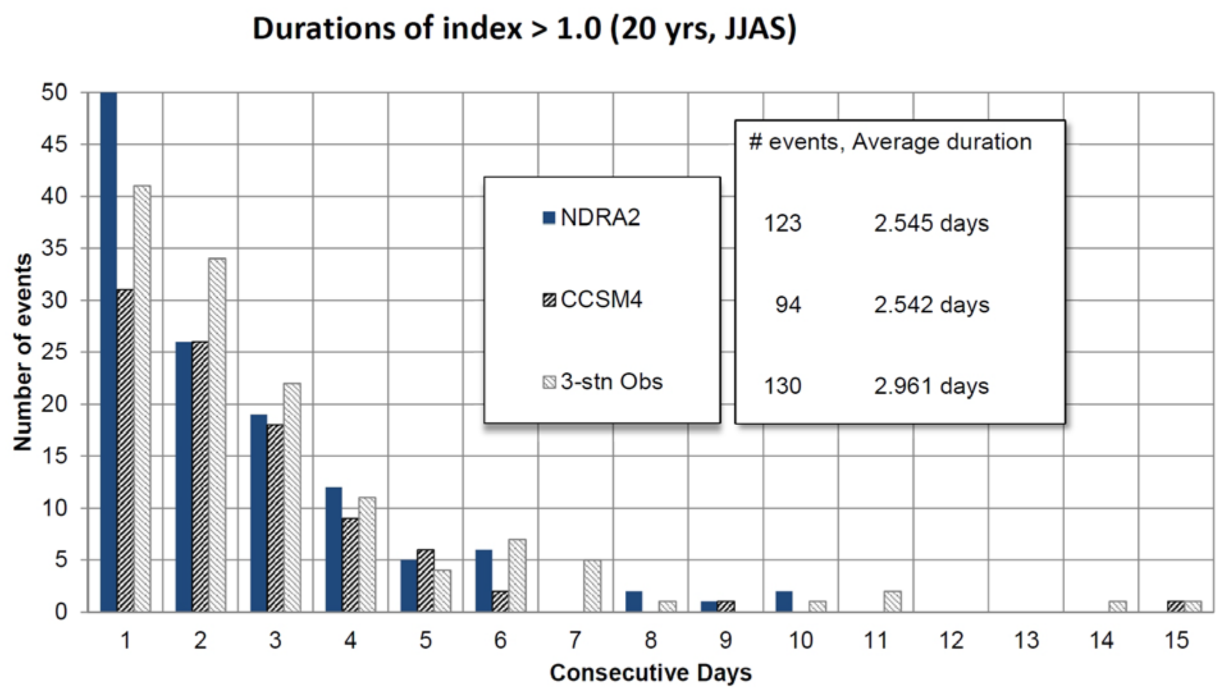

Fig. 6 Duration (in days) above threshold 1.0 for the reanalysis data based circulation index ('NDRA2'), the CCSM4 based circulation index ('CCSM4') and for the observed average normalized maximum temperature anomalies at R, F, and B ('3-stn Obs). While the number of number of events above the threshold varies, the duration periods above this threshold are similar (as measured by average duration) for both circulation indices and for the observed surface maximum temperature (3-stn Obs).