Electricity Markets & Policy
Energy Analysis & Environmental Impacts Division
Lawrence Berkeley National Laboratory

Future projections of wind patterns in California with the Variable-Resolution CESM

A clustering analysis approach

Meina Wang, Paul Ullrich, Dev Millstein

January 2020

This is a pre-print version of a journal article published in *Climate Dynamics.*
DOI: https://dx.doi.org/10.1007/s00382-020-05125-5

This work was funded by the California Energy Commission under the Electric Program Investment Charge Grant, EPC-15-068: Understanding and Mitigating Barriers to Wind Energy Expansion in California.

Future projections of wind patterns in California with the Variable-Resolution CESM

A clustering analysis approach

Meina Wang · Paul Ullrich · Dev Millstein

Received: date / Accepted: date

Abstract Wind energy production is expected to be affected by shifts in wind patterns that will accompany climate change. However, many questions remain on the magnitude and character of this impact, especially on regional scales. In this study, clustering is used to group and analyze large-scale wind patterns in California using model simulations from the Variable-Resolution Community Earth System Model (VR-CESM). Specifically, simulations have been produced that cover historical (1980-2000), mid-century (2030-2050), and end-of-century (2080-2100) time periods. Once clustered, observed changes to wind patterns can be analyzed in terms of both the change in frequency of those clusters and changes to winds within-clusters. Statistically significant capacity factors changes have been found at all five wind plant sites. Decomposition of the capacity factor changes into frequency changes and within-cluster changes enables a better understanding of their drivers. A further examination of the synoptic-scale fields associated with each cluster then provides a better understanding of how changes to large-scale meteorological fields are important for driving changes in localized wind speeds.

Keywords Wind energy · Climate change · Variable-resolution climate modeling · Clustering

1 Introduction

It is expected that wind energy production, as with many other environmentally-sourced renewable energy technologies, will be directly impacted by climate change. However, the highly localized character of wind fields, driven by a strong sensitivity to local topography, makes it difficult to model and project

M. Wang
University of California, Davis, Davis, CA 95616, USA
Tel.: +1(530)574-7972
E-mail: mnawang@ucdavis.edu

wind fields at the scales needed for stakeholders. Nonetheless, a better understanding of the variability of localized wind fields is essential to future wind energy resources planning and could help reduce the risk of selecting future wind project locations.

Even with the known difficulties with modeling wind, some progress has been made in better understanding this important resource. Past studies have focused on analyzing the climate change impact on localized wind fields, and the associated change in wind energy generation potential (Breslow and Sailor, 2002; Miller and Schlegel, 2006; Pryor and Barthelmie, 2010; Wang et al, 2018). Karnauskas et al (2018) analyzed simulations from ten climate models, and found reductions in wind power over Northern Hemisphere mid-latitudes, which can be explained by established features of climate change. Rasmussen et al (2011) employed model data from North American Regional Climate Change Assessment Program (NARCCAP) to project California wind energy change by the mid-century, and detected a decrease of $< 2\%$ in resources at Altamont Pass. Many studies also showed substantial regional and seasonal variations in future wind power change. Wang et al (2018) assessed the climate change impact through mid-century on California wind energy resources, and found that wind speed (and hence wind energy production) is likely to increase in summer, and diminish during fall and winter. Another study by Duffy et al (2014) also concluded that available wind energy in California will decrease in fall and winter. Yu et al (2015) detected upward trends in wind speeds across areas of the US Great Plains and Intermountain West, but downward trends in the east and in some parts of California. Pryor and Barthelmie (2011) found the the simulated future wind resources in the U.S. remain within the historical variability. While a study by Haupt et al (2016) found the future wind speed changes vary by up to 10% depending on different regions and seasons. However, these past studies have only assessed overall trends of wind patterns on seasonal scales, or focused only on one specific type of wind pattern.

In this study, we present a new approach that leverages an unsupervised machine learning algorithm, agglomerative clustering, to group wind patterns from unlabeled data into wind clusters. The unlabeled input data for the clustering algorithm is produced using the Community Earth System Model (CESM), a global climate modeling system that has some demonstrable skill with modeling wind (Wang et al, 2018). More details about the model can be found in Section 2. The agglomerative clustering algorithm is applied to the CESM model output to provide insight into the drivers and variability of different wind patterns. Once clusters have been identified, changes in wind fields between historical and end-of-century are decomposed into change in the cluster frequency and the change within each cluster. The insights gained from this decomposition then serve as our starting point for explaining significant trends that should be expected in the future. We investigate the cause of within-cluster wind speeds change by analyzing synoptic-scale fields associated with each cluster. However, we do not investigate the drivers of future change to the frequency of clusters, as these changes depend on global meteorological patterns that are beyond the scope of this study. Finally, seasonal changes of wind energy are assessed, along with the local impact of observed changes from wind clusters. Given appropriate regional climate data, this technique has the potential to be adapted to essentially any geographic region.

This work builds on a previous study by Millstein et al (2018), who used clustering to identify the characteristics of ten selected clusters over the historical time period. Their study then investigated the wind regime changes over the period of 1980-2015 in California, and further analyzed the impact on local wind energy resources. The present study works to expand the time scope of Millstein et al (2018) to the end of the 21st century, and detect any significant trends associated with the most relevant wind clusters.

For the purposes of this study, we have divided California into two subdomains: the Northern California (NC) domain, which includes Shiloh and Altamont Pass wind plant sites, and the Southern California (SC) domain, which includes Alta, San Gorgonio, and Ocotillo sites (Figure 1)[1]. These five wind plant locations include both wind plant sites currently in service, and wind project sites targeted for future development. The current capacities, according to the United States Wind Turbine Database (USWTDB)(Hoen et al, 2019), at each site is: 1,028 MW at Shiloh, 278 MW at Altamont Pass, 3,118 MW in the greater Tehachapi area, 663 MW in the San Gorgonio region, and 447 MW in the Ocotillo region. The current capacities Due to differences in wind patterns that emerge between NC and SC domains, the clustering algorithm was applied to the two domains separately.

The remainder of this paper is as follows: In section 2 we describe the VR-CESM model setup and the clustering algorithm used in this study. Results are presented in section 3, followed by discussion and conclusions in section 4.

2 Methods

This study uses model output from the Community Earth System Model (CESM), a widely-used global climate model (Neale et al, 2010; Hurrell et al, 2013). Three time periods were separately simulated, including historical (1980-2000), mid-century (2030-2050), and end-of-century (2080-2100). However, the mid-century period that was the focus of Wang et al (2018) is not considered in this study, and is only used to provide additional input for the clustering procedure. All simulations used the same model setup, enabling us to compare across time frames, with differences only in prescribed sea-surface temperatures and greenhouse-gas forcing. Details on model validation, including comparison with observational stations, reanalysis datasets, and other modeling products, can be found in Wang et al (2018).

[1] These wind plants names are representatives of an agglomeration of plants in close proximity to each other. Based on the calssification from California Energy Commission (CEC) (https://ww2.energy.ca.gov/maps/renewable/wind.html), Shiloh represents "Solano Wind Resource Area", Altamont represents "Altamont Wind Resource Area", Tehachapi represents "Tehachapi Wind Resource Area", San Gorgonio represents "San Gorgonio Wind Resource Area", Ocotillo represents "East San Diego Wind Resource Area".

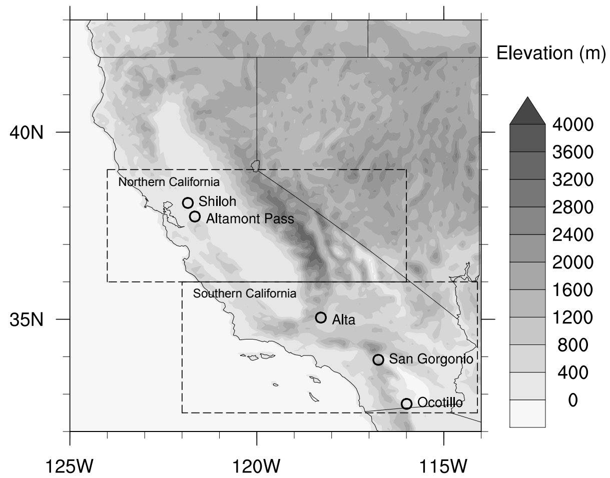

Fig. 1 The Northern California (NC) and Southern California (SC) domains with dash line bounding boxes, along with the five wind plant locations. This figure is a reproduction of Figure 1 from Millstein et al (2018).

2.1 Description of VR-CESM (global climate model product)

CESM version 1.5.5 was used for this study with the F-component set (FAMPIC5), which prescribes sea-surface temperatures and sea ice but dynamically evolves the atmosphere and land surface component models (AMIP protocols) (Gates, 1992). The atmospheric component model is the Community Atmosphere Model, version 5.3 (CAM5) (Neale et al, 2010) with the spectral-element (SE) dynamical core Dennis et al (2012) in its variable-resolution (VR) configuration (Zarzycki et al, 2014b). More details of the CAM5 configuration can be found in Neale et al (2010). The land component model used in this study is the Community Land Model (CLM) version 4.0 (Oleson et al, 2010). The SE dynamical core is employed along with variable resolution grid support. CAM5-SE is built with a continuous Galerkin spectral finite-element method to solve the hydrostatic atmospheric primitive equations. It has several benefits compared with the other CAM dynamical cores, including support of unstructured grids that eliminates grid singularities at higher latitudes, and near perfect multi-processor scalability (Zarzycki et al, 2014b,a; Zarzycki and Jablonowski, 2014; Taylor and Fournier, 2010). Physical parameterizations in CAM5 include aerosols (Ghan et al, 2012), deep convection (Neale et al, 2008), macrophysics (Park et al, 2014), microphysics (Morrison and Gettelman, 2008), radiation (Iacono et al, 2008), and shallow convection (Park and Bretherton, 2009). Further details regarding CAM5-SE can be found in Neale et al (2010). More details on VR-CESM can be found in Rhoades et al (2018b, 2016), and Huang et al (2016). The VR model grid used for this study, depicted in Figure 2, was generated for use in CAM and CLM with the open-source software package SQuadGen (Ullrich, 2014; Guba et al, 2014). This grid has a finest horizontal resolution of 0.125°(∼14km) over the western United States, with a quasi-uniform 1° mesh over the remainder of the globe. Three simulations were conducted on this grid: The historical run covered the period from October 1st, 1979 to December 31st, 2000, with the last three months of 1979 discarded as the spin-up period, for a total of 21-years of three-hourly output. This historical time period was chosen to provide an adequate sampling of the inter-annual variability, as well as coincide with the satellite era for model validation with reanalysis datasets. For projections of future wind energy change, our mid-century and end-of-century simulations ran with the "business as usual" Representative Concentration Pathway 8.5 (RCP8.5) (Taylor et al, 2012) from October 1st, 2029 to December 31st, 2050, and from October 1st, 2079 to December 31st, 2100, respectively. In each case the first three months of the simulation were discarded, yielding two additional 21-year-long simulations. Analogous simulations with VR-CESM have also been conducted by Rhoades et al (2018a) and Huang and Ullrich (2017) for assessing snowpack and future precipitation, respectively. Greenhouse gas (GHG) and aerosol forcings are prescribed based on historical or RCP8.5 concentrations for each simulation. Historically prescribed SST and sea-ice were derived from the Hadley Centre sea ice and SST dataset version 1 (HadISST1) and version 2 of the National Oceanic and Atmospheric Administration (NOAA) weekly optimum interpolation (OI) SST analysis (Hurrell et al, 2008). Future SSTs and sea-ice forcings were derived from a future 1 degree RCP8.5 bias-corrected dataset (Small et al, 2014). Both datasets were developed at NCAR. The historical and mid-century VR-CESM simulations were previously validated and analyzed in Wang et al (2018). Here we expand the time horizon through the end of the 21st century, and analyze the potential changes on localized wind regimes. We also validated the end-of-century simulation from VR-CESM against 33 model projections from CESM LENS (Kay et al, 2015) by comparing the 700hPa geopotential height field, and this comparison indicates the robustness of the projection from VR-CESM (not shown). Note that in Wang et al (2018), we found that although the large-scale patterns are captured, there is nonetheless a low wind speed bias from VR-CESM which leads to an under estimation of capacity factor.

In order to calibrate the wind speed from VR-CESM, we estimated a bias correction factors of 1.3 in Wang et al (2018). This bias-correction factor was calculated based on a comparison between VR-CESM and a high-resolution regional simulation (referred to as DNV GL in Wang et al (2018)). Linear bias correction factors have been applied in past efforts in order to match global modeling or reanalysis outputs with operational data, for example, see Staffell and Pfenninger (2016) and Olauson et al (2017). The use of a linear factor effectively assumes that the dynamics and variability of the atmosphere above the boundary layer are captured well by the model, but that the dominant errors instead emerge from downscaling of the near surface winds to the sub-grid-scale – i.e. from a failure to capture local topographic effects, surface friction, or turbulence. Given that VR-CESM appears to capture the process drivers and dynamical character of the wind field well (Wang et al, 2018; Huang

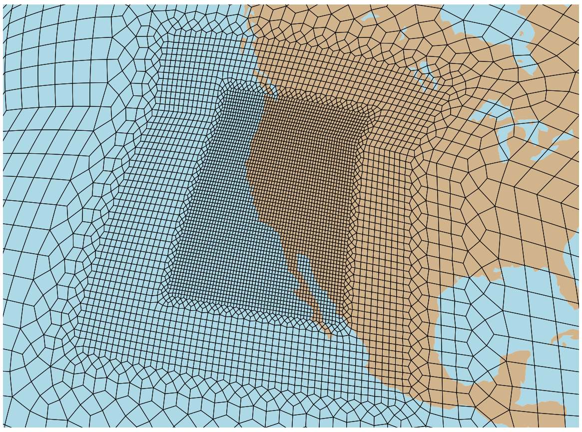

Fig. 2 The VR-CESM grid used in this study, constructed by first successively refining a cubed-sphere grid with a 1°(111km) quasi-uniform resolution to a resolution of 0.125°(∼14km) over the western USA. This figure is a reproduction of Figure 2 from Wang et al (2018).

et al, 2016), we believe this is a reasonable assumption. Capacity factors, which are analyzed in section 3.3, were therefore calculated from the bias-corrected wind speed. We used the capacity factor (CF) to measure the wind energy production. CF is a key concept measuring the ratio (%) of energy generated by a turbine to the energy that same turbine could have generated had it been running at its rated capacity continuously. More details on the calculation of CF can be found in supplement material Section 2.

2.2 Agglomerative clustering

In the nomenclature of machine learning, the output data from the CESM model simulations is referred to as "unlabeled" – namely, there is no prior knowledge of the different wind patterns and their associated frequencies. In order to develop such a labeling, we apply an unsupervised machine learning algorithm to group and distinguish different wind patterns. Specifically, we use the agglomerative clustering algorithm with Ward's method (Ward Jr, 1963) to minimize the total within-cluster variance. Under this algorithm, each data point is initialized as a single-item cluster. At each iteration of the method, smaller nearby clusters are chosen to merge and form larger clusters; the particular choice of merged clusters minimizes a global inter-cluster distances metric (i.e., Ward's method minimizes the variance of clusters being merged). This "bottom-up" algorithm then iterates to create a dendrogram, which is tree-like structure, illustrating the arrangement of clusters. The number of clusters used in the subsequent analysis can then be varied by halting the iteration procedure at a particular level. Typically this choice is made through inspection of the resulting clusters at each iteration, so as to identify the earliest point at which there is sufficient distinction between all clusters in the set.

This algorithm's primary advantage over k-means clustering (Hartigan and Wong, 1979) is that it does not require the parameter k (how many clusters to generate) to be specified beforehand. Since we did not have prior knowledge of the number of distinct wind patterns before execution of the clustering algorithm, agglomerative clustering provided a natural mechanism to tune this value.

In this study, clustering is solely applied to 80m wind vector fields (composed of horizontal and meridional wind magnitudes). This particular height of 80m was chosen as it is typical of the hubs of large wind turbines. The clustering was accomplished through two steps: first, we reduced the dimensionality of the input data using the principal components analysis (PCA); second, we applied the agglomerative clustering algorithm to the principal components. This approach is similar to the steps taken in Ludwig et al (2004), Conil and Hall (2006), Jin et al (2011), Berg et al (2013), and Millstein et al (2018).

For the first step, principal component analysis (PCA) was applied to 3-hourly (eight times daily) 80m wind vector fields to reduce dimensionality. We retained the first ten principal components for clustering, as they accounted for over 80% of the total variance. Then, each day was categorized into a particular cluster based on a set of (8 times daily $\times$ 10 pricipal components) 80 PCA components. For each region (NC and SC), regridded data from all three time periods (historical 1980-2000, mid-century 2030-2050, and end-of-century 2080-2100) was simultaneously provided as input to the clustering algorithm. This was to ensure the consistency of clusters across all three time periods. Then for the second step, we ran the agglomerative clustering algorithm separately on NC and SC domains since the synoptic-scale wind patterns produce distinct localized effects in these regions. The agglomerative clustering is a "bottom-up" approach, which begins with each day classified as its own cluster, then "similar" days are then merged together into larger groups based on minimizing a criterion (Wards method minimizes the variance of the clusters being merged). To determine how many wind patterns would be needed to distinguish wind regimes, we leveraged the dendrogram produced by the agglomerative clustering algorithm and determined the point when distinctly different wind patterns were merged (Wilks, 2011). After examination of the clustering output (wind patterns from each cluster), we concluded that for each of NC and SC domains, ten clusters provided a good representation of different wind regimes – namely, lesser clusters did not sufficiently distinguish various qualitatively different wind patterns, and more clusters produced several instances of cluster pairs with only subtle differences. For example, if we were to keep 5 clusters, then the wind patterns did not portray the full range of patterns we've found from 10 clusters, and the set of 15 clusters contained clusters with similar wind patterns. A quantitative assessment using the CH index (Caliński and Harabasz, 1974), which measures the overall within-cluster variance and the overall between-cluster variance, confirmed the optimality of ten clusters in each region. Namely, ten clusters produced a higher CH index than the index from either five and fifteen clusters – indicating that the clusters have larger between-cluster variance, and smaller within-cluster variance.

Therefore, we determined for both NC and SC domains, ten clusters would work the best in our case. Note that in the remainder of the text the numbers associated with each cluster do not bear meaning, and are only for labeling purposes. Each cluster is labeled by its domain and cluster number (e.g. NC 6 is cluster 6 from NC domain).

2.3 Decomposition of changes in wind clusters

Climate change can impact wind clusters through two principal avenues: First, through the modification of the frequency of the wind cluster, and second, through the modification of the wind patterns within each cluster. The change in either the total wind field or the wind field of each cluster can be decomposed into these two contributions as follows. We denote the historical frequency of a given cluster i as f_i^h, the end-of-century frequency as f_i^e, the historical average wind field within the cluster by U_i^h, and the end-of-century wind field within the cluster by U_i^e. Thus the average historical U^h and end-of-century U^e wind fields can be written as:

$$U^h = \sum_i U_i^h f_i^h, \qquad U^e = \sum_i U_i^e f_i^e. \tag{1}$$

The average frequency of the cluster f_i and average wind field within the cluster U_i (combining both historical and end-of-century) are then given by

$$f_i = \frac{1}{2}(f_i^h + f_i^e), \qquad U_i = \frac{U_i^h f_i^h + U_i^e f_i^e}{f_i^h + f_i^e}. \tag{2}$$

Similarly, the change in cluster frequency and change in wind field within cluster i is defined by $\Delta f_i = f_i^e - f_i^h$ and $\Delta U_i = U_i^e - U_i^h$. Denoting the change in the average wind field by $\Delta U = U^e - U^h$ and making an ansatz that ΔU can be decomposed into a term proportional to $U_i \Delta f_i$, a term proportional to $f_i \Delta U_i$, and some nonlinear leftover term then leads to the decomposition:

$$\Delta U = \sum_i U_i^e f_i^e - U_i^h f_i^h \tag{3}$$

$$= \sum_i \underbrace{U_i \Delta f_i}_{\text{(a)}} + \underbrace{(U_i^e - U_i^h) f_i}_{\text{(b)}} \underbrace{- \frac{\Delta f_i^2 (U_i^e - U_i^h)}{4 f_i}}_{\text{(c)}}. \tag{4}$$

Here (4a) denotes the change in average wind speed due to the change in frequency of cluster i, (4b) denotes the change in average wind speed due to the change in the wind field within each cluster i, and (4c) denotes nonlinear changes associated with simultaneous changes in frequency and wind field. In this wind speed decomposition, U represents the wind speed magnitude from VR-CESM, not the wind vector field. Note that such a decomposition is independent of our choice of clustering technique, and can be performed for any grouping of fields from two periods.

3 Results

Section 3.1 describes the wind patterns associated with each cluster. Section 3.2 then examines the climatological synoptic-scale fields from clusters with significant trends. In section 3.3, we analyze the future projections of wind clusters from the end-of-century VR-CESM simulation, and their impact on wind energy output.

Our results mirror those of previous work on this subject (Wang et al, 2018; Duffy et al, 2014; Miller and Schlegel, 2006) that have found a reduction of overland wind speeds in DJF and an increase in wind speeds in JJA. This change means that, in general, we see a decrease (increase) in the frequency of clusters that have high wind speeds and a decrease (increase) in the wind speeds across clusters in DJF (JJA).

3.1 Trends in cluster frequency

As described in section 2.2, days from historical and end-of-century time periods were grouped into ten clusters per region (NC and SC) based solely on wind vector fields (twenty clusters total). A qualitative summary of these clusters, their dominant seasonality, and end-of-century minus historical frequency change (annual and broken down by season) is given in Table 1. By using a combined dataset of historical and end-of-century daily wind fields as input for the cluster analysis, we would generally expect that changes in cluster frequency will dominate the total change in the wind field. Namely, since the cluster analysis is, in effect, grouping days with similar wind fields, we expect that the wind field for days in a particular cluster to be more similar to one another than to the wind field of days in another cluster. For each of these twenty clusters, Figures S3-S5 show the magnitudes of each of the three terms in Equation (4) for the northern California clusters. In general, we observe that change in cluster frequency is the dominant contributor to change in wind patterns, followed by changes in wind fields within each cluster (except in those cases where the change in cluster frequency is small). In each case the nonlinear term is not a significant contributor to the overall change. The remainder of this section focuses on analysis of select clusters, with additional discussion on the large-scale drivers that could influence the wind climatology in each case.

3.2 Synoptic-scale character of prominent clusters

This section describes the synoptic-scale character of the select clusters from Table 1. We focus on analyzing the mean meteorological fields, including the 700hPa geopotential height, and the wind field at 80m above the ground. The 700hPa geopotential height field was chosen as it is reflective of the general circulation, with wind flow at this level being largely geostrophic but still

Table 1 Top: Dominant seasons, historical frequency, end-of-century frequency changes, and qualitative summary for NC and SC clusters. Bottom: Historical frequency and end-of-century frequency change broken down by season. Frequency changes indicated in bold are significant under the two-proportion z-test at the 95% significance level. The seasonal frequency of these clusters is also depicted in Figures S1 and S2. Seasons are March-April-May (MAM), June-July-August (JJA), September-October-November (SON), and December-January-February (DJF).

Cluster	Dominant Seasons	Annual f_i^h	Δf_i	Qualitative summary
NC 1	DJF MAM	13.6%	**-1.5%**	Westerly wind
NC 2	DJF	10.2%	**-1.3%**	Stronger westerly wind w/ offshore trough
NC 3	DJF SON	11.2%	**- 3.2%**	Offshore blocking
NC 4	SON MAM	13.4%	- 0.5%	Low wind
NC 5	JJA	5.3%	+ 0.3%	Strong northerly wind
NC 6	JJA MAM	12.7%	**+ 2.4%**	Northwesterly wind (marine air penetration)
NC 7	JJA MAM	12.3%	+ 0.2%	Strong northwesterly (marine air penetration)
NC 8	JJA SON	8.0%	**+ 2.1%**	Northerly wind (marine air penetration)
NC 9	DJF MAM	9.2%	+ 0.6%	Low southerly wind
NC 10	JJA	4.0%	**+ 0.8%**	Strongest northwesterly (marine air penetration)
SC 1	MAM DJF	14.1%	**- 1.1%**	Strong alongshore wind
SC 2	JJA SON	23.1%	- 0.3%	Weak onshore flow
SC 3	DJF MAM	12.5%	+ 0.4%	Low wind
SC 4	JJA MAM	15.5%	**+ 2.8%**	Onshore flow
SC 5	DJF	3.8%	- 0.5%	Southwesterly wind
SC 6	DJF SON	8.8%	**- 2.3%**	Santa Ana winds
SC 7	JJA SON	7.3%	**+ 2.0%**	Weakened onshore flow
SC 8	DJF MAM	7.2%	**- 1.7%**	Westerly wind
SC 9	SON MAM	4.9%	**+ 1.0%**	Low wind
SC 10	DJF MAM	2.8%	- 0.4%	Onshore flow

Cluster	MAM f_i^h	MAM Δf_i	JJA f_i^h	JJA Δf_i	SON f_i^h	SON Δf_i	DJF f_i^h	DJF Δf_i
NC 1	17.5%	- 0.9%	1.1%	**- 0.8%**	15.7%	**- 5.1%**	20.5%	+ 0.8%
NC 2	9.3%	**- 2.6%**	0.1%	0.0%	7.0%	-1.1%	24.5%	- 1.3%
NC 3	7.1%	**- 1.7%**	1.0%	**- 0.9%**	15.2%	**- 6.4%**	21.7%	**- 3.9%**
NC 4	17.8%	+ 0.5%	5.8%	**- 4.0%**	20.8%	- 0.1%	9.4%	+ 1.6%
NC 5	2.3%	**+ 1.3%**	15.7%	- 0.9%	3.0%	+ 0.7%	0.0%	+ 0.1%
NC 6	17.5%	-0.5%	19.1%	**+ 6.2%**	11.7%	**+ 3.8%**	2.3%	+ 0.2%
NC 7	11.7%	**+ 2.4%**	27.3%	**- 3.1%**	8.1%	+ 0.7%	1.8%	+ 0.8%
NC 8	4.3%	**+ 1.9%**	16.8%	**+ 3.3%**	10.5%	**+ 3.2%**	0.3%	+ 0.1%
NC 9	10.6%	- 1.1%	0.2%	- 0.1%	6.7%	**+ 2.1%**	19.5%	+ 1.7%
NC 10	1.9%	+ 0.7%	12.9%	+ 0.3%	1.2%	**+ 2.3%**	0.0%	0.0%
SC 1	22.7%	- 1.4%	2.2%	- 0.6%	13.3%	**- 3.5%**	18.4%	+ 1.0%
SC 2	19.5%	**+ 3.1%**	45.9%	**- 6.4%**	21.4%	+ 2.0%	5.2%	0.0%
SC 3	12.2%	**- 2.6%**	0.2%	0.0%	16.5%	- 1.3%	21.4%	**+ 5.5%**
SC 4	17.2%	**+ 4.0%**	30.8%	**+ 5.3%**	12.9%	+ 1.5%	0.8%	+ 0.7%
SC 5	2.5%	+ 0.2%	0.0%	0.0%	1.9%	- 0.5%	10.7%	- 1.7%
SC 6	4.0%	**- 1.7%**	0.0%	+ 0.1%	10.4%	**- 4.4%**	21.1%	**- 3.0%**
SC 7	2.8%	**+ 2.1%**	18.0%	+ 1.9%	7.7%	**+ 4.2%**	0.5%	- 0.2%
SC 8	10.8%	**- 3.9%**	0.2%	+ 0.2%	4.9%	- 0.3%	12.9%	**- 2.8%**
SC 9	5.8%	+ 0.8%	2.7%	- 0.5%	8.8%	**+ 2.8%**	2.3%	**+ 1.1%**
SC 10	2.4%	- 0.5%	0.0%	+ 0.1%	2.0%	- 0.5%	6.8%	- 0.5%

strongly connected with near-surface winds. Because of the terrain-following coordinate, the lowest model level in CESM is everywhere below the 80m level, and so all wind speeds are interpolated. The interpolation procedure is as follows: the CAM5 hybrid coordinates are first converted to pressure coordinates; the height of each pressure surface above ground level (AGL) is computed by subtracting the surface geopotential height from the geopotential height at the model level; two model levels that bound the 80m AGL are used, and logarithmic interpolation is applied to obtain the wind speed at 80m AGL. Specifically, the interpolation was performed by fitting a log equation with the two levels bounding 80m AGL, then interpolating the wind at 80m AGL (Justus and Mikhail, 1976). The figures in each subsection show the meteorological fields for these clusters. For each figure, the top left plot shows the historical mean 700hPa geopotential height; top right shows the historical mean 80m wind field (U_i^h); bottom left shows the change in geopotential height within the cluster; bottom middle shows the end-of-century wind speed change due to the change in cluster frequency $(U_i \Delta f_i)/f_i$ (see section 2.3); and bottom right shows the mean end-of-century 80m wind speed minus mean historical 80m wind field ($U_i^e - U_i^h$).

3.2.1 NC 1 and NC 2: Reduced ventilation from westerly winds

Clusters NC 1 (westerly wind) and NC 2 (stronger westerly wind) in the NC domain are frequent (13.6% and 10.2%) wind patterns that peak in frequency during the winter season (20.5% and 24.5% frequency in DJF). They are accompanied by relatively large annual frequency changes (-1.5% and -1.3%), with the largest decreases occurring in the spring and fall. Further analysis of these patterns is beneficial to explain decreases in wind energy output during DJF, described later in the paper (Table 5).

NC 1 is the most frequent cluster in NC domain (13.6%) (Figure 3), and sees a large frequency decrease of 1.5%. The 700hPa geopotential height field from Figure 3 is a driver for strong alongshore winds, particularly along the coast of central California. The geopotential gradient perpendicular to the coast from NC 1 is significantly smaller than NC 2, and so NC 1 is associated with weaker onshore winds. Comparing end-of-century to historical, the geopotential height increase in the Eastern subtropical Pacific produces a weaker, westerly wind pattern.

Among the two, cluster 2 shows higher wind speed in NC domain than cluster 1. The synoptic-scale fields for NC 2 are depicted in Figure 4. The 700hPa geopotential height field shows a trough over the Gulf of Alaska that promotes flow directed perpendicular to the coast and hence on-shore ventilation through the NC domain. As discussed later, NC 2 tends to produce the highest wind speeds at the Shiloh and Altamont Pass wind plants among all clusters, and so a reduction in the frequency of this pattern will be associated with decreasing NC capacity factors in DJF. Comparing end-of-century to historical within this cluster, two effects appear to be prominent: First there is an increase in the geopotential gradient in the mid-Pacific which drives up

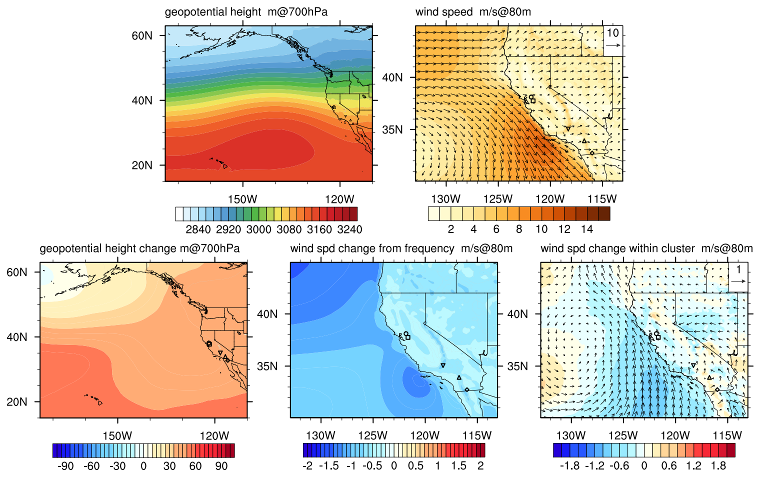

Fig. 3 Meteorological fields from cluster NC 1. (top left) Historical mean 700hPa geopotential height; (top right) 80m historical wind field; (bottom left) 700hPa geopotential height change; (bottom middle) end-of-century minus historical wind speed change due to change in cluster frequency ($U_i \Delta f_i / f_i$); and (bottom right) end-of-century minus historical wind speed change within-cluster ($U_i^e - U_i^h$).

wind speeds over the open ocean. However, simultaneously increased overland temperatures (not shown) appear to be promoting an increase in the overland geopotential height (thicker air masses from warmer temperature). This second factor drives a reduction in onshore flow, and consequently we observe decreasing wind speeds within this cluster across the NC domain.

3.2.2 NC 3: Reduced offshore blocking

Figure 5 depicts the synoptic-scale fields from NC 3, which again peaks in the winter season and exhibits a frequency decrease of 3.2% through end-of-century. This cluster corresponds to offshore blocking along the California coast. In opposition to NC 6 (associated with summertime marine air penetration), this cluster exhibits a pronounced ridge over the Eastern Pacific, leading to a strong northerly wind flow parallel to the California coastline that is associated with the second largest wind speeds at the NC wind plants. Within this cluster, the 700hPa geopotential height field exhibits a broad increase in end-of-century; however, the change in geopotential height is larger at lower latitudes and smaller over the Northern Pacific. This leads to a weakening of the northerly flow, in turn causing an overall decrease in offshore and onshore wind speeds. Overall, the decrease in frequency and character of this pattern drives weaker wind speeds at both Shiloh and Altamont Pass.

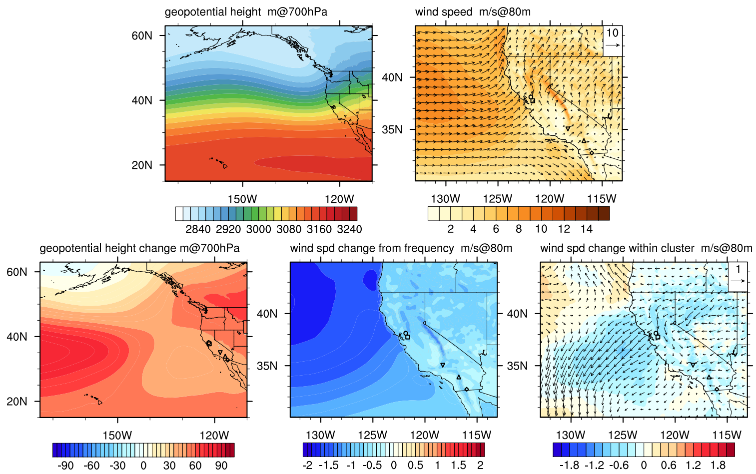

Fig. 4 As Figure 3 but for NC domain cluster 2.

Note that other studies (i.e., Wang and Schubert (2014)) noted an increased trend in blocking over the 20th century, particularly in the Gulf of Alaska, which seems contrary to our observations in this section (particularly given that NC 3 is representative of this offshore blocking pattern). To assess if this trend is present in the VR-CESM data, we counted blocking days at each grid point over each DJF season, defined as days where the geopotential at a given point exceeded the climatological geopotential for that period plus one standard deviation (separately calculated for historical and end-of-century). Note that the blocking days were selected outside the clustering framework, using only the aforementioned criterion. The results of this analysis are plotted in Figure 6, and are inconsistent with an increased blocking frequency.

3.2.3 NC 6-8 and NC 10: Increased summertime marine air penetration (MAP)

Figure 7 depicts the synoptic-scale fields of cluster 6 in the NC domain, which is expected to increase in frequency by 2.4% through end-of-century. The change in frequency of this cluster appears to occur in conjunction with a decreasing frequency of the NC 4 cluster (supplement Figure 6), associated with low wind events. NC 6 is indicative of a typical summertime marine air penetration (MAP) condition (Wang and Ullrich, 2017; Beaver and Palazoglu, 2006; Fosberg and Schroeder, 1966). Clusters NC 7 (supplement Figure 8), NC 8 (supplement Figure 9), and NC 10 (supplement Figure 11) also show an analogous, but stronger synoptic pattern and are depicted in the supplemental materials. Notably, the increasing frequency of summertime MAP events from

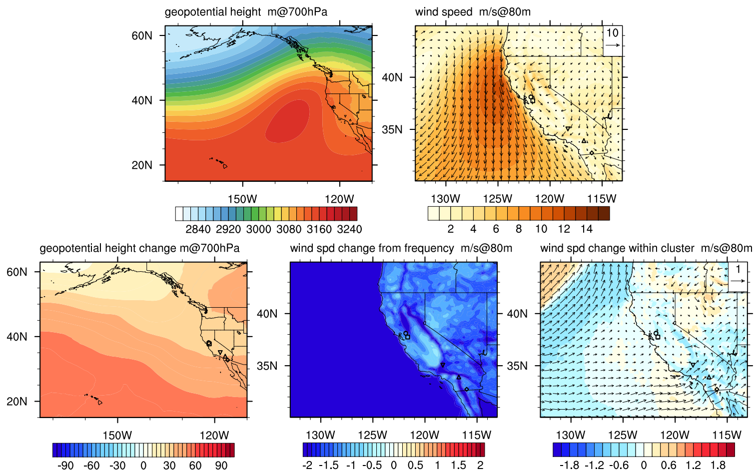

Fig. 5 As Figure 3 but for NC domain cluster 3.

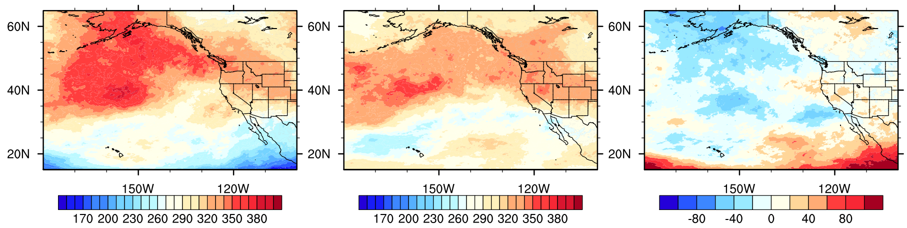

Fig. 6 Total number of days each grid point exceeds the mean plus one standard deviation of 500hPa geopotential height field for (Left) historical and (Center) end-of-century. (Right) Difference between end-of-century and historical.

these clusters agrees with the findings of Wang and Ullrich (2017). MAP events feature an off-shore trough and geopotential height contour lines perpendicular to coastline, allowing cool and moist marine air to penetrate inland. It is the location of the off-shore trough that is directly responsible for driving marine air through the San Francisco Bay Delta.

Within this cluster and relative to the historical period, the magnitude of the 700hPa geopotential height field under the end-of-century increases, as a direct consequence of low-level warming (not shown). This low-level warming drives a thickening of air layers and thus an increase in the 700hPa geopotential height field. However, this increase is less pronounced over the Northern Pacific, which drives a weakening of the typically northerly wind pattern that traces the coastline in Northern California, and an increase in the on-shore flow pattern driven by the general circulation. This in turn leads to an increase in wind speeds through the San Francisco Delta region during MAP days (and at

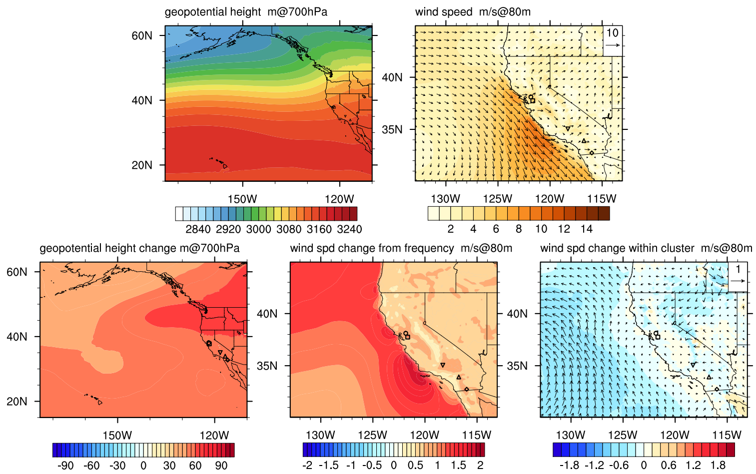

Fig. 7 As Figure 3 but for NC domain cluster 6.

Shiloh and Altamont Pass in NC domain). A shift in this particular synoptic-scale pattern also drives increased ventilation in the SC domain.

These changes to frequency and wind pattern suggest the tendency towards more MAP days and more intense MAP winds are primary drivers for increased summertime wind speeds in the San Francisco Bay region.

3.2.4 SC 1: More seasonally concentrated strong alongshore wind

Moving to the SC domain, cluster SC 1 captures days of strong alongshore wind off the U.S. west coast (Figure 8) that appear most prominently between the fall and spring seasons. The alongshore flow weakens south of the SC domain, leading to alongshore convergence that induces transverse inland flow of the marine air through the Los Angeles region. This pattern is associated with some of the highest historical capacity factors for the Alta wind plant (see table 7). Due to the location of Alta wind plant, which sits in the pass between in the Tehachapi mountains, the ventilation from the San Joaquin valley to the Mojave also contributes to the high capacity factors. It is also a frequent pattern, and one that has been projected to decrease in frequency by 1.1% annually; however, this change in frequency is primarily because of an increase in seasonality – the pattern sees an increase in frequency in DJF but decrease in MAM and SON. Within this cluster, the 700hPa geopotential height field change shows an inhomogenous pattern that favors overland warming, and reduces the alongshore gradient, thus leading to a weakening of the flow. The net result of these changes is a reduction in spring and winter wind speeds in the SC region.

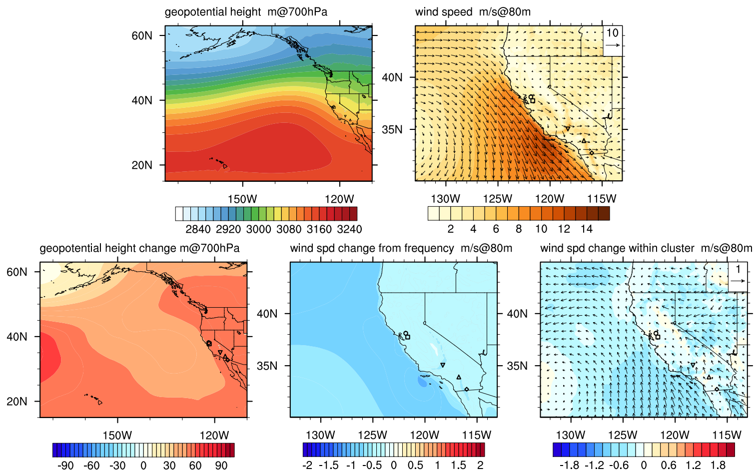

Fig. 8 As Figure 3 but for SC domain cluster 1.

3.2.5 SC 4: Increased summertime marine air penetration

Spring and summertime marine air penetration is also reflected in the SC domain via cluster SC 4, and its increased frequency through end-of-century supports our prior observations with cluster NC 6 (marine air penetration). As shown in Figure 9, a local trough sits off-shore with a 700hPa geopotential contour perpendicular to the shoreline in SC domain, leading to onshore marine air. Within-cluster changes to wind speeds are small (and largely mixed) over California, but the increased frequency of SC 4 suggests increased ventilation of the SC domain. The end-of-century change to the 700hPa geopotential height surface also produces a small enhancement in wind speeds parallel to the shore. Consequently both the increased frequency of SC 4 and slightly increased onshore winds within SC 4 leads to increased ventilation of the SC domain.

3.2.6 SC 5: Less frequent wintertime southwesterly wind

SC 5 represents wintertime southwesterly wind from an offshore trough sitting near the U.S. west coast. This cluster brings relatively high wind speeds, but is becoming less frequent during the winter season. By the end-of-century, the offshore trough intensifies, leading to higher wind speeds over the Pacific. Simultaneously, the 700hPa geopotential height anomaly center over the SC domain acts to block the onshore wind, leading to wind speeds decreasing over almost all areas within California.

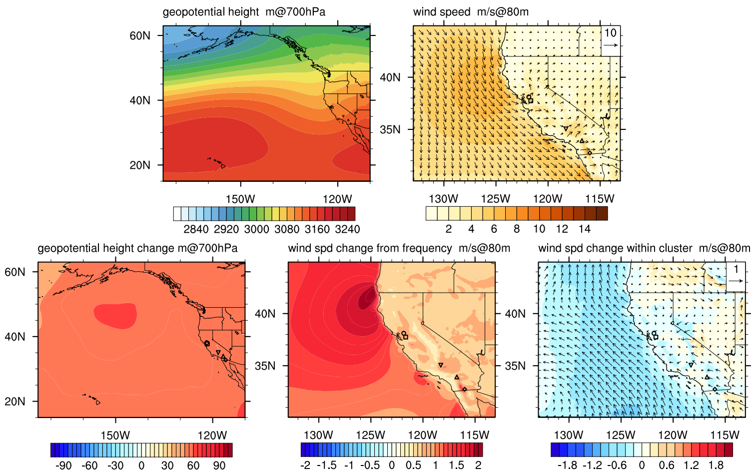

Fig. 9 As Figure 3 but for SC domain cluster 4.

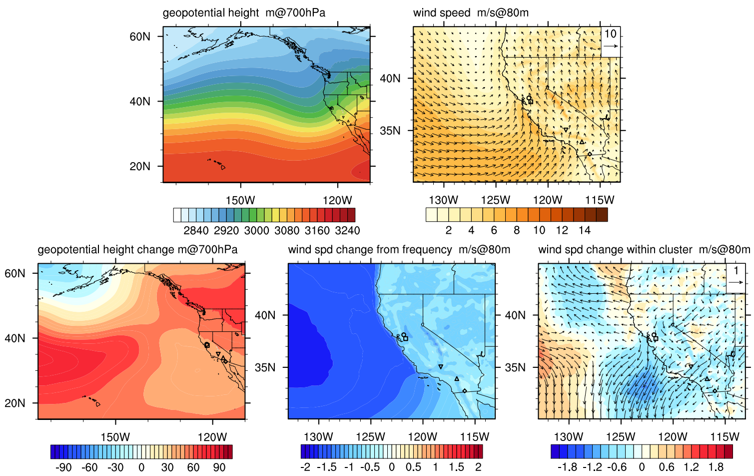

Fig. 10 As Figure 3 but for SC domain cluster 5.

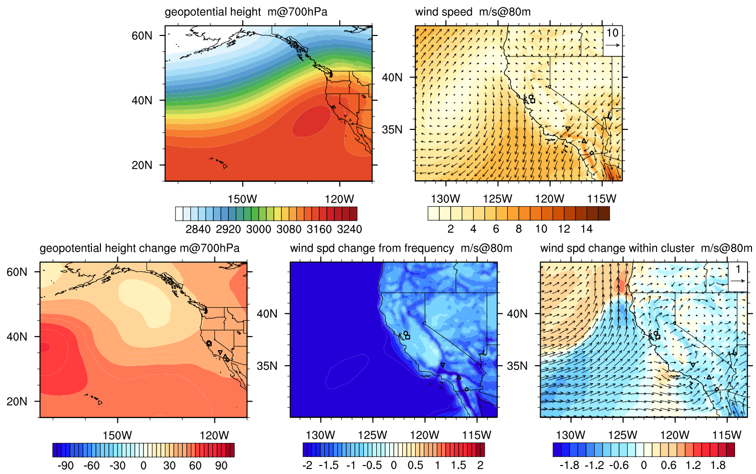

Fig. 11 As Figure 3 but for SC domain cluster 6.

3.2.7 SC 6: Less frequent and weaker Santa Ana winds in fall/winter

The second largest change in cluster frequency for the SC domain occurs in cluster 6, which is 2.3% less frequent by end-of-century. The synoptic fields for these days is depicted in Figure 11, and corresponds to a typical wind pattern from Santa Ana events (Raphael, 2003; Westerling et al, 2004; Li et al, 2016; Millstein et al, 2019; Guzman-Morales and Gershunov, 2019). The relatively high 700hPa geopotential height field over the western US, along with the high center sitting off-shore, leads to the northeasterly wind field throughout the SC region. The end-of-century change in 700hPa geopotential height field indicates a weakening of the onshore ridge, in turn producing slightly weaker winds during Santa Ana events. The decrease in cluster frequency around Fall season is also consistant with findings from Miller and Schlegel (2006), where decreasing frequency of Santa Ana occurrence was also projected in early Fall through the end-of-century.

3.2.8 SC 7: More frequent and less seasonal weakened onshore flow

SC cluster 7, which corresponds to weakened onshore flow in the summer and fall seasons, also shows a significant increase in frequency by 2.0%. The synoptic-scale fields of this cluster are depicted in Figure 12. By the end-of-century, the high 700hPa geopotential height anomaly center sitting offshore to the California coast acts to increase the northerly flow parallel to the coastline in Northern California, and blocks northerly flow in SC domain. This leads to a weakening of the offshore flow throughout the SC domain.

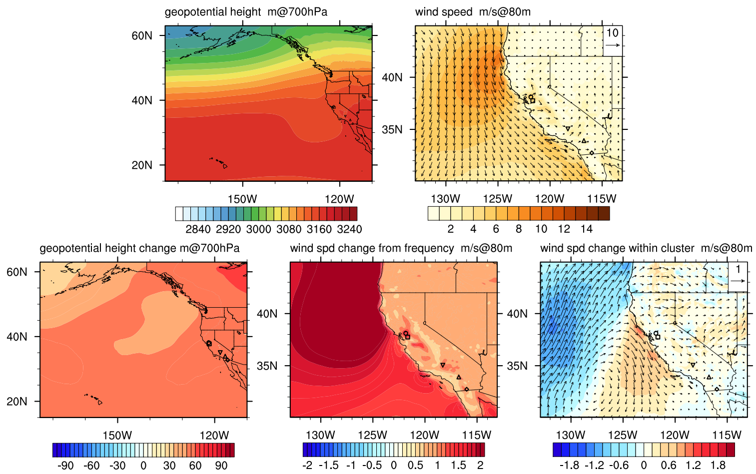

Fig. 12 As Figure 3 but for SC domain cluster 7.

3.2.9 SC 8: Less frequent westerly wind in winter/spring

SC cluster 8 represents a steady westerly marine flow directed onshore (Figure 13), and appears most prominently in the winter season. This cluster is less frequent (7.2%) but has been projected to decrease by 1.7% in its frequency under end-of-century, with most of the decrease occurring in winter and spring. Similar to the previously described clusters, the 700hPa geopotential height field in cluster 8 is also increasing, although with a magnitude that is reduced over the area centered around the offshore region near Baja California. The net result of this change in the geopotential height field is a reduced wind field throughout the whole California, and also a reduction in onshore marine flow. Consequently the changes in this cluster produce a reduction in wind speeds throughout the SC domain.

3.3 Trends in wind energy production

In this section, projected changes in wind energy production are considered in light of the cluster analysis. Before proceeding, we first assess projected changes in wind energy production from model output. Wind fields from VR-CESM runs were interpolated to each wind plant location so as to directly compute wind energy capacity factor (CF in %) changes between historical and end-of-century (details of this calculation can be found in supplement material Section 2). Before calculating CF based on the wind fields from VR-CESM, a constant bias correction factors of 1.3 (Section 2.1) was applied to

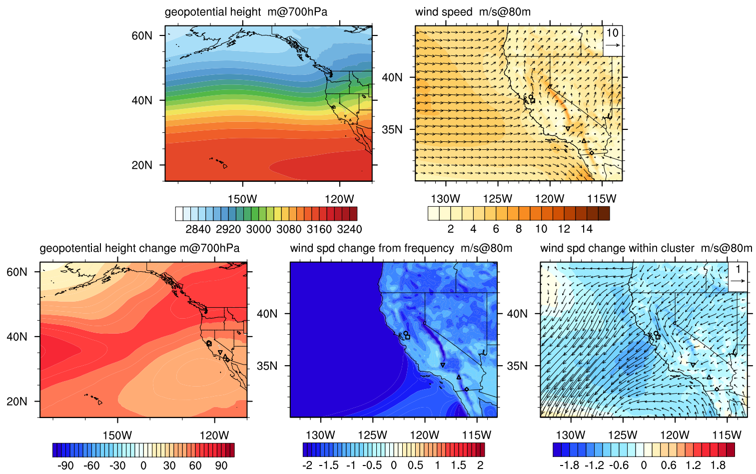

Fig. 13 As Figure 3 but for SC domain cluster 8.

the wind fields to reduce the low wind speed bias from VR-CESM. Then CF were calculated from the bias-corrected wind fields. Table 2 through 8 are all based on the bias-corrected CF values. CFs are commonly defined as actual power output divided by the maximum wind power output that can be generated through the wind turbine system. The relationship between wind speed and CF is nonlinear, and is calculated via different characteristic power curves at each wind plant location (see supplement), and do not include electrical losses during the power generation process. Table 2 lists overall seasonal and annual CF differences at each location without using the clustering methodology. Percentage changes in the lowermost table are calculated with end-of-century CF minus historical CF, divided by historical CF, and written as a percentage change by multiplying 100. Overall, CFs are observed to increase in summer season (JJA), whereas winter (DJF) seasons exhibit a CF decrease. Here the overall seasonal trends from end-of-century during JJA and DJF are consistent with mid-century trends reported in Wang et al (2018), but with an increased magnitude. CF changes based on the original wind fields (without bias correction) are given in section 3 in supplement.

Our goal is to now explain the statistically significant CF changes observed in Table 2. In each of the following subsections we decompose the CF from each wind plant into the contribution from each cluster, and further decompose the change in CF into frequency changes and within-cluster changes following section 2.3. Namely, we apply

$$\Delta CF = \sum_i \underbrace{CF_i \Delta f_i}_{(a)} + \underbrace{(CF_i^e - CF_i^h) f_i}_{(b)} + h.o.t., \tag{5}$$

Table 2 Historical seasonal and annual capacity factor (%) (upper table), absolute change in capacity factors (middle table), and percentage capacity factors changes under end-of-century comparing to historical (lower table) at each wind plant sites across California. Absolute changes are calculated with end-of-century CF minus historical CF. Percentage changes are calculated with end-of-century CF minus historical CF, divided by historical CF, and multiplied by 100 to write as percentages. Shiloh and Altamont Pass are located in NC domain, and the other three wind plants are in SC domain. All CF values are based on bias-corrected wind fields from VR-CESM.

Boldface indicates a percent change above the 95% significance level.

Wind plant	MAM	JJA	SON	DJF	Annual
Shiloh	33.45	50.41	30.60	27.47	35.53
Altamont Pass	23.84	40.67	19.22	14.11	24.52
Alta	44.43	40.02	34.25	38.75	39.38
San Gorgonio	19.87	23.59	12.70	11.77	17.02
Ocotillo	37.06	39.82	20.67	12.09	27.50

Wind plant	MAM	JJA	SON	DJF	Annual
Shiloh	+ 0.98	**+ 2.44**	**- 1.65**	**- 3.68**	- 0.46
Altamont Pass	**+ 1.63**	**+ 3.81**	+ 0.39	**- 1.36**	**+ 1.13**
Alta	- 1.54	+ 1.02	**- 5.29**	**- 3.67**	**- 2.35**
San Gorgonio	+ 0.10	**+ 1.91**	**- 1.32**	**- 2.14**	- 0.35
Ocotillo	+ 1.21	**+ 3.57**	- 1.33	- 0.47	+ 0.76

Wind plant	MAM	JJA	SON	DJF	Annual
Shiloh	+ 2.92%	**+ 4.84%**	**- 5.39%**	**- 13.39%**	- 1.29%
Altamont Pass	**+ 6.82%**	**+ 9.37%**	+ 2.04%	**- 9.65%**	**+ 4.62%**
Alta	- 3.46%	+ 2.54%	**- 15.44%**	**- 9.47%**	**- 5.98%**
San Gorgonio	+ 0.52%	**+ 8.09%**	**- 10.37%**	**- 18.14%**	- 2.04%
Ocotillo	+ 3.27%	**+ 8.97%**	- 6.42%	- 3.89%	+ 2.77%

where CF_i^h and CF_i^e are the historical and end-of-century average CF for cluster i and $CF_i = (CF_i^h + CF_i^e)/2$. Here $h.o.t.$ denotes higher-order terms that are negligible in the decomposition.

3.3.1 NC JJA (Shiloh and Altamont Pass)

Both NC wind plant locations experience a significant increase in JJA CF, driven by essentially two factors. First, from Table 1 we see that there is a significant reduction in the frequency of low wind days (NC 4 as in supplement Figure 6), and an accompanying increase in summertime MAP days (NC 6 and NC 8 as in supplement Figure 9). Second, there is a significant increase in the wind speeds on MAP days (NC 6, 7, and 8), as explained in section 3.2.3 – in fact, the increase in wind speeds actually compensates for a reduced frequency of the NC 7 cluster (supplement Figure 8) of MAP days. Table 3 identifies the 6 clusters responsible for 98.1% and 98.6% of the historical wind energy production for Shiloh and Altamont Pass.

Table 3 Historical mean CF in select clusters (CF_i^h)(%), historical contribution to total seasonal CF ($CF_i^h f_i^h$), end-of-century CF change due to changes in cluster frequency (ΔCF (a)), and within-cluster change in wind speeds (ΔCF (b)) for the NC JJA season. Boldface in the (ΔCF (a)) column indicates clusters with significant change in frequency (see Table 1). Boldface in the (ΔCF (b)) column indicates a significant within-cluster change in CF at the 95% significance level obtained from t-statistics. The values in the "Total" row indicate how much total CF and CF change is attributed to this subset of clusters (compared to Table 2).

NC JJA (top 6 clusters)					
Cluster	Wind plant	CF_i^h	$CF_i^h f_i^h$	ΔCF (a)	ΔCF (b)
4	Shiloh	36.79	2.13	**- 1.55**	+ 0.12
	Altamont Pass	26.80	1.55	**- 1.16**	+ 0.14
5	Shiloh	53.71	8.45	- 0.46	**- 0.34**
	Altamont Pass	25.39	3.99	- 0.22	**- 0.12**
6	Shiloh	52.11	9.95	**+ 3.25**	+ 0.31
	Altamont Pass	49.27	9.41	**+ 3.10**	+ 0.44
7	Shiloh	47.51	12.96	**- 1.52**	**+ 0.80**
	Altamont Pass	52.10	14.21	**- 1.70**	**+ 1.32**
8	Shiloh	60.09	10.08	**+ 2.00**	**+ 0.05**
	Altamont Pass	38.12	6.39	**+ 1.34**	**+ 0.86**
10	Shiloh	45.58	5.87	+ 0.15	**+ 0.32**
	Altamont Pass	35.14	4.53	+ 0.11	**+ 0.09**
Total	Shiloh		49.45	+ 1.85	+ 1.27
	Altamont Pass		40.09	+ 1.47	+ 2.74

3.3.2 NC SON (Shiloh)

In accordance with Table 1, there is a decrease in the frequency of NC 1 and 3, associated with westerly wind and blocked offshore wind, and a compensating increase in the frequency of NC 6, 8, and 9, corresponding to MAP days and low southerly wind. As discussed in sections 3.2.1 and 3.2.2 inhomogeneity in the changing geopotential field has the further effect of reducing the wind speeds within the NC 1 and NC 3 clusters, further driving down CFs. Curiously, Altamont Pass does not experience a corresponding decrease in total CF, as historical CF at this wind plant during NC 1 and NC 3 days are much lower than NC 6 and NC 8 (supplement Figure 9) and so the shifting cluster frequencies actually drive up average CF. Unlike the summer and winter seasons, the transitional fall and spring seasons do not feature a prominent subset of wind clusters. However, low wind days (NC 4 as in supplement Figure 6) are much more likely to occur in the future during these seasons – we thus see that Shiloh is projected to see a decrease in CF in the fall. The breakdown of the contributions from the six most prominent clusters to Shiloh's CF is given in Table 4, which accounts for 72.8% of the wind energy production for this season. However, changes in these six clusters effectively explain the observed change in wind speed in this season.

Table 4 As Table 3, except for NC SON.

NC SON (top 6 clusters)					
Cluster	Wind plant	CF_i^h	$CF_i^h f_i^h$	ΔCF (a)	ΔCF (b)
1	Shiloh	24.66	3.87	**- 1.12**	**- 0.73**
	Altamont Pass	16.39	2.57	**- 0.74**	**- 0.51**
2	Shiloh	38.15	2.67	- 0.39	- 0.35
	Altamont Pass	22.07	1.55	- 0.22	- 0.24
3	Shiloh	38.49	5.84	**- 2.33**	- 0.49
	Altamont Pass	13.76	2.09	**- 0.78**	- 0.36
6	Shiloh	37.67	4.42	**+ 1.42**	- 0.15
	Altamont Pass	33.97	3.98	**+ 1.30**	- 0.01
8	Shiloh	43.05	4.53	**+ 1.33**	- 0.32
	Altamont Pass	25.53	2.68	**+ 0.82**	+ 0.05
9	Shiloh	13.95	0.93	**+ 0.29**	- 0.03
	Altamont Pass	7.77	0.52	**+ 0.16**	+ 0.04
Total	Shiloh		22.27	- 0.80	- 2.06
	Altamont Pass		13.40	+ 0.53	-1.12

Table 5 As Table 3, except for NC DJF.

NC DJF (top 5 clusters)					
Cluster	Wind plant	CF_i^h	$CF_i^h f_i^h$	ΔCF (a)	ΔCF (b)
1	Shiloh	19.96	4.08	+ 0.16	- 0.29
	Altamont Pass	12.24	2.50	+ 0.10	- 0.07
2	Shiloh	48.93	11.98	- 0.62	**- 1.47**
	Altamont Pass	27.62	6.76	- 0.34	**- 1.14**
3	Shiloh	27.14	5.90	**- 1.05**	- 0.05
	Altamont Pass	8.54	1.85	**- 0.34**	+ 0.05
4	Shiloh	11.32	1.06	+ 0.16	**- 0.25**
	Altamont Pass	4.97	0.47	+ 0.08	**- 0.02**
9	Shiloh	19.07	3.72	+ 0.29	**- 0.74**
	Altamont Pass	10.12	1.98	+ 0.16	**- 0.07**
Total	Shiloh		26.74	- 1.06	- 2.80
	Altamont Pass		13.56	- 0.34	- 1.24

3.3.3 NC DJF (Shiloh and Altamont Pass)

Both wind plants experience a significant decline in total CF over this season. The observed change can be largely attributed to a decrease in the frequency of NC 2 and NC 3 (strong westerly wind and blocked offshore wind), which have the highest average CF at Shiloh, and an increase in the frequency of NC 1, 4, and 9 clusters, which are each associated with lower wind speeds and CF. There is further a significant decrease in the wind speeds of cluster NC 2, the most frequent wintertime pattern, as described in section 3.2.1 to be attributed to higher overland pressures. NC wintertime is associated with 5 clusters that describe 97.4% and 96.1% of total seasonal wind energy productions at Shiloh and Altamont Pass, respectively.

Table 6 As Table 3, except for SC JJA.

SC JJA (top 3 clusters)					
Cluster	Wind plant	CF_i^h	$CF_i^h f_i^h$	ΔCF (a)	ΔCF (b)
2	San Gorgonio	19.15	8.78	- 1.33	+ 1.34
	Ocotillo	33.00	15.13	- 2.26	+ 1.89
4	San Gorgonio	32.99	10.16	**+ 1.73**	- 0.19
	Ocotillo	56.36	17.36	**+ 2.99**	+ 0.16
7	San Gorgonio	19.39	3.48	**+ 0.37**	**+ 0.01**
	Ocotillo	29.36	5.27	**+ 0.58**	**+ 0.40**
Total	San Gorgonio		22.42	+ 0.77	+ 1.15
	Ocotillo		37.76	+ 1.31	+ 2.45

3.3.4 SC JJA (San Gorgonio and Ocotillo)

These two wind plants experience a pronounced increase in CF over this season attributed to two factors. First, a strengthening of the onshore flow (when it occurs) that leads to a reclassification of SC 2 days (weak onshore flow) (supplement Figure 12) to SC 4 and SC 7(onshore flow) days (Table 1). Second, an increase in the overall strength of SC 2 (supplement Figure 12) days when they do occur and SC 7 days, generally associated with an increase in onshore flow speeds associated with a stronger land/sea temperature gradient. The three clusters in Table 6 describe 97.1% and 96.9% of total JJA wind energy productions for San Gorgonio and Ocotillo, respectively.

3.3.5 SC SON (Alta and San Gorgonio)

Wind speeds are projected to decrease throughout the SC domain in the fall season leading to a significant decrease in CF at Alta and San Gorgonio. As observed in Table 7 this can be attributed to a widespread drop in wind speeds within essentially all clusters. This is accompanied by a significant drop in frequency of SC 1 (strong alongshore winds) and SC 6 (Santa Ana winds) and accompanying increase in SC 7 (weak onshore wind) and SC 9 (low wind) (supplement Figure 14) – whereas SC 1 and SC 6 days correspond to the highest and third-highest CFs, SC 7 and SC 9 (supplement Figure 14) are the lowest and third lowest producers.

3.3.6 SC DJF (Alta and San Gorgonio)

As in the NC region, overland warming across SC leads to a widespread weakening of the within-cluster winds and a reduction in CF across the board. This process further drives an increase in the frequency of SC 3 (low wind) (supplement Figure 13), which is associated with one of the lowest CF values, at the expense of SC 6 (Santa Ana winds) and SC 8 (westerly winds), which have among the highest CF values. There is further a substantial drop in the within-cluster wind speeds of SC 5 (southwesterly winds), as explained in sec-

Table 7 As Table 3, except for SC SON.

SC SON (top 7 clusters)					
Cluster	Wind plant	CF_i^h	$CF_i^h f_i^h$	ΔCF (a)	ΔCF (b)
1	Alta	61.71	8.20	**- 2.10**	- 0.45
	San Gorgonio	15.77	2.10	**- 0.56**	+ 0.03
2	Alta	38.25	8.19	+ 0.71	**- 1.08**
	San Gorgonio	11.75	2.51	+ 0.23	- 0.11
3	Alta	19.32	3.19	- 0.22	**- 0.71**
	San Gorgonio	4.89	0.81	- 0.06	**- 0.15**
6	Alta	43.08	4.49	**- 1.90**	- 0.05
	San Gorgonio	18.03	1.88	**- 0.74**	- 0.22
7	Alta	16.16	1.24	**+ 0.72**	+ 0.22
	San Gorgonio	7.03	0.54	**+ 0.32**	+ 0.12
8	Alta	40.18	1.98	- 0.09	**- 0.37**
	San Gorgonio	16.89	0.83	- 0.04	- 0.14
9	Alta	22.25	1.97	**+ 0.58**	**- 0.38**
	San Gorgonio	7.93	0.70	**+ 0.19**	**- 0.26**
Total	Alta		29.26	- 2.30	- 2.81
	San Gorgonio		9.37	- 0.66	- 0.72

Table 8 As Table 3, except for SC DJF.

SC DJF (top 6 clusters)					
Cluster	Wind plant	CF_i^h	$CF_i^h f_i^h$	ΔCF (a)	ΔCF (b)
1	Alta	55.26	10.14	+ 0.54	+ 0.06
	San Gorgonio	13.97	2.56	+ 0.13	- 0.24
3	Alta	19.31	4.12	**+ 1.00**	- 0.48
	San Gorgonio	4.20	0.90	**+ 0.21**	- 0.19
5	Alta	43.82	4.67	- 0.65	**- 1.02**
	San Gorgonio	9.31	0.99	- 0.14	**- 0.21**
6	Alta	41.27	8.73	**- 1.23**	- 0.37
	San Gorgonio	18.27	3.86	**- 0.52**	- 0.44
8	Alta	39.31	5.06	**- 1.05**	- 0.39
	San Gorgonio	13.12	1.69	**- 0.32**	- 0.38
9	Alta	19.46	0.44	**+ 0.20**	-0.09
	San Gorgonio	3.48	0.08	**+ 0.04**	+ 0.03
Total	Alta		33.16	- 1.19	- 2.29
	San Gorgonio		11.22	- 0.60	- 1.43

tion 3.2.6. Table 8 identifies the six clusters responsible for 85.6% and 85.7% of wind energy productions at Alta and San Gorgonio, respectively.

4 Discussion and Summary

This study utilized the state-of-the-art climate model CESM in its variable-resolution configuration to analyze California wind patterns change under the future climate. The agglomerative clustering algorithm was applied to the climate model output to group different weather patterns into separate clusters within the NC and SC domains. We defined ten wind clusters from each domain, and analyzed changes to within-cluster wind speeds and also changes to the frequency of occurrence of each cluster by the end-of-century. Additionally, we analyzed the synoptic-scale patterns that accompany each cluster. The changes to these patterns can then be used to identify some of the causes of changes to within-cluster wind speeds. Moreover, some of these synoptic scale changes (e.g., changes to the land – sea temperature contrast) are directly tied to global warming, which allows us to tie a specific portion of the forecasted future change in wind resources directly to identified climate change phenomena.

Below we list the most important changes we observe to clusters by the end-of-century.

4.1 Northern California

Westerly winds (NC 1 and NC 2): These two clusters are among the most frequent winter season cluster, and have been projected to become less frequent with lower within-cluster wind speed. The reduction in within-cluster wind speed is associated with the change in geopotential height field over the Pacific, and overland warming under the future climate. Both factors contribute to the decrease in within-cluster wind speed.

Offshore blocking (NC 3): This is another wintertime cluster with a projected decreasing frequency and weaker within-cluster wind speeds. The latter is related to the change in geopotential height pattern, driving a weaker northerly flow offshore, thus leading to weaker within-cluster wind speeds.

Marine air penetration (NC 6-8 and NC 10): These clusters peak in frequencies during summertime. All have been projected to become more frequent with stronger within-cluster wind speeds. The increase in within-cluster wind speeds is associated with changes in the geopotential height pattern, which leads to a weakening of the offshore northerly wind, and promoting the onshore flow pattern. This increase in wind speeds contributes to the projected greater wind power during the summer season.

4.2 Southern California

Strong alongshore wind (SC 1): This cluster produced some of the highest capacity factors due to its frequent occurrences in all seasons only except summer, and its high within-cluster wind speed. It has been projected to become less frequent during spring and fall seasons, and more frequent in the winter season. For within-cluster wind speeds change, the change in the geopotential height field pattern reduces the alongshore gradient, leading to a weaker alongshore flow, and a decrease in wind speeds statewide.

Marine air penetration (SC 4): This cluster peaks in frequency during summertime. It has been projected to become more frequent with slightly increased onshore winds. The latter is caused by the increase in the geopotential height pattern which drives up wind speeds offshore, creating a better ventilation condition.

Santa Ana winds (SC 6): This is the second most frequent wintertime cluster, and has been projected to decrease in frequency with weaker within-cluster wind speeds. This reduction of the within-cluster wind speeds during Santa Ana events is associated with the weakening of the onshore ridge during end-of-century.

Weakened onshore flow (SC 7): This cluster is the third most frequent summertime cluster, with a projected increase in frequency. Under end-of-century, the geopotential height anomaly acts to strengthen the northerly wind offshore in Northern California, while blocks the offshore flow in Southern California.

Westerly wind (SC 8): This is a prominent cluster during winter and spring seasons, and its frequencies during these two season both decrease under end-of-century, along with weaker within-cluster wind speeds. The latter is driven by large-scale dynamical changes that cause a weakening of wind speeds across California, including suppressed onshore flow in Southern California.

4.3 Changes in capacity factor

Along with changes to cluster frequency and within-cluster wind speeds, we found statistically significant changes to energy generation (specifically to estimated capacity factor, or CF) at all wind plants.

There is an increase in the within-cluster wind speeds during JJA driven by an increase land/sea temperature contrast and a subsequent tendency towards more frequent marine air penetration events for both NC and SC. This increasing frequency in marine air penetration events is accompanied by a frequency decrease from NC 4 (low wind) (supplement Figure 6) and SC 2 (weak onshore flow) (supplement Figure 12). Therefore, beside the within-cluster wind speed increase, this frequency shift from low wind cluster to high wind clusters further contributes to the capacity factors increase during summertime.

This pattern is reversed in the winter season, with a smaller land/sea contrast that contributes to a decrease in within-cluster wind speeds in both NC and SC. During the winter season, we observe an overland warming, that leads to an increase in the geopotential height field, and decrease in wind speeds statewide. The 700hPa geopotential height over Northern Pacific decreases in winter. This change in the general circulation also contributes to the wind speed decrease in winter. There is also a clusters frequency shift from high wind speed clusters to low wind speed clusters during winter season for both two domains (a frequency shift from NC 2 and NC 3 to NC 1, NC 4 (supplement Figure 6) and NC9 in the NC domain, and from SC 6 and SC 8 to SC 3 (supplement Figure 13) in the SC domain). So both the cluster frequency changes, and the within-cluster wind speed changes contribute to the decrease in capacity factors during the winter season.

The overall seasonal CF trends in JJA and DJF from the end-of-century were consistent with the trends from the mid-century (Wang et al, 2018), though the magnitudes of the changes are larger. Findings from this study are also consistent with the increasing frequency of marine air penetration events from Wang and Ullrich (2017), decreasing wind speed during fall and winter seasons from Duffy et al (2014), and decreasing frequency of Santa Ana winds during early fall from Miller and Schlegel (2006).

Much of the forecasted change to wind resources is linked to changing frequency of weather patterns or clusters. The changes to frequency of each cluster type is tied to global circulation patterns, and possibly to climate modes and other teleconnections. Determining the specific mechanisms that cause the shifts to the cluster frequency is therefore out of scope within this study, but remains an intriguing target for future work.

Overall, this study provides a statistical approach to group different wind patterns without requiring prior knowledge of various wind types. The synoptic analysis of wind clusters also improves our understanding of the variability of California wind resources by the end-of-century. Future work may focus on associating the wind speed changes with global teleconnection centers and low-frequency patterns, and investigate the causes of change in cluster frequencies, which consequently would improve the predictability of wind power in California. Potential future study can also focus on developing a machine learning model for wind energy forecasting based on meteorological fields.

Acknowledgements The authors want to thank the Craig Collier, Daran Rife, and Christopher Hayes for the helpful conversations throughout this project. We would further like to thank the two anonymous reviewers for their thorough evaluation of the manuscript and valuable suggestions. This work was funded by the California Energy Commission under the Electric Program Investment Charge Grant, EPC-15-068: Understanding and Mitigating Barriers to Wind Energy Expansion in California. This work was supported by the U.S. Department of Energy (DOE) under Lawrence Berkeley National Laboratory Contract No. DE-AC02-05CH11231. This project is further supported by the National Institute of Food and Agriculture, U.S. Department of Agriculture, hatch project under California Agricultural Experiment Station project CA-DLAW-2203-H. Author Ullrich is supported by Department of Energy Office of Science award number DE-SC0016605, An Integrated Evaluation of the Simulated Hydroclimate System of the Continental US.

References

Beaver S, Palazoglu A (2006) Cluster analysis of hourly wind measurements to reveal synoptic regimes affecting air quality. Journal of Applied Meteorology and Climatology 45(12):1710–1726

Berg N, Hall A, Capps SB, Hughes M (2013) El niño-southern oscillation impacts on winter winds over southern california. Climate dynamics 40(1-2):109–121

Breslow PB, Sailor DJ (2002) Vulnerability of wind power resources to climate change in the continental united states. Renewable Energy 27(4):585–598

Caliński T, Harabasz J (1974) A dendrite method for cluster analysis. Communications in Statistics-theory and Methods 3(1):1–27

Conil S, Hall A (2006) Local regimes of atmospheric variability: A case study of southern california. Journal of Climate 19(17):4308–4325

Dennis JM, Edwards J, Evans KJ, Guba O, Lauritzen PH, Mirin AA, St-Cyr A, Taylor MA, Worley PH (2012) CAM-SE: A scalable spectral element dynamical core for the Community Atmosphere Model. The International Journal of High Performance Computing Applications 26(1):74–89

Duffy PB, Bartlett J, Dracup J, Freedman J, Madani K, Waight K (2014) Climate Change Impacts on Generation of Wind, Solar, and Hydropower in California. California Energy Commission (2014) CEC5002014111

Fosberg MA, Schroeder MJ (1966) Marine air penetration in central California. Journal of Applied Meteorology 5(5):573–589

Gates WL (1992) An ams continuing series: Global change–amip: The atmospheric model intercomparison project. Bulletin of the American Meteorological Society 73(12):1962–1970

Ghan SJ, Liu X, Easter RC, Zaveri R, Rasch PJ, Yoon JH, Eaton B (2012) Toward a minimal representation of aerosols in climate models: Comparative decomposition of aerosol direct, semidirect, and indirect radiative forcing. Journal of Climate 25(19):6461–6476

Guba O, Taylor MA, Ullrich PA, Overfelt JR, Levy MN (2014) The spectral element method on variable resolution grids: Evaluating grid sensitivity and resolution-aware numerical viscosity. Geosci Model Dev Discuss 7:4081–4117

Guzman-Morales J, Gershunov A (2019) Climate change suppresses santa ana winds of southern california and sharpens their seasonality. Geophysical Research Letters 46(5):2772–2780

Hartigan JA, Wong MA (1979) Algorithm as 136: A k-means clustering algorithm. Journal of the Royal Statistical Society Series C (Applied Statistics) 28(1):100–108

Haupt SE, Copeland J, Cheng WY, Zhang Y, Ammann C, Sullivan P (2016) A method to assess the wind and solar resource and to quantify interannual variability over the united states under current and projected future climate. Journal of Applied Meteorology and Climatology 55(2):345–363

Hoen B, Diffendorfer J, Rand J, Kramer L, Garrity C, Hunt H (2019) United states wind turbine database. u.s. geological survey, american wind energy association, and lawrence berkeley national laboratory data release: Uswtdb v2.1. United States Wind Turbine Database

Huang X, Ullrich PA (2017) The changing character of twenty-first-century precipitation over the western united states in the variable-resolution cesm. Journal of Climate 30(18):7555–7575

Huang X, Rhoades AM, Ullrich PA, Zarzycki CM (2016) An evaluation of the variable-resolution CESM for modeling California's climate. Journal of Advances in Modeling Earth Systems 8(1):345–369

Hurrell JW, Hack JJ, Shea D, Caron JM, Rosinski J (2008) A new sea surface temperature and sea ice boundary dataset for the community atmosphere model. Journal of Climate 21(19):5145–5153

Hurrell JW, Holland MM, Gent PR, Ghan S, Kay JE, Kushner PJ, Lamarque JF, Large WG, Lawrence D, Lindsay K, et al (2013) The community earth system model: a framework for collaborative research. Bulletin of the American Meteorological Society 94(9):1339–1360

Iacono MJ, Delamere JS, Mlawer EJ, Shephard MW, Clough SA, Collins WD (2008) Radiative forcing by long-lived greenhouse gases: Calculations with the aer radiative transfer models. Journal of Geophysical Research: Atmospheres 113(D13)

Jin L, Harley RA, Brown NJ (2011) Ozone pollution regimes modeled for a summer season in californias san joaquin valley: A cluster analysis. Atmospheric environment 45(27):4707–4718

Justus C, Mikhail A (1976) Height variation of wind speed and wind distributions statistics. Geophysical Research Letters 3(5):261–264

Karnauskas KB, Lundquist JK, Zhang L (2018) Southward shift of the global wind energy resource under high carbon dioxide emissions. Nature Geoscience 11(1):38

Kay JE, Deser C, Phillips A, Mai A, Hannay C, Strand G, Arblaster JM, Bates S, Danabasoglu G, Edwards J, et al (2015) The community earth system model (cesm) large ensemble project: A community resource for studying climate change in the presence of internal climate variability. Bulletin of the American Meteorological Society 96(8):1333–1349

Li AK, Paek H, Yu JY (2016) The changing influences of the amo and pdo on the decadal variation of the santa ana winds. Environmental Research Letters 11(6):064,019

Ludwig FL, Horel J, Whiteman CD (2004) Using eof analysis to identify important surface wind patterns in mountain valleys. Journal of Applied Meteorology 43(7):969–983

Miller NL, Schlegel NJ (2006) Climate change projected fire weather sensitivity: California Santa Ana wind occurrence. Geophysical Research Letters 33(15)

Millstein D, Solomon-Culp J, Wang M, Ullrich P, Collier C (2018) Wind energy variability and links to regional and synoptic scale weather. Climate Dynamics pp 1–16

Millstein D, Solomon-Culp J, Wang M, Ullrich P, Collier C (2019) Wind energy variability and links to regional and synoptic scale weather. Climate Dynamics 52(7-8):4891–4906

Morrison H, Gettelman A (2008) A new two-moment bulk stratiform cloud microphysics scheme in the community atmosphere model, version 3 (cam3). part i: Description and numerical tests. Journal of Climate 21(15):3642–3659

Neale RB, Richter JH, Jochum M (2008) The impact of convection on enso: From a delayed oscillator to a series of events. Journal of climate 21(22):5904–5924

Neale RB, Chen CC, Gettelman A, Lauritzen PH, Park S, Williamson DL, Conley AJ, Garcia R, Kinnison D, Lamarque JF, et al (2010) Description of the NCAR community atmosphere model (CAM 5.0). NCAR Tech Note NCAR/TN-486+ STR

Olauson J, Edström P, Rydén J (2017) Wind turbine performance decline in sweden. Wind Energy 20(12):2049–2053

Oleson KW, Lawrence DM, Gordon B, Flanner MG, Kluzek E, Peter J, Levis S, Swenson SC, Thornton E, Feddema J, et al (2010) Technical description of version 4.0 of the community land model (clm)

Park S, Bretherton CS (2009) The university of washington shallow convection and moist turbulence schemes and their impact on climate simulations with the community atmosphere model. Journal of Climate 22(12):3449–3469

Park S, Bretherton CS, Rasch PJ (2014) Integrating cloud processes in the community atmosphere model, version 5. Journal of Climate 27(18):6821–6856

Pryor S, Barthelmie R (2010) Climate change impacts on wind energy: A review. Renewable and sustainable energy reviews 14(1):430–437

Pryor S, Barthelmie R (2011) Assessing climate change impacts on the near-term stability of the wind energy resource over the united states. Proceedings of the National Academy of Sciences

Raphael M (2003) The santa ana winds of california. Earth Interactions 7(8):1–13

Rasmussen D, Holloway T, Nemet G (2011) Opportunities and challenges in assessing climate change impacts on wind energy – a critical comparison of wind speed projections in California. Environmental Research Letters 6(2):024,008

Rhoades AM, Huang X, Ullrich PA, Zarzycki CM (2016) Characterizing Sierra Nevada snowpack using variable-resolution CESM. Journal of Applied Meteorology and Climatology 55(1):173–196

Rhoades AM, Ullrich PA, Zarzycki CM (2018a) Projecting 21st century snowpack trends in western usa mountains using variable-resolution cesm. Climate Dynamics 50(1-2):261–288

Rhoades AM, Ullrich PA, Zarzycki CM, Johansen H, Margulis SA, Morrison H, Xu Z, Collins WD (2018b) Sensitivity of mountain hydroclimate simulations in variable-resolution cesm to microphysics and horizontal resolution. Journal of Advances in Modeling Earth Systems 10(6):1357–1380

Small RJ, Bacmeister J, Bailey D, Baker A, Bishop S, Bryan F, Caron J, Dennis J, Gent P, Hsu Hm, et al (2014) A new synoptic scale resolving global climate simulation using the community earth system model. Journal of Advances in Modeling Earth Systems 6(4):1065–1094

Staffell I, Pfenninger S (2016) Using bias-corrected reanalysis to simulate current and future wind power output. Energy 114:1224–1239

Taylor KE, Stouffer RJ, Meehl GA (2012) An overview of CMIP5 and the experiment design. Bulletin of the American Meteorological Society 93(4):485

Taylor MA, Fournier A (2010) A compatible and conservative spectral element method on unstructured grids. Journal of Computational Physics 229(17):5879–5895

Ullrich PA (2014) SQuadGen: Spherical quadrilateral grid generator. University of California, Davis, Climate and Global Change Group software URL http://climate.ucdavis.edu/squadgen.php

Wang H, Schubert SD (2014) Causes of the extreme dry conditions over california during early 2013

Wang M, Ullrich PA (2017) Marine air penetration in California's Central Valley: Meteorological drivers and the impact of climate change. Journal of Applied Meteorology and Climatology (2017)

Wang M, Ullrich PA, Millstein D (2018) The future of wind energy in california: Future projections with the variable-resolution cesm. Renewable Energy 127:242–257

Ward Jr JH (1963) Hierarchical grouping to optimize an objective function. Journal of the American statistical association 58(301):236–244

Westerling AL, Cayan DR, Brown TJ, Hall BL, Riddle LG (2004) Climate, santa ana winds and autumn wildfires in southern california. Eos, Transactions American Geophysical Union 85(31):289–296

Wilks DS (2011) Cluster analysis. In: International geophysics, vol 100, Elsevier, pp 603–616

Yu L, Zhong S, Bian X, Heilman WE (2015) Temporal and spatial variability of wind resources in the United States as derived from the Climate Forecast System Reanalysis. Journal of Climate 28(3):1166–1183

Zarzycki CM, Jablonowski C (2014) A multidecadal simulation of atlantic tropical cyclones using a variable-resolution global atmospheric general circulation model. Journal of Advances in Modeling Earth Systems 6(3):805–828

Zarzycki CM, Jablonowski C, Taylor MA (2014a) Using variable-resolution meshes to model tropical cyclones in the community atmosphere model. Monthly Weather Review 142(3):1221–1239

Zarzycki CM, Levy MN, Jablonowski C, Overfelt JR, Taylor MA, Ullrich PA (2014b) Aquaplanet experiments using CAMs variable-resolution dynamical core. Journal of Climate 27(14):5481–5503