Linking large-scale circulation patterns to low-cloud properties

Timothy W. Juliano[1] and Zachary J. Lebo[2]

[1]Research Applications Laboratory, National Center for Atmospheric Research, Boulder, CO 80301

[2]Department of Atmospheric Science, University of Wyoming, Laramie, WY 82071

Correspondence: Timothy W. Juliano (tjuliano@ucar.edu)

Abstract. The North Pacific High (NPH) is a fundamental meteorological feature present during the boreal warm season. Marine boundary layer (MBL) clouds, which are persistent in this oceanic region, are influenced directly by the NPH. In this study, we combine 11 years of reanalysis and an unsupervised machine learning technique to examine the gamut of 850-hPa synoptic-scale circulation patterns. This approach, which yields the frequency at which these regimes occur, reveals two

5   distinguishable patterns — a dominant NPH setup and a land-falling cyclone — and in between a spectrum of regimes. We then use satellite retrievals to elucidate for the first time the explicit dependence of MBL cloud properties (namely cloud droplet number concentration and cloud droplet effective radius) on 850-hPa circulation patterns over the northeast Pacific Ocean. Moreover, we find that shortwave cloud radiative forcing ranges from -144.0 to -117.5 W m^{-2}, indicating that the range of MBL cloud properties must be accounted for in global and regional climate models. Our results demonstrate the value of

10   combining reanalysis and satellite observations to help clarify the relationship between synoptic-scale dynamics and cloud microphysics.

1   Introduction

Low, stratiform clouds that develop in the marine boundary layer (MBL) are of significant interest to the atmospheric science community because they impact meteorological forecasts and, ultimately, a host of human activities (e.g., Koraĉin and Dorman,

15   2017). These cloud types are widespread (coverage on the order of one-third of the globe at any given time; e.g., Hartmann et al., 1992) in the subsiding branch of the Hadley circulation (e.g., Wood, 2012) due to a separation of the cool, moist MBL and the warm, dry free troposphere by a strong (∼10 K) and sharp $\mathcal{O}(100-500$ m$)$ thermal inversion (e.g., Parish, 2000). Despite their substantive role in the radiation budget (global shortwave cloud radiative forcing (CRF$_{SW}$) of ∼60–120 W m^{-2}; e.g., Yi and Jian, 2013), MBL clouds and their radiative response to changes in the climate system are not simulated accurately

20   by global climate models (e.g., Palmer and Anderson, 1994; Delecluse et al., 1998; Bachiochi and Krishnamurti, 2000; Bony and Dufresne, 2005; Webb et al., 2006; Lin et al., 2014); however, results from regional climate models are more encouraging (e.g., Wang et al., 2004, 2011).

During boreal summer, the northeast Pacific Ocean is home to one of the largest MBL stratiform cloud decks (e.g., Klein and Hartmann, 1993). Differential heating of land and ocean masses during the warm season leads to the development of the North Pacific High (NPH) and the desert thermal low over the southwest United States. Classical descriptions in the literature often treat the mean summertime location of the NPH to be far offshore (thousands of kilometers) of the western United States coastline. However, several studies have examined NPH strengthening as it meanders toward the north and east (e.g., Mass and Bond, 1996; Fewings et al., 2016; Juliano et al., 2018). These events are typically associated with an increase in the offshore flow and a clearing of the cloud deck (e.g., Kloesel, 1992; Crosbie et al., 2016), and they may lead to a complete reversal of the alongshore pressure gradient (e.g., Nuss et al., 2000).

Often called coastally trapped disturbances (CTDs), these mesoscale phenomena are usually characterized by a redevelopment of the stratiform cloud deck (e.g., Thompson et al., 2005; Parish et al., 2008). Recent work using satellite observations suggests that MBL clouds accompanying CTDs are more polluted [increased cloud droplet number concentration (N) and smaller cloud droplet effective radius (r_e)] than those forming under typical northerly flow conditions due to aerosol-cloud interactions (Juliano et al., 2018). Offshore flow, which is a requirement for the initiation of a CTD, likely enhances the transport of aerosol from the continent to the ocean. These results motivate the present study because in this study we consider data over a relatively long time span to identify objectively the most prevalent synoptic-scale dynamical regimes during boreal summer. We aim to improve the current understanding of the relationship between these synoptic-scale patterns, mesoscale cloud microphysics, and CRF_{SW} over the northeast Pacific Ocean — an issue identified previously as "vital" (Stevens and Feingold, 2009).

To diagnose the various NPH regimes, we first use the NCEP North American Regional Reanalysis (NARR) to develop a self-organizing map (SOM). We then examine measurements from the *Aqua* Moderate Resolution Imaging Spectroradiometer (MODIS). Two important variables — r_e and optical thickness (τ) — are passively retrieved by MODIS. r_e is defined as the ratio of the third moment to the second moment of the cloud droplet size distribution, while optical thickness (τ) is defined as the line integral of the extinction (absorption plus scattering) coefficient between two levels (in this case, the passive sensor and some distance into the cloud deck). Values of N may then be estimated from r_e and τ. For discussion of the MODIS retrievals, we focus on N and r_e because these two variables most clearly accentuate the connection between large-scale dynamics and MBL cloud properties.

2 Methods and data

2.1 Pattern identification

The SOM is a type of neural network that uses a competitive, unsupervised machine learning technique (e.g., Reichstein et al., 2019) to develop a 2-dimensional topology (map) of nodes that represents the *n*-dimensional input data[1]. Because this dimensionality reduction method falls under the category of unsupervised learning, the user does not need to teach the

---

[1]The term "data" is used loosely because the input variable(s) need not be actual data and may be, for example, model output.

algorithm on a separate training data set. Instead, the user is required to provide only a 2-dimensional array, and the node

55 topology organizes itself to mimic the input data. That is, each node will represent a group of similar input vectors.

We use the Matlab SOM Toolbox (version 2.1) to generate the SOM using the batch algorithm. This algorithm follows the well-known Kohonen technique (Kohonen, 1990). The SOM batch training procedure can be described as follows:

1. Define the number of nodes and iterations (one iteration is defined as a single pass through all of the input data vectors), in addition to the neighborhood radius

60 2. Initialize the SOM node weights linearly along its greatest eigenvectors

3. Present all vectors from the input data and calculate the Euclidean distance between each input vector and each node

4. Update the neighborhood radius

5. Determine the node that most closely matches each input vector; the winning node is characterized by the minimum Euclidean distance[2]

65 6. Update the weight of each node[3] after a single iteration

7. Repeat steps 3-6 for n iterations

Choosing the number of nodes is critical because a map with too few nodes yields larger sample sizes but insufficient detail, while one with too many nodes yields greater detail but insufficient sample sizes. For the present study, a series of sensitivity tests is conducted using different node map sizes to determine an optimal number of nodes (Fig. 1). Quantization

70 and topographic errors (QE and TE, respectively) for each map are calculated. The QE, which is a measure of map resolution, is equal to the average distance between each input vector and the best matching node, while the TE indicates map topology preservation by determining the percentage of input vectors whose first and second best matching nodes are not adjacent. As the number of nodes increases, the QE decreases typically at the cost of sacrificing node topology. This trade-off is shown quite well in Fig. 1: the QE decrease is most pronounced as the number of nodes increases from approximately 9 to 20, and the TE

75 increase is most notable above approximately 30 nodes. Moreover, using a nonuniform (rectangular) map appears to reduce the TE, which supports previous work showing the superiority of rectangular maps over square maps (e.g., Ultsch and Herrmann, 2005). Due to the TE minimum at 20 nodes and a relatively marginal decrease in QE after 20 nodes, in addition to ample pattern detail and sufficient sample sizes, for this study we choose to use a 4×5 node map. Moreover, we choose to iterate 5,000 times and use an initial neighborhood radius of 4. The neighborhood radius, which determines the number of nodes surrounding the

80 winning node that nudge toward the input vector, slowly reduces to one (only the winning node is nudged) through the training period. Overall, our choices are similar to and follow guidelines outlined in prior SOM studies that focus on vertical sounding classification problems (e.g., Jensen et al., 2012; Nowotarski and Jensen, 2013; Stauffer et al., 2017) and synoptic meteorology

---

[2]Given two points a and b in $\{x, y\}$ space, the Euclidean distance, $d(a, b) = \sqrt{(x_a - x_b)^2 + (y_a - y_b)^2}$.

[3]The new weight for each node is equal to the weighted average of each input data vector to which that node or any nodes in its neighborhood responded.

pattern recognition (e.g., Cassano et al., 2015; Ford et al., 2015; Mechem et al., 2018). Additionally, we find that changing the initial parameters (iterations and neighborhood radius to 25,000 and 2, respectively) has a relatively small impact on the final

85   node topology similar to other studies (e.g., Cassano et al., 2006; Skific et al., 2009). Once training is complete, and the node topology has organized itself to best represent the input array, each input vector is associated with one of the map nodes.
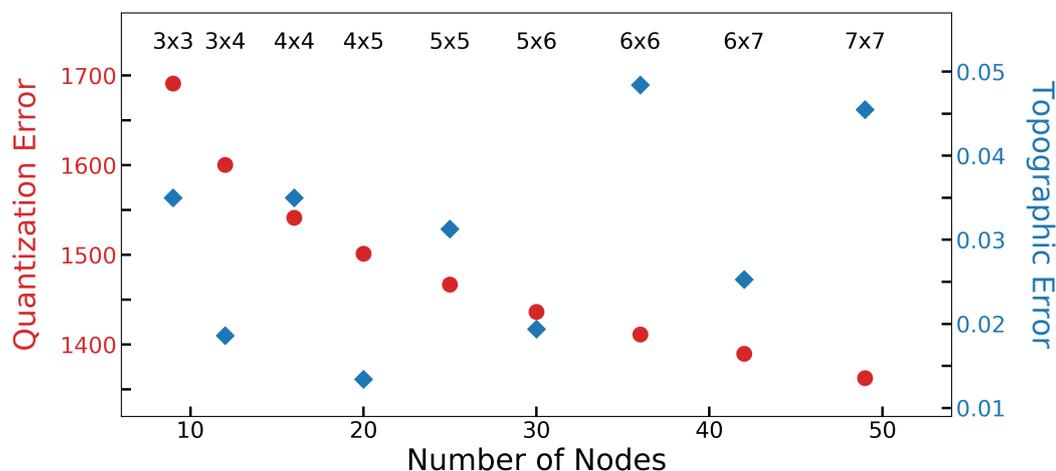
Figure 1. The quantization (red circles; left axis) and topographic (blue diamonds; right axis) errors for each SOM configuration tested in this study. SOM node topologies (rows × columns) range from 3 × 3 to 7 × 7; we choose to use a 4 × 5 node map.

Similar to previous work (e.g., Cavazos, 2000; Tymvios et al., 2010; Mechem et al., 2018), we choose an isobaric height field as our input data. Specifically, we use the 850-hPa spatial anomaly height field because we expect this variable to most accurately represent the location and strength of the NPH[4]. For the 2-dimensional input array, we use the 0000 UTC NCEP

90   North American Regional Reanalysis (NARR) 32-km product for each day during the months June through September from 2004 through 2014. Spatial anomalies are calculated for each day by subtracting the domain-averaged 850-hPa height from the 850-hPa height at each grid point. Each row of the input array represents one day from our data set, while each column represents a grid box from our NARR domain. The dimensions (rows × columns) of our input array are 1,342 × 6,952.

2.2   Data sets

95   In this study, we consider afternoon satellite measurements from *Aqua* MODIS because we use 0000 UTC NARR grids to generate the SOM. The satellite images, which are typically retrieved between 2030 and 2330 UTC, are paired with the NARR grid for the next day. For instance, we link the MODIS retrieval from 2200 UTC on 5 July 2010 to the NARR grid from 0000 UTC on 6 July, 2010. Even in the instance where the time difference between a MODIS image and NARR grid is a maximum (approximately 3.5 hours), we expect the influence of time mismatch to be minimal because we focus on the synoptic scale

---

[4]We also use the sea-level pressure (SLP) field as our input data; however, the result is an inaccurate representation of the different NPH patterns because there are many regions over the western United States where the SLP is extrapolated using a standard atmosphere assumption due to elevated terrain.

100 over relatively short time periods. Moreover, any two consecutive images (∼5 minutes apart) are stitched together and counted
as one sample. The MODIS files provide cloud information at 1-km horizontal spacing. We then interpolate these data to a
uniform $1/10° \times 1/10°$ (∼10 × 10 km) grid to be closer to the native horizontal grid spacing (∼32 km) of the NARR output
without losing too much detail.

2.3 MODIS processing

105 For the MODIS retrievals, values of r_e and τ are calculated utilizing a bispectral solar reflectance method (Nakajima and King,
1990), whereby extinction information is gleaned at 0.75- and 3.7-μm. We choose to interrogate retrievals from the 3.7-μm
channel because these data best represent the actual value of r_e at cloud top (Platnick, 2000; Rausch et al., 2017). Cloud liquid
water path (LWP) may then be inferred from the r_e and τ retrievals by the equation $LWP = C\rho_l r_e \tau$, where C is a function of
the assumed vertical distribution of cloud liquid water, and ρ_l is the density of liquid water (e.g., Miller et al., 2016). For the
110 calculation of LWP, we assume that the cloud vertical profile is approximately adiabatic ($C = 5/9$; e.g., Wood and Hartmann,
2006) and that N is approximately constant with height. Values of N may be estimated from observations of τ and LWP
after assuming an adiabatic cloud model (Bennartz, 2007). Moreover, we estimate fractional cloud albedo (α_c) using MODIS
retrievals of τ and following Lacis and Hansen (1974): $\alpha_c \approx \frac{0.13\tau}{1+0.13\tau}$, where τ is optical thickness. The top of the atmosphere
(TOA) CRF$_{SW}$ may then be calculated as CRF$_{SW} = \left(\frac{S_o}{4}\right)(\alpha_o - \alpha_c)$, where S_o is the solar constant (1370 W m^{-2}), and α_o is
115 the ocean albedo [0.10 (10%)]. The MODIS techniques are expounded in Juliano et al. (2018).

3 Results

3.1 Synoptic meteorological conditions

We now use the SOM output to investigate the various NARR 850-hPa meteorological patterns that are present during boreal
summer over the northeast Pacific Ocean from 2004 to 2014 (Fig. 2). There are several key features to discuss. The leftmost
120 part of the map (nodes 1, 6, 11, and 16) represents regimes where the NPH is relatively suppressed and a land-falling low-
pressure system is dominant. In general, strong, onshore flow is noticeable, and the flow diverges near the coastline. Relatively
cool temperatures related to the cyclonic circulation are present across the domain over the ocean and close to the shoreline
over land. Combined, these patterns account for approximately 22.8% of days in the data set. Moving from left to right across
the map, there is a smooth transition between regimes, and the presence of the NPH becomes more noticeable. The rightmost
125 portion of the map (nodes 5, 10, 15, and 20) represents synoptic-scale patterns where the NPH is dominant, and the nearshore
850-hPa flow is relatively weak or even directed offshore. Interestingly, there is a cyclonic circulation centered around 36° N,
127° W in node 5. For all of these nodes, and especially node 5, relatively high temperatures are observed along the coastline.
Approximately 26.0% of all days in the data set fall under these four nodes with a dominant NPH. Overall, the SOM appears
to capture well the spectrum of large-scale meteorological conditions that are present during the warm season months over the
130 northeast Pacific Ocean.
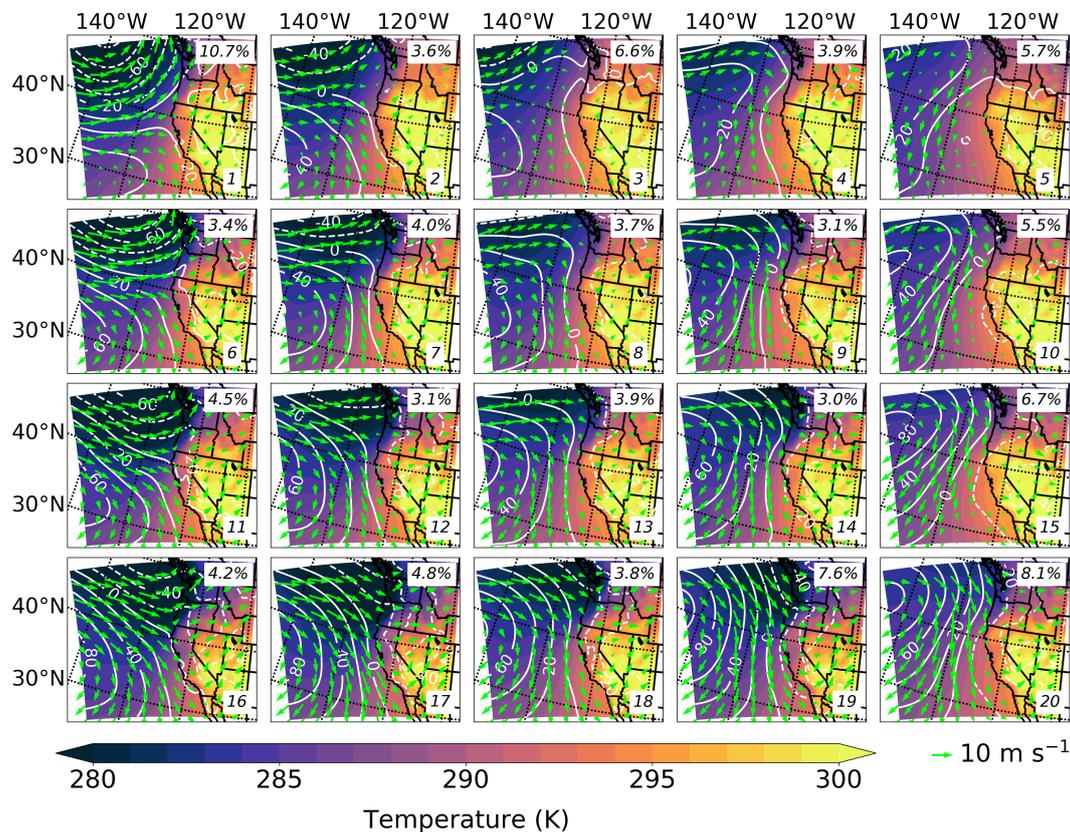
Figure 2. The synoptic-scale 850-hPa height (white lines contoured every 20 m), temperature (contoured every 1 K according to colorbar), and wind vector (lime green arrows with 10-m s^{-1} reference vector) fields as represented by the 4 × 5 SOM node topology. Each panel represents the mean conditions for each node (number displayed in the bottom right corner), while the percentage frequency of occurrence is displayed in the top right corner.

Large-scale regimes associated with both offshore continental flow driven by the NPH (e.g., node 5) and onshore continental flow driven by a land-falling cyclone (e.g., node 16) at 850 hPa often cause the near-surface alongshore flow to become southerly along the California coastline, as depicted by the observations from buoys 46013 and 46022 (Fig. 3). Offshore flow generates a weakening or reversal in the alongshore pressure gradient that drives southerly flow, while onshore flow is blocked

135   by the coastal terrain thereby forcing the flow to diverge in the meridional direction. The location and intensity of the NPH are main factors in dictating the northward extent and strength of the southerly flow for the 850-hPa offshore flow events. Similarly, the location and intensity of a land-falling cyclone control the location of alongshore flow bifurcation.

  Measurements from buoy 46013 (Bodega Bay), which is located just northwest of Point Reyes, California, suggest that southerly flow is present for a substantial number of hours (∼38.5%, ∼39.2%, ∼34.3%, and ∼29.3%) that fall within nodes

140   5, 10, 15, and 20, respectively. Meanwhile, buoy observations just northwest of Cape Mendocino (buoy 46022, Eel River) show strong influence from the land-falling cyclone (onshore flow) patterns; ∼28.3%, ∼22.7%, ∼23.8%, and ∼48.2% of the hours for nodes 1, 6, and 11, and 16, respectively, are characterized by southerly flow. The dependence of these regional flow

conditions on the synoptic-scale forcing regime is important for various meteorological applications such as ocean upwelling and offshore wind energy forecasting.

145   Table 1 lists the total and monthly frequencies of occurrence for each node. In general, the majority of days that are represented by the land-falling cyclone regime (nodes 1, 6, 11, and 16) are in early summer (June) and early fall (September). This is not surprising because these systems are more common during transition seasons than during summer (e.g., Reitan, 1974). Additionally, we find that the dominant NPH regime (nodes 5, 10, 15, and 20) occurs most often in July, August, or September. We also note that node 5, which represents a regime characterized by a weak regional height gradient, shows a strong increase

150   in frequency of occurrence over time (frequencies of 15.8%, 14.5%, 28.9%, and 40.8% in June, July, August, and September, respectively).

Percentage of Hours with Southerly Flow

Buoy 46013

Buoy 46022

Figure 3. The percentage of southerly flow hours recorded at each buoy site along the California coastline for each regime in the 4 × 5 SOM node topology: (left panel) buoy 46013 (Bodega Bay; 38.238° N 123.307° W) and (right panel) buoy 46022 (Eel River; 40.712° N 124.529° W).

Due to the nature of the SOM, adjacent synoptic-scale patterns are similar to one another, and there is a gradual transition between different regimes as one moves across the SOM. The SOM patterns farther left on the map are associated with generally strong westerly flow offshore and divergent flow near the coastline due to a dominant cold-core land-falling cyclone.

155   Conversely, those patterns toward the right feature northerly, and even northeasterly, flow offshore due to a dominate warm-core NPH. Moreover, several of the regimes (nodes 3, 4, and 5) feature a noticeably weak 850-hPa height gradient; on average, the winds over the ocean at this level are <5 m s^{-1}. In general, the top-right SOM nodes feature the most notable offshore continental flow (and associated weak nearshore winds at southern latitudes in the domain) because the 850-hPa height contours are oriented northeast-southwest and the wind vectors have pronounced south and west components. Therefore, one might

160   expect to see relatively high N and small r_e values dominate in these regimes because they appear to be influenced strongly by continental air masses.

Table 1. Summary statistics for SOM node frequency. Total and monthly frequencies over the 11-year period are shown.

Node	Freq., counts	Freq., %	Monthly Freq., %			
			Jun	Jul	Aug	Sep
1	143	10.7	27.3	23.1	15.4	34.2
2	48	3.6	20.8	27.1	33.3	18.8
3	89	6.6	10.1	28.1	27.0	34.8
4	52	3.9	9.6	32.7	34.6	23.1
5	76	5.7	15.8	14.5	28.9	40.8
6	46	3.4	34.8	17.4	23.9	23.9
7	54	4.0	29.6	31.5	25.9	13.0
8	50	3.7	16.0	22.0	36.0	26.0
9	42	3.1	11.9	31.0	38.1	19.0
10	74	5.5	10.8	27.0	33.8	28.4
11	61	4.5	40.9	19.7	19.7	19.7
12	41	3.1	34.2	26.8	24.4	14.6
13	53	3.9	20.7	30.2	34.0	15.1
14	40	3.0	15.0	20.0	45.0	20.0
15	90	6.7	6.7	44.4	23.3	25.6
16	56	4.2	51.8	7.1	14.3	26.8
17	65	4.8	53.8	15.4	10.8	20.0
18	51	3.8	33.3	29.4	27.5	9.8
19	102	7.6	41.2	13.7	21.6	23.5
20	109	8.1	15.6	39.5	22.9	22.0

3.2 MODIS cloud retrievals

Figures 4 and 5 show the mean N and r_e values from MODIS that are associated with each node. In Fig. 4, the red (yellow) end of the colorbar corresponds to relatively low (high) N, and in Fig. 5, the red (yellow) end of the colorbar corresponds to relatively small (large) r_e. Therefore, yellow regions in Fig. 4 and red regions in Fig. 5 indicate a potential influence of continental and/or shipping aerosol sources on marine clouds.

Although the MODIS retrievals are not used directly to generate the SOM, and instead are simply associated with the corresponding days in each node, there is an apparent connection between the various synoptic-scale patterns in the 850-hPa height fields (which are used to generate the SOM) depicted in Fig. 2 and the MODIS cloud properties shown in Figs. 4 and 5. Generally, there are more regions of high N and smaller r_e as one moves from left to right across the SOM; that is, nodes to the left (right) on the SOM represent days where marine clouds are, on average, composed of less numerous and larger (more
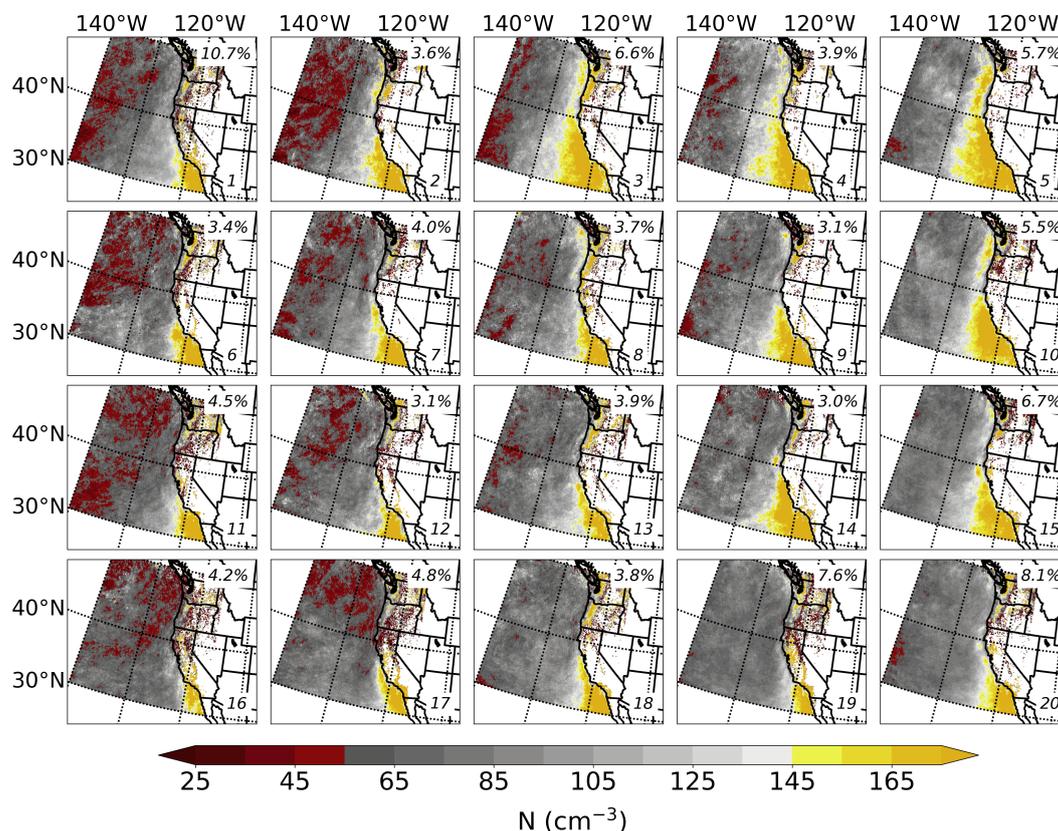
Figure 4. The MODIS estimation of N (contoured every 10 cm^{-3} according to the colorbar) for each regime in the 4 × 5 SOM node topology. Each panel represents the mean conditions for each node (number displayed in the bottom right corner), while the percentage frequency of occurrence is displayed in the top right corner.

numerous and smaller) droplets. Through a visual inspection, node 5 appears to be most representative of cases where marine stratiform clouds have more numerous and smaller droplets. As shown in the analysis in Fig. 2, node 5 is characterized by distinct offshore continental flow at 850 hPa, in addition to very weak flow (and even southerly flow) near the shoreline. These

175 results highlight the utility of using reanalysis to define modes of large-scale pressure patterns and subsequently incorporate other data sets — satellite observations in the case here — to understand interactions across spatial scales that could not otherwise be gleaned from the original reanalysis products themselves with sufficient accuracy.

Evident in all of the SOM nodes is a region of high N and small r_e south of the pronounced coastal bend near Point Conception, California (approximately 34.4° N, 120.5° W). This nearshore oceanic region is likely polluted due to its proximity

180 to population centers (namely Los Angeles, San Diego, and Tijuana) and wildfire activity (e.g., Duong et al., 2011; Metcalf et al., 2012; Zauscher et al., 2013). Also, this area serves as a major port for international trade and it hosts numerous refineries (e.g., Ault et al., 2009; Ryerson et al., 2013). In this region, transport of aerosol is governed typically by the synoptic-scale conditions and mesoscale land-sea breeze processes (e.g., Agel et al., 2011; Naifang et al., 2013); however, previous work
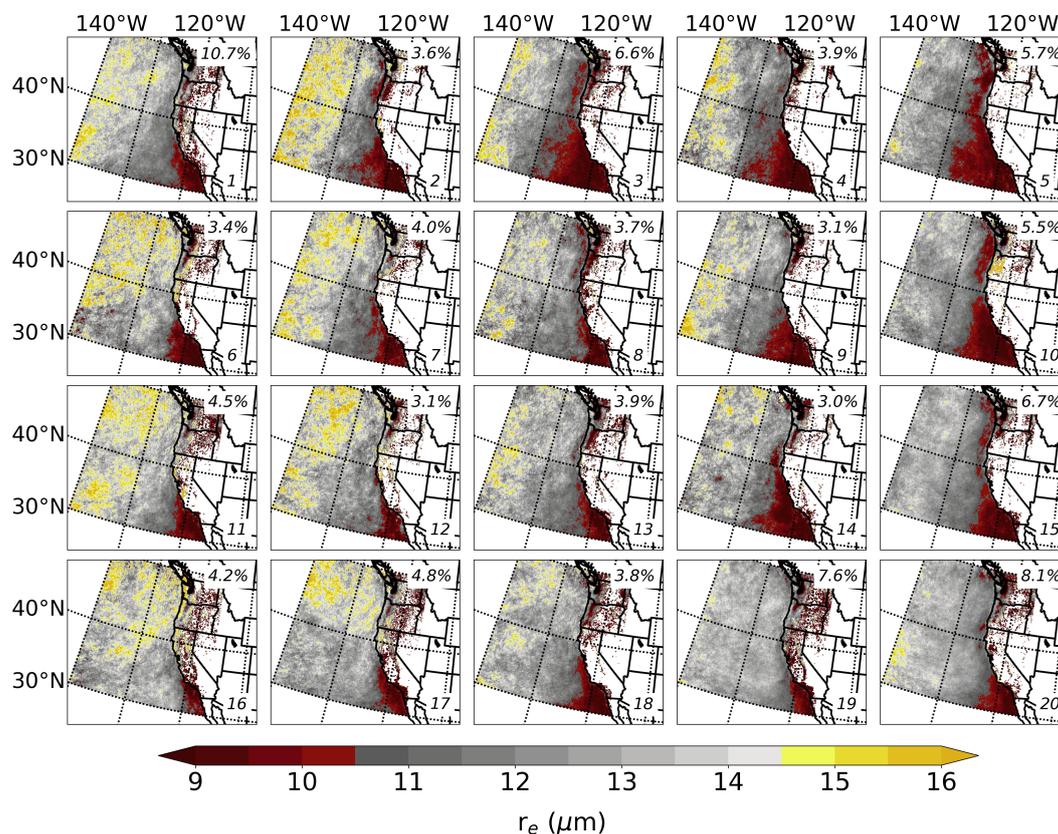
Figure 5. As in Fig. 4 except for the MODIS retrieval of r$_e$ (contoured every 0.5 μm according to the colorbar).

suggests that the pervasive Catalina Eddy — a phenomenon linked to the generation of CTDs (e.g., Skamarock et al., 2002) —

185 may transport pollution offshore and toward the north (Wakimoto, 1987). We hypothesize that the MODIS retrievals presented here show clearly that the 1$^{\text{st}}$ aerosol indirect effect (Twomey, 1977) is more pronounced in the nodes farther to the right on the map due to this complex combination of atmospheric processes that impacts marine clouds through aerosol-cloud interactions. Specifically, we hypothesize that the transport of continental aerosol (e.g., nitrates, sulfates, biogenic organics, and dust) into the marine environment, in addition to the interaction of ship track aerosol (e.g., sulfates) and marine aerosol (e.g., sea salt),

190 increases the number of cloud condensation nuclei (CCN) and therefore cloud droplets. These effects are most notable within several hundred kilometers of the western United States and Baja California coastlines; however, remote oceanic locations also appear to be influenced strongly by the NPH regime. Additionally, the areas likely affected by pollution sources extend along nearly the entire coastline in the nodes to the right on the SOM, while the nodes to the left on the SOM show a much more confined region of polluted clouds due to strong, onshore flow. In general, the nodes display varying extensions according to

195 the synoptic-scale regime.

Frequency distributions reveal that between the various SOM nodes, N, r_e, and CRF$_{SW}$ exhibit a broad range that is dependent on the prevailing synoptic-scale pattern (Fig. 6; cf. Table 2 for median values). The distributions confirm that node
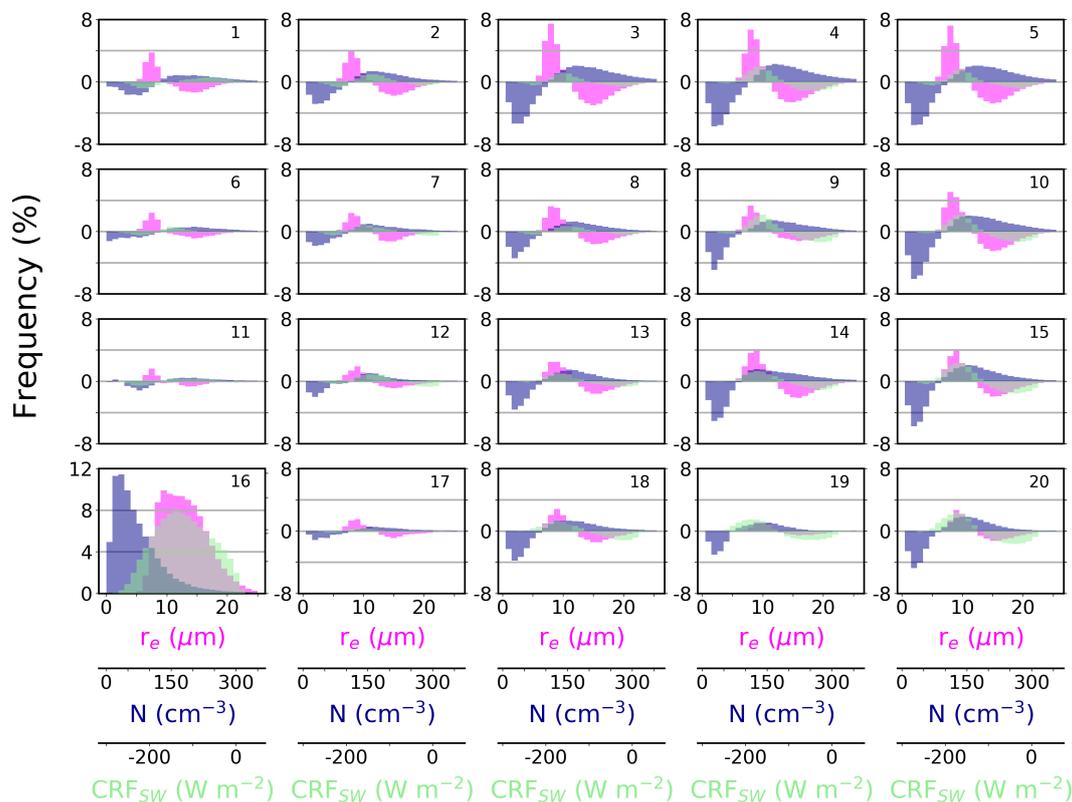
Figure 6. Frequency distributions of r_e (units in μm; colored magenta), N (units in cm^{-3}; colored dark blue), and CRF$_{SW}$ (units in W m^{-2}; colored light green) for each regime in the 4×5 SOM node topology. The node number is shown in the top right corner of each subplot. The actual distributions are shown for node 16, while the difference relative to node 16 (node x minus node 16 where x is a given node) is shown for all other nodes. Therefore, smaller (larger) values indicate a deficit (surplus) relative to node 16. The distributions of r_e and N are generated from the plan views in Figs. 4 and 5. The distribution of CRF$_{SW}$ is calculated from the plan view of τ (cf. Section 2.3), which is generated similarly to r_e and N but is not shown here. Also, median values of each distribution are documented in Table 2.

5, in addition to nodes 3, 4, and 10, represent the scenarios where MBL clouds are characterized by relatively high N, small r_e, and strong CRF$_{SW}$ compared to the other meteorological regimes. The median values of N, r_e, and CRF$_{SW}$ are 93.2 cm^{-3},

200   10.4 μm, and -129.7 W m^{-2} for node 3, 93.3 cm^{-3}, 10.6 μm, and -138.1 W m^{-2} for node 4, 92.8 cm^{-3}, 10.5 μm, and -132.3 W m^{-2} for node 5, and 90.6 cm^{-3}, 10.8 μm, and -139.1 W m^{-2} for node 10. For most of the other nodes, the frequency distributions of N and r_e are shifted toward the left and right, respectively, indicative of fewer and larger cloud droplets. In the patterns that are much different than nodes 3, 4, 5, and 10 — for example, node 16, in addition to nodes 6, 11, and 17 — the distributions are shifted appreciably such that the median values of N, r_e, and CRF$_{SW}$ are 62.5 cm^{-3}, 12.0 μm, and

205   -123.9 W m^{-2} for node 6, 57.8 cm^{-3}, 12.3 μm, and -121.7 W m^{-2} for node 11, 57.2 cm^{-3}, 12.5 μm, and -124.1 W m^{-2} for node 16, and 61.9 cm^{-3}, 12.0 μm, and -125.4 W m^{-2} for node 17. We note that in some regimes (e.g., node 19), CRF$_{SW}$ does not correlate as well with N and r_e, which suggests that other factors not accounted for here such as aerosol composition, turbulence, and sea surface temperature may play an important role.

Table 2. Summary statistics for SOM node meteorological and cloud properties. We tabulate median values, in addition to the difference relative to node 16 (node x minus node 16, where x is a given node) in parenthesis, of the frequency distributions of r_e, N, and CRF_{SW} (cf. Fig. 6) as well as those of meteorological variables LCF, LTS, and q_v (not shown).

Node	LCF, %	LTS	q_v, g kg^{-1}	r_e, μm	N, cm^{-3}	CRF_{SW}, W m^{-2}
1	63 (+2)	13.4 (-0.4)	4.6 (+0.2)	11.7 (-0.8)	63.6 (+6.4)	-117.5 (+6.6)
2	61 (0)	13.4 (-0.4)	4.9 (+0.3)	11.3 (-1.2)	73.0 (+15.8)	-126.0 (-1.9)
3	63 (+2)	13.3 (-0.5)	4.8 (+0.4)	10.4 (-2.1)	93.2 (+36.0)	-129.7 (-5.6)
4	62 (+1)	13.3 (-0.5)	5.0 (+0.6)	10.6 (-1.9)	93.3 (+36.1)	-138.1 (-14.0)
5	64 (+3)	13.3 (-0.5)	4.7 (+0.3)	10.5 (-2.0)	92.8 (+35.6)	-132.3 (-8.2)
6	61 (0)	13.7 (-0.1)	4.7 (+0.3)	12.0 (-0.5)	62.5 (+5.3)	-123.9 (+0.2)
7	61 (0)	13.7 (-0.1)	4.9 (+0.5)	11.8 (-0.7)	66.6 (+9.4)	-127.0 (-2.9)
8	63 (+2)	13.4 (-0.4)	4.9 (+0.5)	11.4 (-1.1)	72.4 (+15.2)	-126.4 (-2.3)
9	63 (+2)	13.5 (-0.3)	5.0 (+0.6)	11.5 (-1.0)	77.3 (+20.1)	-138.4 (-14.3)
10	64 (+3)	13.3 (-0.5)	4.9 (+0.5)	10.8 (-1.7)	90.6 (+33.4)	-139.1 (-15.0)
11	60 (-1)	13.7 (-0.1)	4.7 (+0.3)	12.3 (-0.2)	57.8 (+0.6)	-121.7 (+2.4)
12	62 (+1)	13.7 (-0.1)	4.8 (+0.4)	12.0 (-0.5)	65.2 (+8.0)	-127.9 (-3.8)
13	63 (+2)	13.6 (-0.2)	4.9 (+0.5)	11.5 (-1.0)	74.0 (+16.8)	-132.3 (-8.2)
14	63 (+2)	13.4 (-0.4)	5.0 (+0.6)	11.1 (-1.4)	81.7 (+24.5)	-135.4 (-11.3)
15	64 (+3)	13.5 (-0.3)	5.0 (+0.6)	11.2 (-1.3)	85.0 (+27.8)	-141.6 (-17.5)
16	61 (—)	13.8 (—)	4.4 (—)	12.5 (—)	57.2 (—)	-124.1 (—)
17	60 (-1)	13.9 (+0.1)	4.5 (+0.1)	12.0 (-0.5)	61.9 (+4.7)	-125.4 (-1.3)
18	61 (0)	13.7 (-0.1)	4.9 (+0.5)	11.6 (-0.9)	75.0 (+17.8)	-136.0 (-11.9)
19	61 (0)	13.8 (0)	4.6 (+0.2)	12.1 (-0.4)	68.5 (+11.3)	-139.3 (-15.2)
20	63 (+2)	13.7 (-0.1)	4.9 (+0.5)	11.7 (-0.8)	77.8 (+20.6)	-144.0 (-19.9)

To explore the potential impact of the regional meteorology associated with each of the synoptic-scale regimes — compared
210 to simply the abundance of aerosol — on the observed cloud properties, we also examine low cloud fraction (LCF), lower tropospheric stability [LTS; Klein and Hartmann (1993)], and 850-700 hPa mean water vapor mixing ratio (q_v) from the NARR grids (Table 2). In general, LCF increases, LTS decreases, and q_v increases from left to right across the SOM; however, these relationships do not appear to explain all of the variability in the observed cloud properties among the various nodes. Therefore, we attribute the variability in the satellite-retrieved cloud microphysical and radiative properties to aerosol forcing
215 (first order effect) as opposed to meteorological factors (second order effect).

Overall, our SOM results elucidate the apparent coupling between NPH dynamics and mesoscale MBL cloud properties. That is, generally weak flow and/or an enhancement in offshore continental flow at 850 hPa (e.g., node 5) likely augments aerosol transport into the marine layer, thereby increasing both the number of CCN and the brightness (reflection) of MBL clouds. Moreover, we hypothesize that a weaker regional pressure gradient allows for the transport of aerosol by the coastal jet

220 due to the dominance of localized land-sea breeze circulations, which may advect continental aerosol offshore (e.g., Lawrence and Lelieveld, 2010; Loughner et al., 2014; Mazzuca et al., 2017). In contrast, a suppression in offshore continental flow (e.g., node 16) likely inhibits continental aerosol transport into the marine layer, thereby decreasing both the number of CCN and the brightness (reflection) of MBL clouds.

4   Discussion and conclusions

225 Through the use of a SOM, we show that the location and intensity of the NPH, as well as the presence of land-falling low-pressure systems, play a significant role in modifying MBL cloud microphysical properties offshore of the western United States during boreal summer. The 850-hPa height field is chosen as the meteorological input variable for the SOM algorithm because it represents well the large-scale circulation over the northeast Pacific Ocean. Results from the SOM reveal several distinct synoptic patterns present during the Northern Hemisphere warm season as well as their frequencies of occurrence;

230 however, most notable is the high frequency of pronounced offshore continental flow and generally weak flow. Incorporating MODIS observations into the analysis yields a connection between the synoptic-scale dynamics and mesoscale cloud micro-physics. Specifically, more instances of polluted clouds (as indicated by high N and small r_e) that are highly reflective (strong CRF_{SW}) are found during node patterns with offshore flow or weak flow at 850 hPa (e.g., nodes 3, 4, 5, and 10).

The findings reported here may be of significant interest to atmospheric science communities utilizing climate models

235 (CMs) because the synoptic-scale flow–cloud microphysics relationship from the SOM may be used to test CMs and probe uncertainties in their simulation of aerosol effects. For instance, the SOM results may be used to better understand if CMs are capable of reproducing similar patterns between large-scale circulation and cloud microphysics/radiative forcing. One could then quantify the impact of using the radiative forcing from the observed SOM relationship with the modeled 850 hPa height field rather than the model-predicted radiative forcing over the semi-permanent marine stratiform regions. Also, this analysis

240 could be extended to evaluate in a statistical sense the ability of long-term simulations to replicate each large-scale regime.

Moreover, most CMs have difficulty with accurately representing MBL clouds — which are susceptible to aerosol effects — because they often use a horizontal grid spacing that is too large ($\gtrsim$10 km). However, reproducing large-scale meteorological fields, such as pressure or isobaric height, is typically easier for CMs. Here, we demonstrate a "proof of concept" study of a novel method to link well-resolved synoptic-scale features to cloud microphysics and shortwave radiative forcing. Because the

245 approach is relatively simple to implement, it may be applied to other problems in atmospheric science involving interactions between spatial scales.

While the results presented here are promising, a data set spanning a longer time period is required to develop a robust analysis that evaluates the ability of CMs to reproduce the observed synoptic-scale weather patterns and mesoscale cloud properties. In general, using machine learning techniques to connect large-scale circulation patterns to cloud microphysics, which

250 is challenging using solely observations from field campaigns or modeling case studies, is important for accurate predictions of future atmospheric climate. The results presented here may not be applicable to all marine stratiform cloud decks owing to potential differences in the frequency, strength, and location of the respective high pressure circulation, as well as differences

in, for example, coastal geometry and topography, continental land use, aerosol sources, and sea surface temperature. Future work will explore the application of the methodology outlined herein to the other dominant MBL cloud regions of the world using global reanalysis products and model output.

255

Code and data availability. NARR reanalysis are available from the National Oceanic and Atmospheric Administration (NOAA) National Centers for Environmental Information (NCEI) website (https://www.ncdc.noaa.gov/), *Aqua* MODIS Level 2 satellite retrievals are available from the National Aeronautics and Space Administration (NASA) Earthdata website (https://search.earthdata.nasa.gov/), and buoy observations are available from the NOAA National Data Buoy Center (NDBC) website (https://www.ndbc.noaa.gov/). The Matlab SOM Toolbox code is available for download courtesy of the Helsinki University of Technology (http://www.cis.hut.fi/projects/somtoolbox/). Additional codes are available upon request.

260

Author contributions. TWJ designed the study, developed the code, performed the analysis, and wrote the manuscript. ZJL made substantial contributions to the analysis and revised the manuscript.

Competing interests. The authors declare that they have no conflict of interest.

265   *Acknowledgements.* The authors are grateful for support in part from the State of Wyoming, the Carlton R. Barkhurst Fellowship, and NCAR through the National Science Foundation (TWJ) in addition to the Department of Energy through grant DE-SC0016354 (ZJL). We would also like to acknowledge high-performance computing support from Cheyenne (doi:10.5065/D6RX99HX) provided by NCAR's Computational and Information Systems Laboratory and sponsored by the National Science Foundation. The authors thank Hugh Morrison, Andrew Gettelman, Kevin Reed, and Stefan Rahimi for providing valuable comments on an earlier version of the manuscript.

References

Agel, L., Lopez, V., Barlow, M., and Colby, F.: Regional and large-scale influences on summer ozone levels in Southern California, *J. Appl. Meteor. Clim.*, 50, 800–805, 2011.

Ault, A. P., Moore, M. J., Furutani, H., and Prather, K. A.: Impact of emissions from the Los Angeles Port region on San Diego air quality during regional transport events, *Environ. Sci. Technol.*, 43, 3500–3506, 2009.

Bachiochi, D. R. and Krishnamurti, T. N.: Enhanced low-level stratus in the FSU coupled ocean—atmosphere model, *Mon. Wea. Rev.*, 128, 3083–3103, 2000.

Bennartz, R.: Global assessment of marine boundary layer cloud droplet number concentration from satellite, *J. Geophys. Res.*, 112, D02201, 2007.

Bony, S. and Dufresne, J.-L.: Marine boundary layer clouds at the heart of tropical cloud feedback uncertainties in climate models, *Geophys. Res. Let.*, 32, l20806, 2005.

Cassano, E. N., Lynch, A. H., Cassano, J. J., and Koslow, M. R.: Classification of synoptic patterns in the western Arctic associated with extreme events at Barrow, Alaska, USA, *Clim. Res.*, 30, 83–97, 2006.

Cassano, E. N., Glisan, J. M., Cassano, J. J., Gutowski, W. J. Jr., and Seefeldt, M. W.: Self-organizing map analysis of widespread temperature extremes in Alaska and Canada, *Clim. Res.*, 62, 199–218, 2015.

Cavazos, T.: Using self-organizing maps to investigate extreme climate events: An application to wintertime precipitation in the Balkans, *J. Climate*, 13, 1718–1732, 2000.

Crosbie, E., Wang, Z., Sorooshian, A., Chuang, P. Y., Craven, J. S., Coggon, M. M., Brunke, M., Zeng, X., Jonsson, H., Woods, R. K., Flagan, R. C., and Seinfeld, J.: Stratocumulus cloud clearings and notable thermodynamic and aerosol contrasts across the clear—cloudy interface, *J. Atmos. Sci.*, 73, 1083–1099, 2016.

Delecluse, P., Davey, M. K., Kitamura, Y., Philander, S. G. H., Suarez, M., and Bengtsson, L.: Coupled general circulation modeling of the tropical Pacific, *J. Geophys. Res.*, 103, 14 357–14 373, 1998.

Duong, H. T., Sorooshian, A., Craven, J. S., Hersey, S. P., Metcalf, A. R., Zhang, X., Weber, R. J., Jonsson, H., Flagan, R. C., and Seinfeld, J. H.: Water-soluble organic aerosol in the Los Angeles Basin and outflow regions: Airborne and ground measurements during the 2010 CalNex field campaign, *J. Geophys. Res.*, 116, D00V04, 2011.

Fewings, M. R., Washburn, L., Dorman, C. E., Gotschalk, C., and Lombardo, K.: Synoptic forcing of wind relaxations at Pt. Conception, California, *J. Geophys. Res. Oceans*, 121, 5711–5730, 2016.

Ford, T. W., Quiring, S. M., Frauenfeld, O. W., and Rapp, A. D.: Synoptic conditions related to soil moisture-atmosphere interactions and unorganized convection in Oklahoma, *J. Geophys. Res. Atmos.*, 120, 11,519–11,535, 2015.

Hartmann, D. L., Ockert-Bell, M. E., and Michelsen, M. L.: The effect of cloud type on Earth's energy balance: Global analysis, *J. Climate*, 5, 1281–1304, 1992.

Jensen, A. A., Thompson, A. M., and Schmidlin, F. J.: Classification of Ascension Island and Natal ozonesondes using self-organizing maps, *J. Geophys. Res.*, 11, D04302, 2012.

Juliano, T. W., Lebo, Z. J., Thompson, G., and Rahn, D. A.: A new perspective on coastally trapped disturbances using data from the satellite era, *Bull. Amer. Meteor. Soc.*, 2018.

Klein, S. A. and Hartmann, D. L.: The seasonal cycle of low stratiform clouds, *J. Climate*, 6, 1587–1606, 1993.

Kloesel, K. A.: Marine stratocumulus cloud clearing episodes observed during FIRE, *Mon. Wea. Rev.*, 120, 565–578, 1992.

Kohonen, T.: The self-organizing map, *Proc. IEEE*, 78, 1464–1480, 1990.

Koraĉin, D. and Dorman, C. E.: Marine Fog: Challenges and Advancements in Observations, Modeling, and Forecasting, Springer, New
York, 2017.

310 Lacis, A. A. and Hansen, J. E.: A parameterization for the absorption of solar radiation in the Earth's atmosphere, *J. Atmos. Sci.*, 31, 118–133,
1974.

Lawrence, M. G. and Lelieveld, J.: Atmospheric pollutant outflow from southern Asia: a review, *Atmos. Chem. Phys.*, 10, 11 017–11 096,
2010.

Lin, J., Qian, T., and Shinoda, T.: Stratocumulus clouds in southeastern Pacific simulated by eight CMIP5-CFMIP global climate models, *J.*
315 *Climate*, 27, 3000–3022, 2014.

Loughner, C., Tzortziou, M., Follette-Cook, M., Pickering, K. E., Goldberg, D., Satam, C., Weinheimer, A., Crawford, J. H., Knapp, D. J.,
Montzka, D. D., Diskin, G. S., and Dickerson, R. R.: Impact of bay-breeze circulations on surface air quality and boundary layer export,
J. Appl. Meteor. Clim., 53, 1697–1713, 2014.

Mass, C. F. and Bond, N. A.: Coastally trapped wind reversals along the United States West Coast during the warm season. Part II: Synoptic
320 evolution, *Mon. Wea. Rev.*, 124, 446–461, 1996.

Mazzuca, G. M., Pickering, K. E., Clark, R. D., Loughner, C. P., Fried, A., Zweers, D. C. S., Weinheimer, A. J., and Dickerson, R. R.: Use of
tethersonde and aircraft profiles to study the impact of mesoscale and microscale meteorology on air quality, *Atmos. Environ.*, 149, 55–69,
2017.

Mechem, D. B., Wittman, C. S., Miller, M. A., Yuter, S. E., and de Szoeke, S. P.: Joint Synoptic and Cloud Variability over the Northeast
325 Atlantic near the Azores, *J. Appl. Meteor. Clim.*, 57, 1273–1290, 2018.

Metcalf, A. R., Craven, J. S., Ensberg, J. J., Brioude, J., Angevine, W., Sorooshian, A., Duong, H. T., Jonsson, H. H., Flagan, R. C., and
Seinfeld, J. H.: Black carbon aerosol over the Los Angeles Basin during CalNex, *J. Geophys. Res.*, 117, D00V13, 2012.

Miller, D. J., Zhang, Z., Ackerman, A. S., Platnick, S., and Baum, B. A.: The impact of cloud vertical profile on liquid water path retrieval
based on the bispectral method: A theoretical study based on large-eddy simulations of shallow marine boundary layer clouds, *J. Geophys.*
330 *Res. Atmos.*, 121, 4122–4141, 2016.

Naifang, B., Li, G., Zavala, M., Barrera, H., Torres, R., Grutter, M., Gutiérrez, W., García, M., Ruiz-Suarez, L. G., Ortinez, A., Guitierrez, Y.,
Alvarado, C., Flores, I., and Molina, L. T.: Meteorological overview and plume transport patterns during Cal-Mex 2010, *Atmos. Environ.*,
70, 477–489, 2013.

Nakajima, T. and King, M. D.: Determination of the optical thickness and effective particle radius of clouds from reflected solar radiation
335 measurements. Part I: Theory, *J. Atmos. Sci.*, 47, 1878–1893, 1990.

Nowotarski, C. and Jensen, A. A.: Classifying proximity soundings with self-organizing maps toward improving supercell and tornado
forecasting, *Weather Forecasting*, 28, 783–801, 2013.

Nuss, W. A., Bane, J. M., Thompson, W. T., Holt, T., Dorman, C. E., Ralph, F. M., Rotunno, R., Klemp, J. B., Skamarock, W. C., Samelson,
R. M., Rogerson, A. M., Reason, C., and Jackson, P.: Coastally trapped wind reversals: Progress toward understanding, *Bull. Amer. Meteor.*
340 *Soc.*, 81, 719–743, 2000.

Palmer, T. N. and Anderson, D. L. T.: The prospect for seasonal forecasting — A review paper, *Quart. J. Roy. Meteor. Soc.*, 120, 755–793,
1994.

Parish, T. R.: Forcing of the summertime low-level jet along the California coast, *J. Appl. Meteor.*, 39, 2421–2433, 2000.

Parish, T. R., Rahn, D. A., and Leon, D.: Aircraft observations of a coastally trapped Wind reversal off the California coast, *Mon. Wea. Rev.*,
345   136, 644–663, 2008.

Platnick, S.: Vertical photon transport in cloud remote sensing problems, *J. Geophys. Res. Atmos.*, 105, 22 919–22 935, 2000.

Rausch, J., Meyer, K., Bennartz, R., and Platnick, S.: Differences in liquid cloud droplet effective radius and number concentration estimates
       between MODIS collections 5.1 and 6 over global oceans, *Atmos. Meas. Tech.*, 10, 2105–2116, 2017.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., and Prabhat: Deep learning and process understanding
350   for data-driven Earth system science, *Nature*, 566, 195–204, 2019.

Reitan, C. H.: Frequencies of cyclones and cyclogenesis for North America, 1951–1970, *Mon. Wea. Rev.*, 102, 861–868, 1974.

Ryerson, T. B., Andrews, A. E., Angevine, W. M., Bates, T. S., Brock, C. A., Cairns, B., Cohen, R. C., Cooper, O. R., de Gouw, J. A.,
       Fehsenfeld, F. C., Ferrare, R. A., Fischer, M. L., Flagan, R. C., Goldstein, A. H., Hair, J. W., Hardesty, R. M., Hostetler, C. A., Jimenez,
       J. L., Langford, A. O., McCauley, E., McKeen, S. A., Molina, L. T., Nenes, A., Oltmans, S. J., Parrish, D. D., Pederson, J. R., Pierce,
355   R. B., Prather, K., Quinn, P. K., Seinfeld, J. H., Senff, C. J., Sorooshian, A., Stutz, J., Surratt, J. D., Trainer, M., Volkamer, R., Williams,
       E. J., and Wofsy, S. C.: The 2010 California Research at the Nexus of Air Quality and Climate Change (CalNex) field study, *J. Geophys.
       Res. Atmos.*, 118, 5830–5866, 2013.

Skamarock, W. C., Rotunno, R., and Klemp, J. B.: Catalina eddies and coastally trapped disturbances, *J. Atmos. Sci.*, 59, 2270–2278, 2002.

Skific, N., Francis, J. A., and Cassano, J. J.: Attribution of projected changes in atmospheric moisture transport in the Arctic: A self-organizing
360   map perspective, *J. Climate*, 22, 4135–4153, 2009.

Stauffer, R. M., Thompson, A. M., Oltmans, S. J., and Johnson, B. J.: Tropospheric ozonesonde profiles at long-term U.S. monitoring sites:
       2. Links between Trinidad Head, CA, profile clusters and inland surface ozone measurements, *J. Geophys. Res. Atmos.*, 122, 1261–1280,
       2017.

Stevens, B. and Feingold, G.: Untangling aerosol effects on clouds and precipitation in a buffered system, *Nature*, 461, 607–613, 2009.

365 Thompson, W. T., Burke, S. D., and Lewis, J.: Fog and low clouds in a coastally trapped disturbance, *J. Geophys. Res.*, 110, D18213, 2005.

Twomey, S.: The influence of pollution on the shortwave albedo of clouds, *J. Atmos. Sci.*, 34, 1149–1152, 1977.

Tymvios, F., Savvidou, K., and Michaelides, S. C.: Association of geopotential height patterns with heavy rainfall events in Cyprus, *Adv.
       Geosci.*, 23, 73–78, 2010.

Ultsch, A. and Herrmann, L.: The architecture of emergent self-organizing maps to reduce projection errors, in: ESANN, 2005.

370 Wakimoto, R. M.: The Catalina Eddy and its effect on pollution over Southern California, *Mon. Wea. Rev.*, 115, 837–855, 1987.

Wang, L., Wang, Y., Lauer, A., and Xie, S.: Simulation of seasonal variation of marine boundary layer clouds over the Eastern Pacific with a
       regional climate model, *J. Climate*, 24, 3190–3210, 2011.

Wang, Y., Xu, H., and Xie, S.: Regional model simulations of marine boundary layer clouds over the Southeast Pacific off South America.
       Part II: Sensitivity experiments, *Mon. Wea. Rev.*, 132, 2650–2668, 2004.

375 Webb, M. J., Senior, C. A., Sexton, D. M. H., Ingram, W. J., Williams, K. D., Ringer, M. A., McAvaney, B. J., Colman, R., Soden, B. J.,
       Gudgel, R., Knutson, T., Emori, S., Ogura, T., Tsushima, Y., Andronova, N., Li, B., Musat, I., Bony, S., and Taylor, K. E.: On the
       contribution of local feedback mechanisms to the range of climate sensitivity in two GCM ensembles, *Climate Dynamics*, 27, 17–38,
       2006.

Wood, R.: Stratocumulus clouds, *Mon. Wea. Rev.*, 140, 2373–2423, 2012.

380 Wood, R. and Hartmann, D. L.: Spatial variability of liquid water path in marine low cloud: The importance of mesoscale cellular convection,
       J. Climate, 19, 1748–1764, 2006.

Yi, Z. and Jian, L.: Shortwave cloud radiative forcing on major stratus cloud regions in AMIP-type simulations of CMIP3 and CMIP5 models, *Adv. Atmos. Sci.*, 30, 884–907, 2013.

Zauscher, M. D., Wang, Y., Moore, M. J. K., Gaston, C. J., and Prather, K. A.: Air quality impact and physicochemical aging of biomass
385   burning aerosols during the 2007 San Diego wildfires, *Environ. Sci. Technol.*, 47, 7633–7643, 2013.