We thank both reviewers for their valuable comments that have improved this version of the manuscript. Individual comments are addressed below, with the original comments/suggestions in black where possible and our point-by-point responses in blue italic font.

Before we address the reviewers' comments, we would like to make it clear that we discovered a MODIS processing error. When processing the Level 2 MODIS files, we accidentally shifted the days by one so that, for each year in the analysis, we processed from May 31st through September 29th instead of from June 1st through September 30th. Consequently, we have updated all figures, tables, and any text that refer to numerical values influenced by this mistake. Overall, the results of the study are not affected considerably and the main findings remain the same.

Additionally, we want to point out that we unintentionally miscalculated the meteorological variables (LCF, LTS, and q_v) in Table 2 because we originally included regions over land. This mistake has been fixed, and the correct values for LCF, LTS, and q_v are now shown in Table 2. The differences are not large, and so our discussion regarding these numbers has not changed.

In response to comments from both Reviewers 1 and 2 regarding the appropriateness of using 20 different meteorological regimes, we agree with the reviewers that this nomenclature is not realistic. Our intention is not to classify each node as a separate regime, and we apologize for this confusion. While 20 nodes may seem like a lot, it is customary in the literature (e.g, Cassano et al. 2006, 2015; Mechem et al. 2018) to use a relatively large number of nodes in order to reveal patterns that are otherwise not exposed when using a relatively small number of nodes. Our intention here is to highlight the two main regimes (the NPH and a land-falling cyclone) that are present during the boreal summer in this region; however, the patterns that occur in between these two main regimes in the SOM space are also important and should be discussed. If we use too few nodes, then the pattern variability within each regime will be too large and therefore the mean states of each node is not well representative of each individual case. Moreover, this variability means that differences between the nodes may not be statistically significant if too few nodes are used. In light of each reviewer's point concerning this topic, we have removed most instances of the word "regime". We reserve its use when talking about the NPH or land-falling cyclone circulation pattern, and we use the word "node" for all other patterns.

References:
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.

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 Atlantic near the Azores, J. Appl. Meteor. Clim., 57, 1273–1290, 2018.

Reviewer #1:

I have reviewed the manuscript "Linking large-scale circulation patterns to low-cloud properties" by Juliano and Lebo. The manuscript describes a study to classify the meteorology in the northeastern Pacific Ocean using self-organizing maps, an unsupervised machine learning algorithm. Based on this classification, the authors infer influences of continental aerosols on marine clouds in their study region. As the authors conclude, this study could provide valuable knowledge about ACI in this region, and it could provide a test for model representation of ACI. However, before the study can do so, I believe the authors would first need to address two major concerns:

We thank you for your thoughtful comments that strengthen the manuscript. We have addressed your individual comments below.

1. Now that even the most complex (opaque?) machine learning algorithms are available as off-the-shelf black boxes, the temptation is great to use them on any problem that comes along. To be a bit blunt, I think this work illustrates the dangers of doing so without carefully considering potential pitfalls. I am worried that the 20 meteorological regimes (really, 20? I am struggling to see the differences between many of them) are simply not robust. If the training dataset were slightly different (included an extra year at either end, an extra degree of latitude or longitude, ...), would the regimes look the same? Given that the manuscript's conclusions recommend these regimes be used for model evaluation, I think this is an important question to address; otherwise, if the models do not reproduce the regimes, we might end up falsely blaming the models for not including some non-robust idiosyncrasies of the training dataset that the machine learning happened to pick up on. In light of this (and in general), the authors' statement that unsupervised learning does not require a validation dataset is simply wrong.

For the first part of this comment, please see our general comment.

In response to the reviewer's concern about the robustness of the 20 different circulation patterns, we have generated three additional SOMs (Figs. R1-R3). As per the reviewer's suggestion, Fig. R1 shows the SOM output using NARR grids from years 2005-2013 (excluding years 2004 and 2014, which are included in Fig. 3 in the manuscript), and Figs. R2 and R3 show the SOM output using a smaller (reduce the domain by 3 grid points, or 96 km, in all four directions) and larger (increase the domain by 3 grid points, or 96 km, in all four directions) NARR region, respectively. In general, all three SOMs show a very similar range of circulation patterns. It is interesting to note that in Fig. R1, each node represents the same pattern as in Fig. 3 except with slight differences in the meteorological fields. However, in Figs. R2 and R3, while the patterns revealed by the SOM are similar, the position of the nodes has shifted so that the NPH and land falling cyclone regimes are represented by the nodes in the bottom left and top right corners, respectively, compared to the top right and bottom left corners in Figs. 3 and R1. We are unsure of the reason for these differences in Figs. R2 and R3; however, the position of the nodes does not appear to be important because all of the

circulation patterns are still represented by the SOM compared to the SOM output in Fig. 3. We have made a brief note about this sensitivity test at the end of Section 2.1.
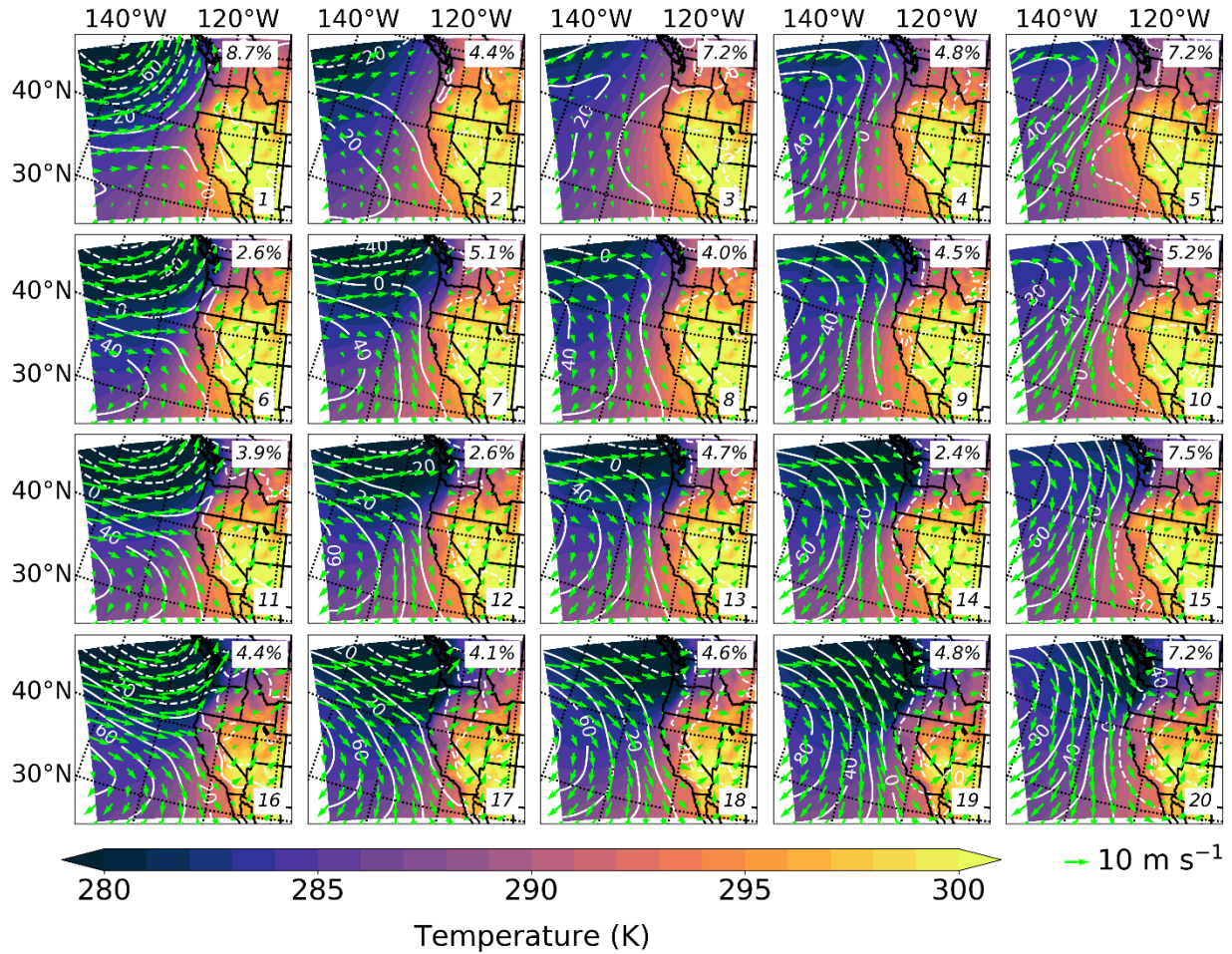
Figure R1. As in Fig. 3, except using NARR grids from years 2005-2013.
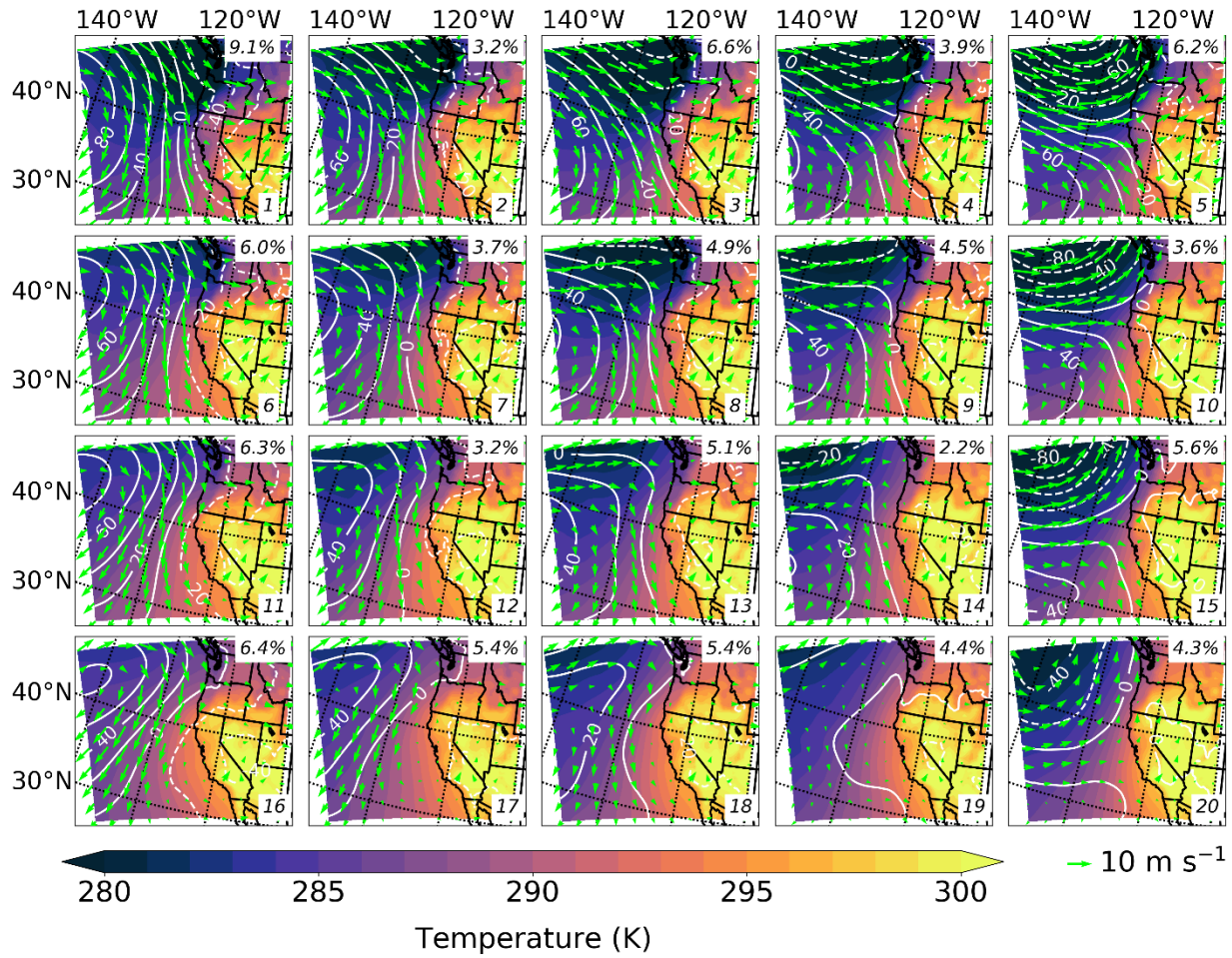
Figure R2. As in Fig. 3, except using a smaller NARR domain.
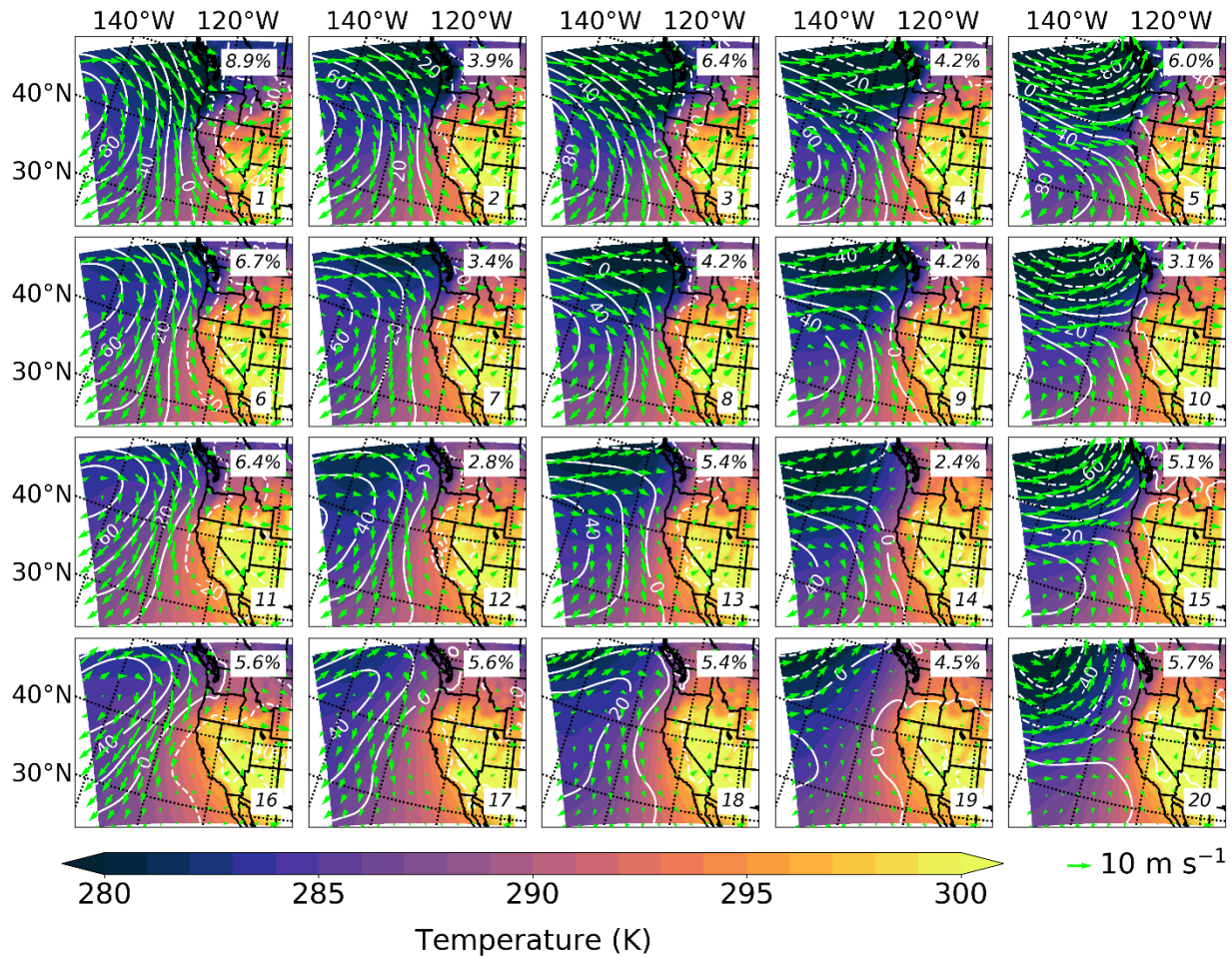
Figure R3. As in Fig. 3, except using a larger NARR domain.

Concerning the reviewer's last point, unfortunately it appears as though our writing was misleading and has unintentionally lead the reviewer astray. Lines 52-54 stated: "Because this dimensionality reduction method falls under the category of unsupervised learning, the user does not need to teach the algorithm on a separate training data set." In no way were we alluding to a validation data set when we mentioned a "separate training data set". Rather, we were making an (admittedly ambiguous) attempt to compare supervised to unsupervised machine learning techniques. The former technique requires the user to provide both inputs and outputs of a data set, and the machine learning model learns the relationship between the provided inputs and outputs. The learned relationships may then be used to predict the outputs of a separate, new data set where only the inputs are known. In unsupervised learning, only the inputs are known, and so the machine learning model is expected to reveal the structure of the data set. Therefore, in supervised machine learning, two data sets are needed – one for training and one for testing – while in unsupervised machine learning, only one data set is needed (for training).

We have therefore modified the text: "In unsupervised learning, the machine learning model is expected to reveal the structure of the input data set. Specifically in the case of training a SOM, the user is required to provide a 2-dimensional input array (typically time x space), and the node topology organizes itself to mimic the input data. That is, each SOM node represents a group of similar input vectors where an input vector is a single data sample. In the case here, we provide one input vector for each time of interest."

2. Independently of the methods, I am suspicious of the authors' conclusions about the influence of continental aerosols. The conclusion that "we attribute the variability in the satellite-retrieved cloud microphysical and radiative properties to aerosol forcing (first order effect) as opposed to meteorological factors (second order effect)" (l. 214 ff) would raise all kinds of red flags even if it were the result of careful quantitative analysis, as I struggle to think of any situation where aerosols have a first-order effect on cloud radiative properties on a regional scale. Here, it is presented on the basis of a number of "appears to" and "does not appear to" statements that leave me unconvinced.

We appreciate the reviewer's thought provoking discussion on this topic. To approach the topic from a different angle, and as per the reviewer's suggestion in the supplemental document, we have calculated various susceptibility relationships following Platnick and Twomey (1994). The MODIS variables that we consider for this analysis include cloud droplet number concentration (N), liquid water path (LWP), and shortwave cloud radiative effect (CRE_{SW}). [Note that we do not include MODIS cloud fraction (CF) due to our MODIS processing techniques. For our estimations of H and N, we use only pixels where both liquid phase and single-layer clouds are detected so that our estimations are as accurate as possible. Therefore, the CF retrievals considering only these pixels will lead to an underestimation of true CF because there may be multi-layer liquid pixels that we do not consider for our other variables.]

In light of the discussion regarding the influence of meteorology versus aerosol on cloud radiative properties, we are especially interested in the relationship between LWP and CRE_{SW} as well as that between N and CRE_{SW}. Using the three variables N, LWP, and CRE_{SW}, we calculate three susceptibility parameters: $\frac{\partial ln(LWP)}{\partial ln(N)}$, $\frac{\partial ln(CRE_{SW})}{\partial ln(LWP)}$, and $\frac{\partial ln(CRE_{SW})}{\partial ln(N)}$. The latter two relationships represent the meteorological and (presumed) aerosol forcings, respectively, on the shortwave cloud radiative effect. In general, susceptibility decreases from left to right on the node map, as one moves from the land-falling cyclone regime to the dominant NPH regime, e.g., from node 16 to node 12 to node 8 to node 5. The strong susceptibility signal represented by the $\frac{\partial ln(CRE_{SW})}{\partial ln(LWP)}$ relationship (values ranging from 0.65 to 0.75) suggest that changes in CRE_{SW} are strongly and positively related to changes in LWP, which is likely mainly due to meteorological forcing, as the $\frac{\partial ln(LWP)}{\partial ln(N)}$ relationship is relatively weak (values ranging from 0.03 to 0.16). While in all of the nodes the meteorological forcing does dominate over the aerosol forcing in the context of CRE_{SW}, we point out that the magnitude of the aerosol forcing on shortwave cloud radiative properties, represented by $\frac{\partial ln(CRE_{SW})}{\partial ln(N)}$, ranges

from approximately 40-49% of the meteorological forcing depending on the large-scale circulation pattern. Our results suggest that both the meteorological and aerosol forcings are first-order effects that cannot be neglected when examining the influence of large-scale circulation patterns on MBL cloud properties over the northeast Pacific Ocean.

We have added this text into Section 3.2 of the manuscript.

Additionally, to examine the relative roles of meteorology versus aerosol forcings, we have created a new figure (Fig. 8 in the manuscript; reproduced here as Fig. R4) that shows the scatter plot of LWP versus N, colored by CRE_{SW} for each of the nodes.
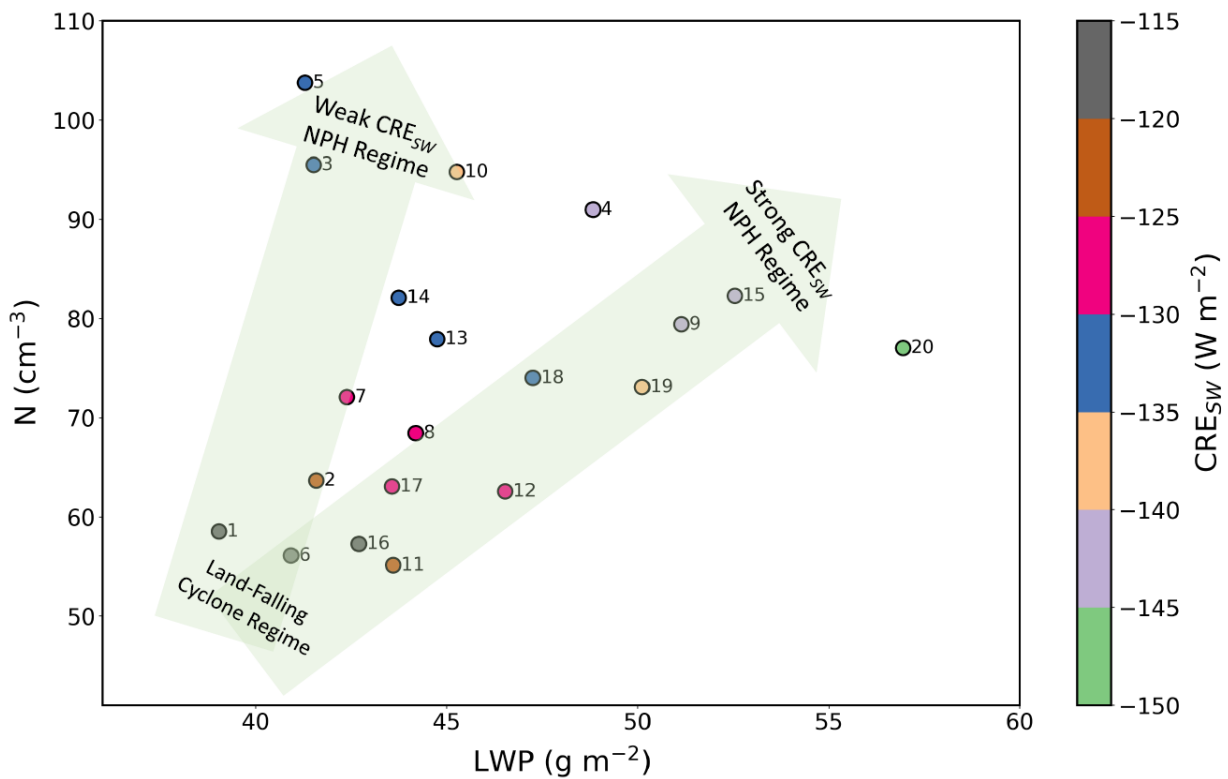
Figure R4. Scatter plot of LWP versus N, colored by CRE_{SW} for each of the nodes. Each value represents the median, and values are taken from Table 2.

We have added the following text to the manuscript to discuss this figure:
"In Fig. 8, we summarize the relationship been N, LWP, and CRE_{SW}. In the large-scale circulation patterns where the NPH is relatively suppressed and onshore flow dominates due to a land-falling cyclone (nodes 1, 6, 11, and 16), N is relatively small (ranging from 55.1 and 58.5 cm^{-3}) and LWP is relatively low (39.0 to 43.6 $g\ m^{-2}$), corresponding to relatively weak CRE_{SW} (-120.7 to -115.5 $W\ m^{-2}$). For the circulation patterns where the NPH is controlling the synoptic-scale setup (nodes 5, 10, 15, 20), while N is relatively large, LWP is relatively high,

and CRE_{SW} is relatively strong compared to the land-falling cyclone, these three variables exhibit much larger spread within the NPH regime compared to within the land-falling cyclone regime. For instance, median values for N, LWP, and CRE_{SW} are 103.7 cm^{-3}, 41.3 $g\ m^{-2}$, and -133.0 $W\ m^{-2}$ for node 5 and 77.0 cm^{-3}, 56.9 $g\ m^{-2}$, and -146.8 $W\ m^{-2}$ for node 20. Moreover, we again highlight the importance of both meteorology and aerosol on cloud properties: positive changes in LWP of ~19% and N of ~7% from node 1 to node 5 lead to +11% more reflective clouds, and positive changes in LWP of ~6% and N of ~77% from node 1 to node 5 lead to +15% more reflective clouds. Our results suggest that both the meteorological and aerosol forcings are first-order effects that neither cannot be neglected when examining the influence of large-scale circulation patterns on MBL cloud properties over the NEP.

We emphasize that Fig. 8 elucidates the large range in LWP for the NPH regime, which is in stark contrast to the relatively narrow range in LWP for the land-falling cyclone regime. As a result, for low LWP, the transition from the land-falling cyclone to NPH regime results in little to no change in LWP but a drastic increase in N and a commensurate increase in CRE_{SW}; this indicates that the meteorology exhibits little control on CRE_{SW} in the low-LWP NPH regime, and instead it is predominately driven by changes in N, which we relate to the offshore transport of continental aerosol. On the contrary, when moving from the land-falling cyclone regime to the high-LWP NPH regime, there is an increase in LWP as well as N, indicating the role of both factors in CRE_{SW}, as discussed above."

Responses to comments in the supplementary material

[We note that the reviewer highlighted some words or phrases without a comment. In most cases, the reviewer's comment is obvious (e.g., a misspelling); however, we are unable to respond to a few comments due to a lack of information.]

1. L29-31: We have reworded these two sentences to be clearer: "These shifts in the NPH are typically associated with […]" and "[…] these mesoscale phenomena develop in response to the reversed pressure gradient and are characterized by southerly MBL flow and a redevelopment of the stratiform cloud deck […]".

2. L38-40: We have changed all instances of CRF_{SW} to CRE_{SW} as per the reviewer's suggestion.

3. L46: We have removed the definitions of r_e and τ in response to Reviewer #2.

4. *Footnote on P2: We have removed this footnote and the associated text.*

5. *L54: Please see our response to the reviewer's first major comment.*

6. *L58: We have reworded two of the previous sentences to be clearer: "That is, each SOM node represents a group of similar input vectors where an input vector is a single data sample. In the case here, we provide one input vector for each time of interest."*

7. *Footnotes on P3: We have removed these footnotes and have added the information into the main text.*

8. *Figure 1 (now Figure 2): We choose to keep this figure in the main text because it shows clearly how the output SOM topological configuration is sensitive to the user's choice of node structure. The figure is also evidence that we have conducted a series of sensitivity tests, which addresses part of the reviewer's first major comment.*

9. *L91: This is the time period that we consider for both the NARR and MODIS data sets. The NARR data set is available from 1979 – present, while the MODIS data set is available from 2004 – present. This particular time period is chosen because we have already collected and processed these data from a previous study that was part of the first author's Ph.D. dissertation. The work presented here is currently unfunded and so additional data acquisition and analyses are not possible at this time.*

10. *L92: We have now added a figure (Fig. 1; reproduced here as Fig. R5) showing the geographical region.*
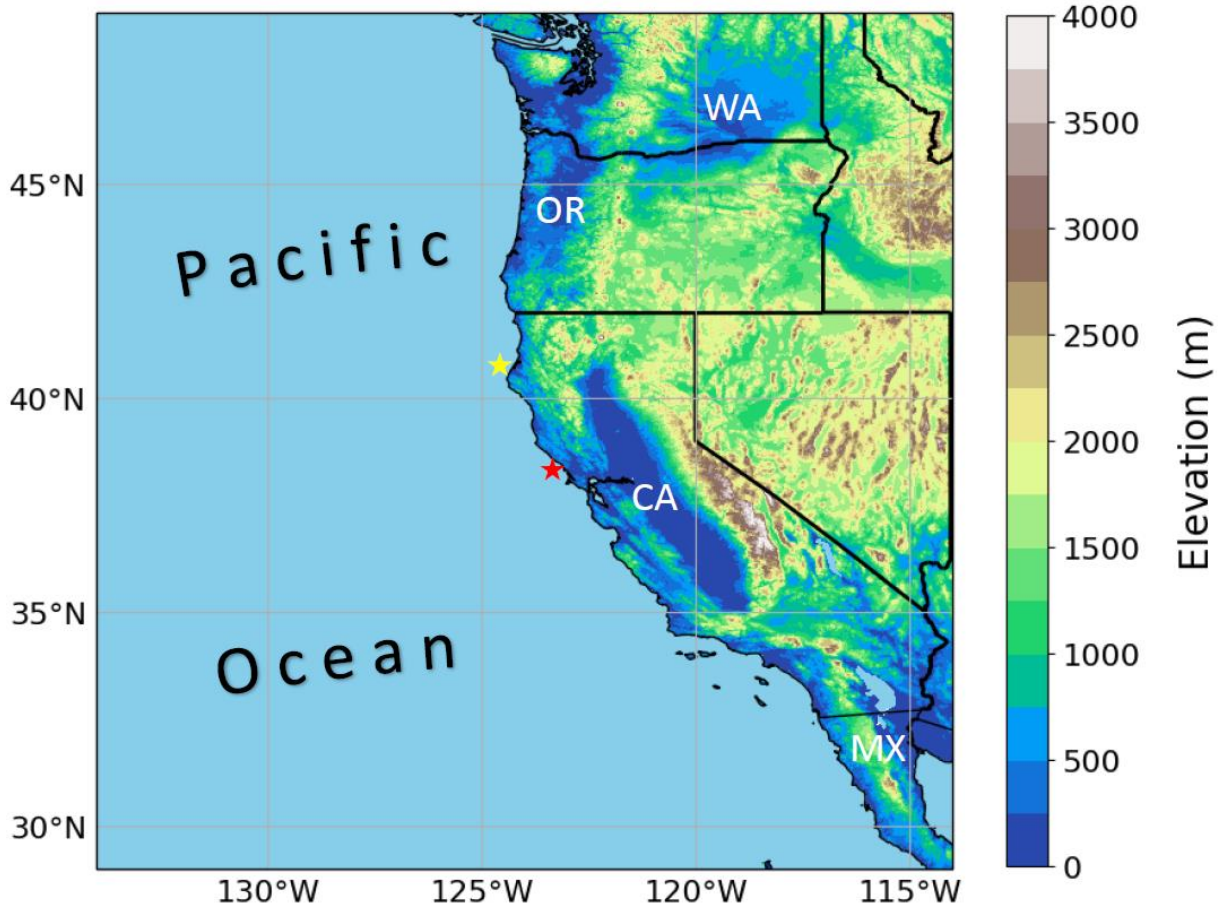
Figure R4. Study region covering the western United States and the NEP for the SOM analysis. United States abbreviations WA, OR, and CA represent Washington, Oregon, and California, respectively, while MX represents Mexico. The red and yellow stars denote the locations of buoys 46013 (Bodega Bay) and 46022 (Eel River), respectively.

11. *L101: We have also interpolated the MODIS data set to the same grid as the NARR output, and the results are qualitatively similar. Between the finer (~10 x 10 km) and coarser (~32 x 32 km) interpolated outputs, differences in the summary statistics shown in Table 2 are typically less than 5%. We have noted this in the text.*

12. *L106: Accepted.*

13. *L114: A Google Scholar search suggests that both 1360 and 1370 W m^{-2} are currently acceptable values. Therefore, we choose to maintain our use of 1370 W m^{-2}.*

14. L127: We have clarified our statement to be more specific: "For all of these nodes, and especially node 5, relatively warm temperatures are observed along the coastline near northern California, Oregon, and Washington."

15. L128-130: Because this is the first study to examine explicitly the morphology of warm-season large-scale circulation patterns over the northeast Pacific Ocean, we have reworded this sentence: "Overall, during the warm season months over the northeast Pacific Ocean, the SOM reveals two pronounced regimes – the dominant NPH and the land-falling cyclone – and in between a spectrum of large-scale meteorological conditions."

16. Figure 2 (now Figure 3): This is precisely what we are trying to show: the variability in land-sea contrast between the nodes that is related to the variability in circulation patterns and therefore low-level flow.

17. L132-133: Here we are referring to Figure 3 (now Figure 4); the text states "[…] as depicted by the observations from buoys 46013 and 46022 (Fig. 4)." Either way, it is likely that we do not see this in Fig. 2 because the 850-hPa flow, which is typically above the MBL top, is not necessarily indicative of the near-surface flow in the MBL. Moreover, and as you state, another plausible reason is because Fig. 3 shows the mean conditions.

18. L138: We agree and have added a topographical map of the study area showing the buoy locations, orography relevant to the flow pattern, and the state names (Fig. 1).

19. L185-192: We have not tested these hypotheses because such an analysis is beyond the scope of this study. However, we have modified the text to provide some thoughts for future work: "While we do not test these hypotheses here, we recommend that future studies examine them using, for example, backward trajectories, similar to previous studies that consider individual cases (e.g., Albrecht et al. 2019; Juliano et al. 2019a, b; Painemal et al. 2015), in combination with numerical simulations that explicitly treat aerosol.

20. *Figure 6: (a) Node 16 is certainly one of the cleanest regimes. Therefore, we choose to plot the actual frequency distributions for node 16 to highlight the apparent relationship between the meteorological pressure pattern and the associated cloud physical properties. (b) Yes, the statistics are over the ocean only. We have noted this in the figure caption. (c) The relatively low r_e/weak CRE regimes are likely due to a stronger impact of meteorological forcing, which is now discussed in the text.*

21. *L207: Please see our response to general comment #2.*

22. *L210: Please see our response to general comment #2.*

23. *L213: Please see our response to general comment #2.*

24. *L214-215: Please see our response to general comment #2.*

25. *L226: We have removed the word "significant" because it is difficult to quantify the role that the NPH location and intensity have on MBL cloud microphysical properties.*

26. *L230: Yes, the SOM reveals the frequency of occurrence of particular NPH regimes, which to our knowledge has not been reported in the literature.*

27. *L237: Please see our response to general comment #1.*

28. *L247: Please see our response to specific comment #9.*

29. *L249-251: We have added the following sentence to the end of the first paragraph in section 2: "While previous approaches have typically relied upon field studies or modeling case studies to highlight the relationship between synoptic-scale circulation*

patterns and cloud physics, the SOM exposes potential linkages through NPH regime classification."

30. L251: Please see our response to general comment #1.

Reviewer #2:

1 General review

This study introduces an analysis of cloud properties in the northeast Pacific region based on an unsupervised classification of 850 hPa-pressure situations.

The paper is well-written, has a clear structure, and figures are of good quality. The approach appears useful and promising in principle. In my view, the major potential for improvement lies in strengthening the link between the unsupervised classification obtained from the SOM analysis and the physical system. More in the details below.

We thank you for your thoughtful comments that strengthen the manuscript. We have addressed your individual comments below.

2 Details

• line 9: "must be accounted for" – why?

Global and regional climate models must account for the observed range of albedo/shortwave cloud radiative effect (CRE_{SW}) characteristic of marine stratiform clouds because these cloud types are critically important to the radiation budget. If the models do not replicate a similar range of CRE_{SW}, then they are not capturing the radiative impact of these cloud types properly (likely due to inaccurate physical parameterizations), and their representation of these cloud types in future climate scenarios is not trustworthy.

• 20: The literature cited here is several years old. What is the current state?

We have added three more recent studies that also suggest that climate models continue to struggle with MBL clouds.

• 28: offshore flow – where / what region?

We have modified the text for clarification: "[…] offshore component of the wind along the western United States coastline […]"

• 28: offshore flow – wind?

Please see our response to the previous comment.

• 31: why 'trapped'?

These mesoscale features are considered 'trapped' because 1) the southerly flow is forced along the coastal topography (to the east in this particular region) due to the Coriolis effect, and 2) the flow response is contained to within one Rossby radius of deformation of the coastline in the cross-coast direction (e.g., Nuss et al. 2000).

• 34: Is this about pollution aerosol from land? Please state explicitly.

We have modified the text to be more explicit: "Offshore flow […] likely enhances the transport of pollution aerosol from the continent to the ocean."

• 41ff: This paragraph seems to belong in the methods section

We agree and have moved this paragraph to the beginning of the methods section.

• 43ff: Are the definitions relevant here?

We have removed the definitions to improve readability.

• 46: How is N estimated and what are the uncertainties?

N is estimated using retrievals of τ and r_e after assuming an adiabatic cloud model (Bennartz 2007). This is the same method employed by Juliano et al. (2019a), who discuss the methods in more detail. We have added this information into the text.

According to Bennartz and Rausch (2017), Aqua-MODIS N uncertainties are of the order of 30% for marine stratocumulus.

• Footnote page 2: model output are 'data' as well.

We have removed the footnote.

• 56: Why SOM?

We have modified the text to include brief justification: "We choose to employ the SOM technique due to its ability to group similar patterns and therefore reveal dissimilar patterns that may be hidden in the large NARR data set considered here. To this end, we use the Matlab SOM Toolbox […]"

• 59: Explain "neighbourhood radius"

Neighborhood radius is explained in the paragraph below the procedure steps: "The neighborhood radius, which determines the number of nodes surrounding the winning node that nudge toward the input vector, slowly reduces to one (only the winning node is nudged) through the training period."

• 60: Explain "greates eigenvectors"

This step has been reworded to the following: "2. From the data set, determine the two eigenvectors that have the largest eigenvalues; initialize the SOM node weights linearly along these eigenvectors to provide a first approximation of the input data set"

• 61: To what end?

It is unclear what the reviewer is looking for here. This step is taken to aid with step 5.

• 62: On which basis?

The new neighborhood radius is a function of only time. Please see our response to the comment above, re: 'Explain "neighbourhood radius"'.

• 66: How is n determined?

The number of iterations, n, is determined a priori by the user. The choice is therefore subjective; however, the positions of the SOM nodes will be relatively stable after a certain number of iterations. In this particular instance, the results are negligible between a training with 5,000 iterations versus one with 25,000, suggesting that the solution converges before 5,000 iterations.

• 101: which cloud information? Which product level and collection?

This sentence has been modified: "The MODIS files (specifically, Level 2 MYD06 product) provide cloud information (namely, τ, r_e, cloud phase, cloud top pressure, and cloud top temperature) at 1-km horizontal spacing."

• Figure 2 etc.: You address the 20 different situations displayed here as "regimes". What is your definition of regime in this context? I am not convinced that all of these 20 situations are really fundamentally distinct in terms of processes. Are you? Why? Why do you see this as a useful categorization for your study? Beyond your study?

Please see our general comment.

• 177: What is 'sufficient accuracy' to you? The analysis you present above this statement is mostly a qualitative discussion. Can you generalize these findings?

We have removed the phrase "sufficient accuracy" because its presence did not contribute positively to the sentence. While the analysis in this paragraph is mostly qualitative, we believe that is sets the stage for the more quantitative analysis that follows.

3 Technicalities

• 27: meanders – moves?

Accepted.

• 36: ... motivate the present study. We consider...

We have modified the text: […] motivate the present study. Here, we consider […]"

• 43: remove "passively"

Accepted.

• 45: remove "passive"

We have removed this sentence.

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 distinguishable ~~patterns~~ regimes — a dominant NPH setup and a land-falling cyclone — and in between a spectrum of ~~regimes~~ large-scale patterns. 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~~, liquid water path, and shortwave cloud radiative effect (CRE_{SW}) 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~~ NEP. We find that CRE_{SW} spans from -146.8 to -115.5 W m^{-2}, indicating that the range of _observed_ MBL cloud properties must be accounted for in global and regional climate models. Our results demonstrate the value of combining reanalysis and satellite ~~observations~~ retrievals to help clarify the relationship between synoptic-scale dynamics and cloud ~~microphysics~~ physics.

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, 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 ($\sim$10 K) and sharp $\mathcal{O}(100-500\text{ m})$ 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)~~ (global shortwave cloud radiative effect (CRE_{SW}) of $\sim$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 by global climate models ~~(e.g., Palmer and Anderson, 1994; Del~~

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

During boreal summer, the northeast Pacific Ocean (NEP) 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~~ moves toward the north and east ~~(e.g., Mass and Bond, 1996; Fewings et al., 2016; ?). These events~~ (e.g., Mass and Bond, 1996; Fewings et al., 2016; Juliano et al., 2019b). These shifts in the NPH are typically associated with an increase in the offshore ~~flow~~ component of the wind along the western United States coastline 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~~ develop in response to the reversed pressure gradient and are characterized by southerly MBL flow and 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)] and reflective (stronger CRE_{SW}) than those forming under typical northerly flow conditions due to aerosol-cloud interactions ~~(?)~~ (Juliano et al., 2019b). Offshore flow, which is a requirement for the initiation of a CTD, likely enhances the transport of pollution aerosol from the continent to the ocean. These results motivate the present study ~~because in this study~~. Here, we consider data over a relatively long time span to identify objectively the ~~most prevalent~~ spectrum of synoptic-scale dynamical ~~regimes~~ patterns during boreal summer. We aim to improve the current understanding of the relationship between these synoptic-scale patterns, mesoscale cloud microphysics, and ~~CRF~~ CRE_{SW} over the ~~northeast Pacific Ocean~~ NEP — an issue identified previously as "vital" (Stevens and Feingold, 2009).

2 Methods and data

To diagnose the various NPH ~~regimes~~ circulations, we first use the ~~NCEP~~ National Centers for Environmental Prediction (NCEP) North American Regional Reanalysis (NARR) to develop a self-organizing map (SOM) ~~.~~ covering the western United States and the NEP. Our study domain is shown in Fig. 1. 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~~, LWP, and CRE_{SW} because these variables most clearly accentuate the connection between large-scale dynamics and MBL cloud properties. While previous approaches have

typically relied upon field studies or modeling case studies to highlight the relationship between synoptic-scale circulation and

55 cloud physics, the SOM exposes potential linkages through NPH pattern classification over longer time periods.

3 ~~Methods and data~~
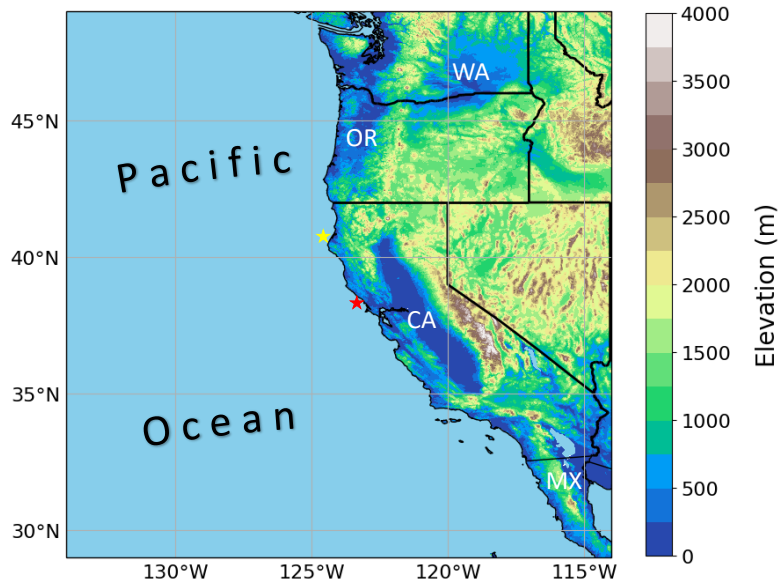
Figure 1. Study region covering the western United States and the NEP for the SOM analysis. United States abbreviations WA, OR, and CA represent Washington, Oregon, and California, respectively, while MX represents Mexico. The red and yellow stars denote the locations of buoys 46013 (Bodega Bay) and 46022 (Eel River), respectively.

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~~

60 ~~this dimensionality reduction method falls under the category of~~. In unsupervised learning, the ~~user does not need to teach the algorithm on a separate training~~ machine learning model is expected to reveal the structure of the input data set. ~~Instead~~Specifically in the case of training a SOM, the user is required to provide ~~only~~ a 2-dimensional ~~array~~input array (typically time × space), and the node topology organizes itself to mimic the input data. That is, each ~~node will represent~~ SOM node represents a group of similar input vectors ~~.~~ where an input vector is a single data sample. In the case here, we provide one input vector for each

65 time of interest.

---

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

We choose to employ the SOM technique due to its ability to group similar patterns and therefore reveal dissimilar patterns that may be hidden in the large NARR data set considered here. To this end, 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:

70

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

2. ~~Initialize the~~ From the data set, determine the two eigenvectors that have the largest eigenvalues; initialize the SOM node weights linearly along ~~its greatest eigenvectors~~ these eigenvectors to provide a first approximation of the input data set

3. Present all vectors from the input data and calculate the Euclidean distance between each input vector and each node, where the Euclidean distance, $d(a,b) = \sqrt{(x_a - x_b)^2 + (y_a - y_b)^2}$ given two points a and b in $\{x,y\}$ space

75

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 [1]

6. Update the weight of each node [1] — where the new weight is equal to the weighted average of each input data vector to which that node or any nodes in its neighborhood responded — after a single iteration

80

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~~2). Quantization 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~~2: the QE decrease is most pronounced as the number of nodes increases from approximately 9 to 20, and the TE 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

85

90

---

[1] ~~Given two points a and b in $\{x,y\}$ space, the Euclidean distance, $d(a,b) = \sqrt{(x_a - x_b)^2 + (y_a - y_b)^2}$.~~
[1] ~~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.~~

95 surrounding the 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 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

100 small impact on the final 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~~data set, each input vector is associated with one of the map nodes.
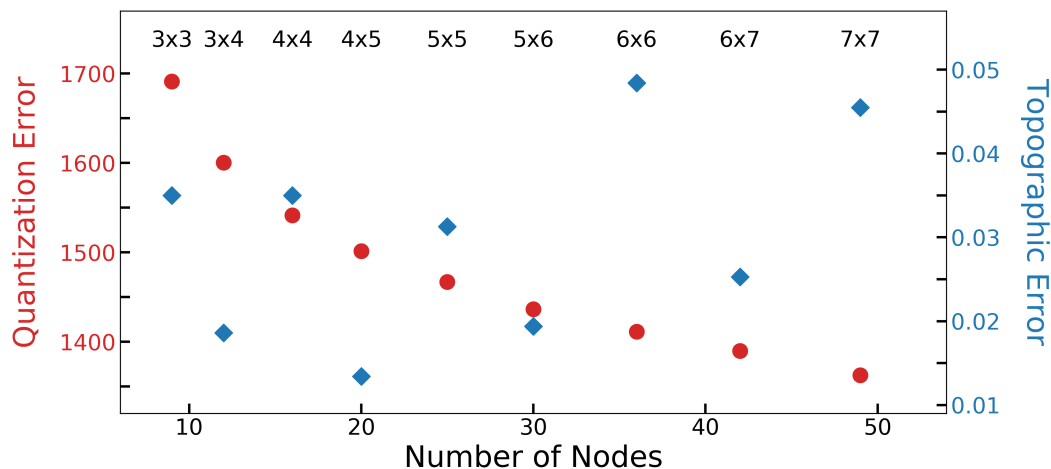
Figure 2. The quantization (red circles; left axis) and topographic (blue diamonds; right axis) errors for each SOM configuration tested in this study. SOM node topologies (rows $\times$ columns) range from 3×3 to 7×7; we choose to use a 4×5 node map.

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

105 accurately represent the location and strength of the NPH[1]~~.~~. We note that we also explore using the sea-level pressure 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 sea-level pressure is extrapolated using a standard atmosphere assumption due to elevated terrain. While this extrapolation does also occur at 850 hPa, there are far fewer locations whose surface pressure is often lower than 850 hPa. For the 2-dimensional input array, we use the 0000 UTC ~~NCEP North American Regional Reanalysis~~

110 ~~(NARR)~~ NARR 32-km product for each day during the months June through September from 2004 through 2014. The use of a single time per day is permissible due to the time scale at which the pressure patterns of interest evolve and to also correspond with the satellite data. Spatial anomalies are calculated for each day by subtracting the domain-averaged 850-hPa height from

---

[1] ~~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.~~

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. We note that our SOM results are largely insensitive to the years (excluding 2004 and 2014) and the domain size (±3 grid points (96 km) in all four directions) considered (not shown).

2.2 Data sets

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 over relatively short time periods. Moreover, any two consecutive images (∼5 minutes apart) are stitched together and counted as one sample. The MODIS files (specifically, the Level 2 MYD06 product) provide cloud information (namely, τ, r_e, cloud phase, cloud top pressure, and cloud top temperature) at 1-km horizontal spacing. We then interpolate these data to a uniform $1/10° × 1/10°$ (∼10 × 10 km) grid to be closer to the native horizontal grid spacing (∼32 km) of the NARR output without losing too much detail. We note that interpolating the MODIS data set to the NARR grid yields qualitatively and quantitatively similar results: differences in the summary statistics shown in Table 2 are typically less than 5% (not shown).

2.3 MODIS processing

For the MODIS retrievals, values of r_e and τ are calculated utilizing a bispectral solar reflectance method (Nakajima and King, 1990), whereby ~~extinction~~ reflectance 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 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~~ r_e after assuming an adiabatic cloud model (Bennartz, 2007), similar to the method used in Juliano et al. (2019b). 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~~ CRE_{SW} may then be calculated as ~~$CRF_{SW} = \left(\frac{S_o}{4}\right)(\alpha_o - \alpha_c)$~~ $CRE_{SW} = \left(\frac{S_o}{4}\right)(\alpha_o - \alpha_c)$, where S_o is the solar constant (1370 W m^{-2}), and α_o is the ocean albedo [0.10 (10%)]. The MODIS processing techniques are expounded in ~~?~~Juliano et al. (2019b).

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 NEP from 2004 to 2014 (Fig. 3). There are several key features to discuss. The leftmost part of the map (nodes 1, 6, 11, and 16) represents a regime 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 low 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 patterns, and the presence of the NPH becomes more noticeable. The rightmost portion of the map (nodes 5, 10, 15, and 20) represents a regime 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, especially node 5, relatively high temperatures are observed along the coastline in the northern portion of the domain. Approximately 26.0% of all days in the data set fall under these four nodes with a dominant NPH. Overall, during the warm season months over the NEP, the SOM reveals two pronounced regimes — the dominant NPH and the land-falling cyclone — and in between a spectrum of large-scale meteorological conditions.

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. 4; cf. Fig. 1). Offshore flow generates a weakening or reversal in the alongshore pressure gradient that drives southerly flow, while onshore flow is blocked 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 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 is important for various meteorological applications, such as ocean upwelling and offshore wind energy forecasting.

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).
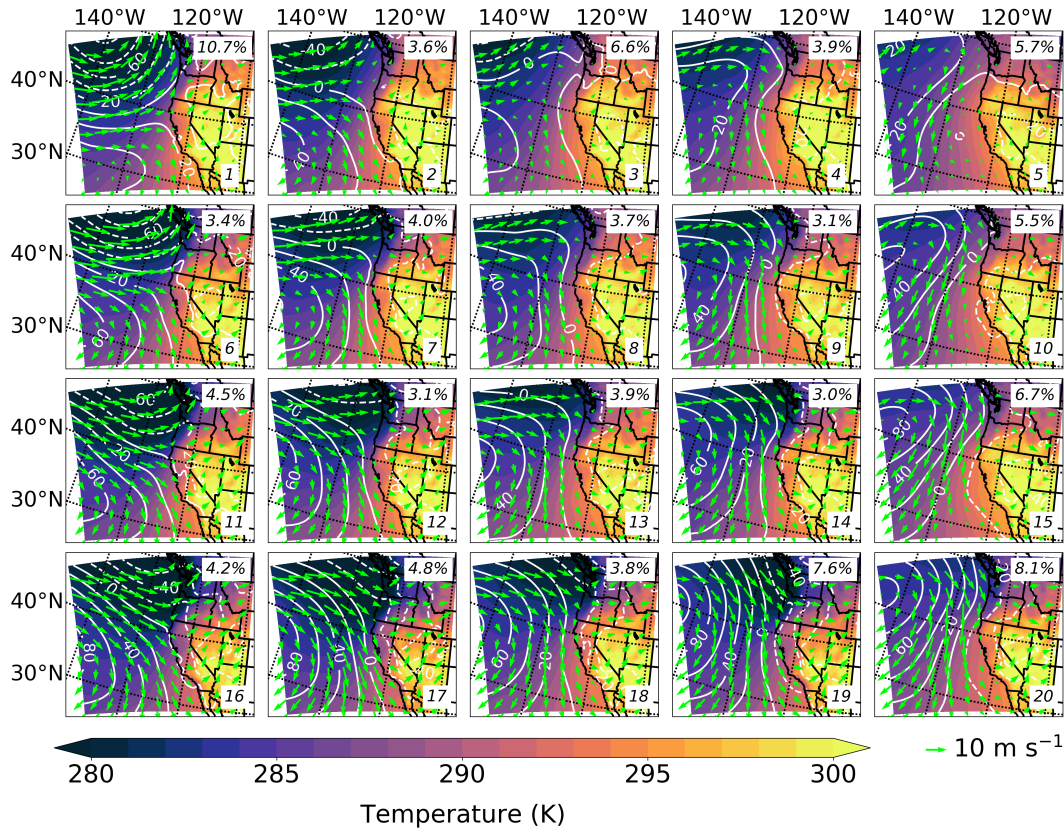
Figure 3. ~~The synoptic-scale~~ Synoptic-scale 850-hPa spatial anomaly height field (white lines contoured every 20 m), ~~temperature~~ in addition to temporal mean temperatures (contoured every 1 K according to colorbar) ~~,~~ and wind ~~vector~~ vectors (lime green arrows with 10-m s^{-1} reference vector) ~~fields~~ as represented by the 4 × 5 SOM node topology. Each panel represents ~~the mean conditions for each~~ a node (number displayed in the bottom right corner), while the percentage frequency of occurrence is displayed in the top right corner.

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~~ features a weak regional height gradient, shows a strong increase in frequency of occurrence over time (frequencies of 15.8%, 14.5%, 28.9%, and 40.8% in June, July, August, and September, respectively).

180   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~~ the nodes 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. Conversely, those patterns toward the right feature northerly, and even northeasterly, flow offshore due to a ~~dominate~~ dominant warm-core NPH. Moreover, several of the ~~regimes (nodes~~ nodes (3, 4, and 5) feature a noticeably weak 850-hPa

185 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.

Percentage of Hours with Southerly Flow

Buoy 46013

7.1	1.5	5.4	7.4	38.5
10.1	3.5	10.2	8.8	39.2
7.0	3.9	12.8	21.3	34.3
14.0	12.1	21.7	16.5	29.3

Buoy 46022

28.3	6.1	1.9	3.8	7.6
22.7	1.8	3.4	1.6	10.8
23.8	5.3	6.8	2.7	16.6
48.2	9.4	20.1	10.4	18.4

Figure 4. ~~The percentage~~ Percentage of southerly flow hours recorded at each buoy site along the California coastline for each ~~regime~~ node 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).

Therefore, one might expect to see relatively high N and small r_e values dominate in these ~~regimes~~ nodes because they appear to be influenced strongly by continental air masses.

3.2 MODIS cloud retrievals

Figures ~~4 and~~ 5 and 6 show the mean N and r_e values from MODIS that are associated with each node. In Fig. ~~4~~5, the red (yellow) end of the colorbar corresponds to relatively low (high) N, and in Fig. ~~5~~6, the red (yellow) end of the colorbar corresponds to relatively small (large) r_e. Therefore, yellow regions in Fig. ~~4~~5 and red regions in Fig. ~~5~~6 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~~3 and the MODIS cloud properties shown in Figs. ~~4 and 5.~~ 5 and 6. 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 numerous and smaller) cloud 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~~3, node 5 is characterized by distinct offshore continental flow at 850 hPa, in addition to very weak ~~flow (and even southerly flow)~~ northerly or southerly flow near the shoreline. These 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

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

205   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 to population centers (namely Los Angeles, California, San Diego, ~~and Tijuana~~California, and Tijuana, Mexico) and wild-

210   fire 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 suggests that the pervasive Catalina Eddy — a phenomenon linked to the generation of CTDs (e.g., Skamarock et al., 2002) — may transport pollution offshore and toward the north (Wakimoto, 1987).

215   We hypothesize that the MODIS retrievals presented here show clearly that the 1st aerosol indirect effect (Twomey, 1977)
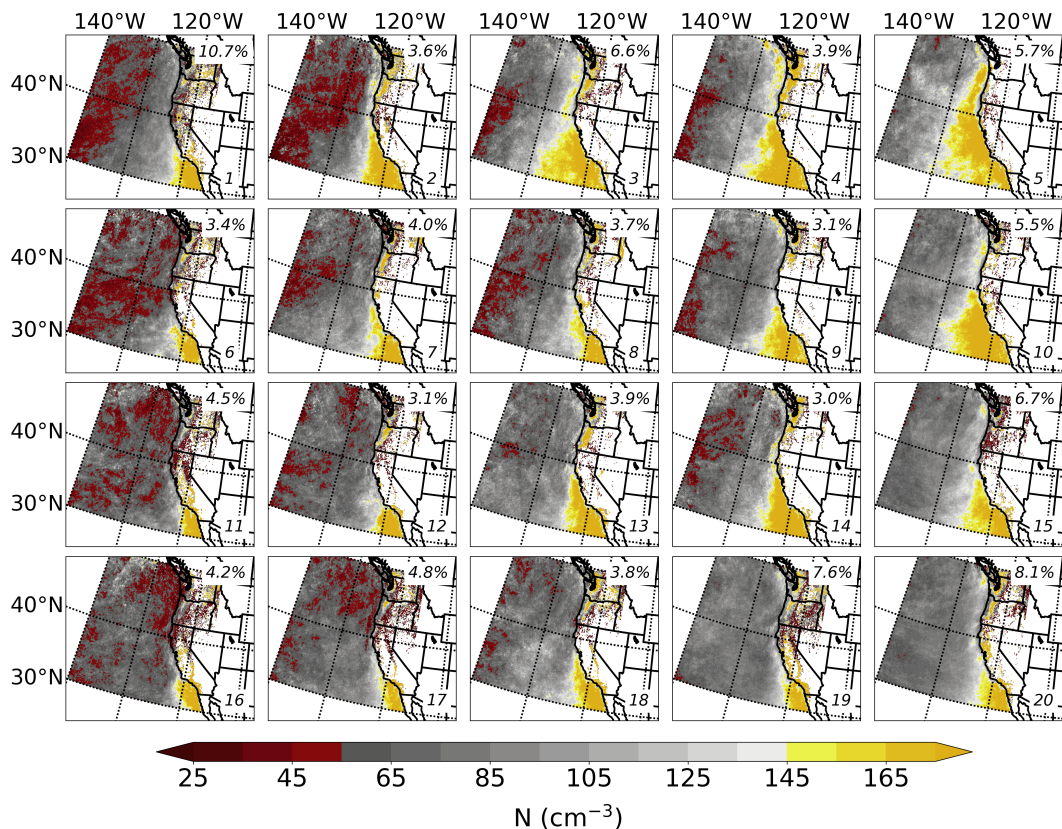
Figure 5. ~~The~~ MODIS estimation of N (contoured every 10 cm^{-3} according to the colorbar) for each ~~regime~~ node in the 4 × 5 SOM ~~node~~ topology. Each panel represents the mean conditions for each node (number displayed in the bottom right corner), while the percentage frequency of occurrence is displayed in the top right corner.

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), increases the number of cloud condensation nuclei

220 (CCN) and therefore cloud droplets. ~~These effects~~ While we do not test these hypotheses here, we recommend that future studies examine them using, for example, backward trajectories — similar to previous studies that consider individual cases (e.g., Painemal et al., 2015; Albrecht et al., 2019; Juliano et al., 2019a, b) — in combination with numerical simulations that explicitly treat aerosol. These potential aerosol-cloud interactions 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

225 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
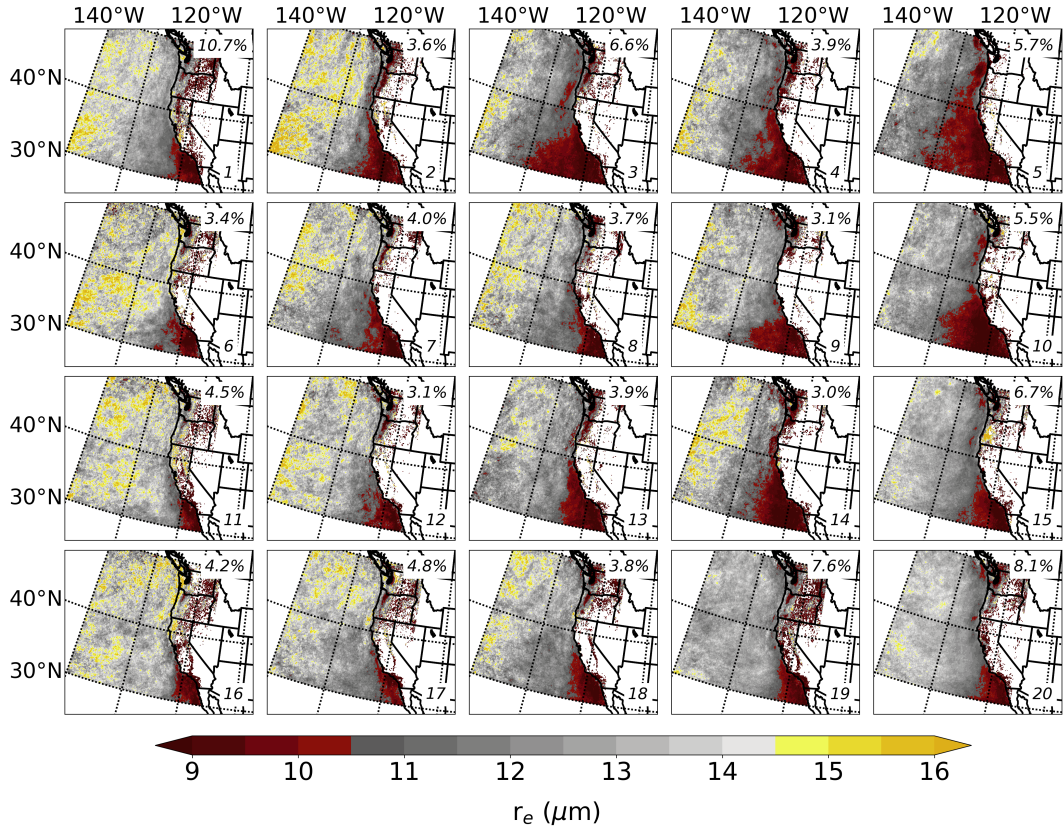
Figure 6. As in Fig. ~~4~~ 5 except for the MODIS retrieval of r$_e$ (contoured every 0.5 μm according to the colorbar).

due to strong, onshore flow. In general, the nodes display varying extensions according to the synoptic-scale ~~regime~~circulation pattern.

Frequency distributions reveal that between the various SOM nodes, ~~N, r_e, and CRF$_{SW}$~~ cloud micro- and macrophysical properties exhibit a broad range that is dependent on the prevailing synoptic-scale pattern (Fig. ~~6~~7; cf. Table 2 for median values). The distributions confirm that node 5, in addition to nodes 3, 4, and 10, represent the scenarios where MBL clouds are characterized by relatively high N ~~, and~~ small r_e ~~, and strong CRF$_{SW}$~~ compared to the other meteorological ~~regimes~~nodes. The median values of N ~~, r_e, and CRF$_{SW}$ are 93.2~~ and r_e are 95.5 cm^{-3} ~~, 10.4~~ and 10.3 μm ~~, and -129.7 W m^{-2}~~ for node 3, ~~93.3~~91.0 cm^{-3} ~~, 10.6~~ and 10.9 μm ~~, and -138.1 W m^{-2}~~ for node 4, 92.8 103.7 cm^{-3} ~~, 10.5~~ and 10.0 μm ~~, and -132.3 W m^{-2}~~ for node 5, and ~~90.6~~94.8 cm^{-3} ~~, 10.8~~ and 10.5 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~~ 12 — the distributions are shifted appreciably such that the median values of N ~~, r_e, and CRF$_{SW}$ are 62.5~~ and r_e are 56.1 cm^{-3} ~~,~~ ~~12.0~~ and 12.3 μm ~~, and -123.9 W m^{-2}~~ for node 6, ~~57.8~~55.1 cm^{-3} ~~, 12.3~~ and 12.5 μm ~~, and -121.7 W m^{-2}~~ for node 11, ~~57.2~~
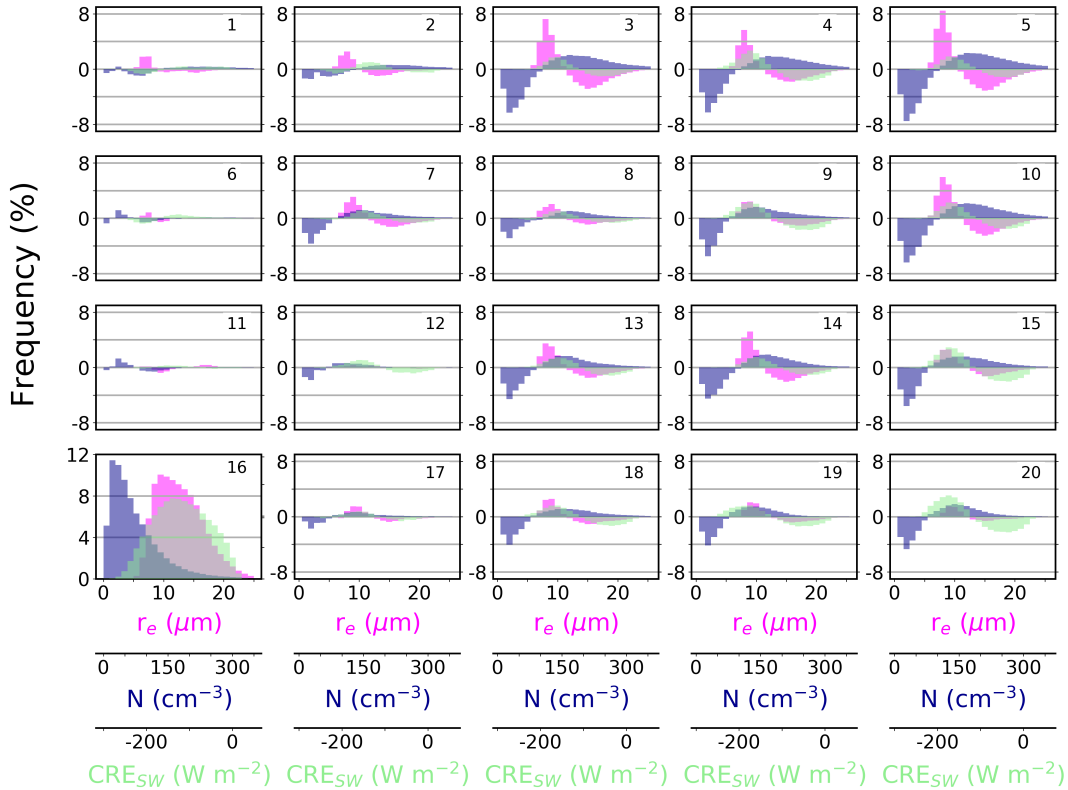
Figure 7. Frequency distributions of r_e (units in μm; colored magenta), N (units in cm^{-3}; colored dark blue), and ~~CRF~~CRE$_{SW}$ (units in W m^{-2}; colored light green) for each ~~regime~~ node over the ocean in the 4×5 SOM ~~node~~ topology. The node number is shown in the top right corner of 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~~CRE$_{SW}$ 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.

240   62.6 cm^{-3} ~~, 12.5~~ and 12.2 μm ~~, and -124.1 W m^{-2} for node 16, and 61.9~~ for node 12, and 57.3 cm^{-3} ~~, 12.0~~ and 12.3 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.~~ for node 16.

     To explore the potential impact of the regional meteorology associated with each of the synoptic-scale ~~regimes~~ patterns

245   — compared 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~~ meteorological variables do not show much variability among

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~~ CRE_{SW} (cf. Fig. ~~6~~7), LWP, and H as well as those of meteorological variables LCF, LTS, and q_v ~~(not~~. Distributions are calculated over the oceanic area ~~)~~ in Figs. 3, 5, and 6.

Node	LCF, %	LTS	q_v, g kg^{-1}	r_e, μm	LWP, g m^{-2}	H, m	N, cm^{-3}	CRE_{SW}, W m^{-2}
1	75	15.0	4.9	12.0	39.0	213.0	58.5	-115.5
2	74	15.2	5.0	11.8	41.6	223.2	63.7	-122.5
3	75	14.8	5.0	10.3	41.5	223.1	95.5	-132.1
4	74	15.0	5.2	10.9	48.8	241.2	91.0	-141.7
5	75	14.8	4.9	10.0	41.3	221.9	103.7	-133.0
6	74	15.5	4.8	12.3	40.9	219.5	56.1	-117.2
7	74	15.7	5.0	11.4	42.4	225.9	72.1	-126.4
8	76	15.1	5.0	11.6	44.2	228.4	68.4	-127.1
9	75	15.2	5.1	11.5	51.1	248.1	79.4	-140.7
10	75	15.1	5.1	10.5	45.3	232.7	94.8	-137.2
11	73	15.6	4.8	12.5	43.6	229.2	55.1	-120.7
12	75	15.7	4.9	12.2	46.5	238.3	62.6	-128.6
13	75	15.6	5.0	11.2	44.8	232.8	77.9	-132.1
14	74	15.2	5.2	10.9	43.8	231.0	82.1	-131.5
15	75	15.4	5.1	11.5	52.5	248.8	82.3	-143.7
16	73	15.6	4.5	12.3	42.7	227.4	57.3	-119.8
17	73	15.9	4.5	11.9	43.6	232.1	63.1	-125.5
18	74	16.0	5.0	11.5	47.3	240.5	74.0	-134.7
19	74	15.8	4.7	11.7	50.1	250.5	73.1	-137.5
20	75	15.8	5.0	11.9	56.9	261.4	77.0	-146.8

250 the different nodes. MODIS retrievals of LWP and cloud depth (H), which are macrophysical quantities that may serve as indicators of meteorological forcing, show a weak trend of increasing values as one moves from left to right across the map; however, there is large spread in LWP and H between the NPH regime patterns. To further investigate the relative influences of meteorological versus aerosol forcings on the satellite-retrieved cloud microphysical and radiative properties, we examine various susceptibility relationships following Platnick and Twomey (1994).

255   For this susceptibility analysis, we consider the MODIS variables N, LWP, and CRE_{SW}. Using these three variables, we calculate three susceptibility parameters: $\partial\ln(\text{LWP})/\partial\ln(N)$, $\partial\ln(CRE_{SW})/\partial\ln(\text{LWP})$, and $\partial\ln(CRE_{SW})/\partial\ln(N)$ (Table 3). The latter two relationships represent the meteorological and aerosol forcings, respectively, on the shortwave CRE. In general, susceptibility decreases from left to right on the node map, as one moves from the land-falling cyclone regime to the dominant NPH regime, e.g., from node 16 to node 12 to node 8 to node 5. The strong susceptibility signal represented

260 by the $\partial\ln(CRE_{SW})/\partial\ln(\text{LWP})$ relationship (values ranging from 0.65 to ~~aerosol forcing~~ 0.75) suggests that changes in CRE_{SW} are strongly and positively related to changes in LWP, which is likely mainly due to meteorological forcing, as the

Table 3. Susceptibility parameters calculated for each node. Here we use the same data that are used to calculate the values in Table 2.

Node	$\partial\ln(\text{LWP})/\partial\ln(N)$	$\partial\ln(\text{CRE}_{SW})/\partial\ln(\text{LWP})$	$\partial\ln(\text{CRE}_{SW})/\partial\ln(N)$
1	0.11	0.75	0.33
2	0.06	0.72	0.31
3	0.09	0.71	0.32
4	0.04	0.66	0.28
5	0.03	0.69	0.29
6	0.11	0.74	0.33
7	0.09	0.72	0.32
8	0.10	0.72	0.32
9	0.09	0.68	0.31
10	0.04	0.68	0.29
11	0.12	0.74	0.33
12	0.12	0.72	0.33
13	0.08	0.69	0.31
14	0.08	0.70	0.32
15	0.03	0.66	0.27
16	0.16	0.74	0.36
17	0.16	0.74	0.36
18	0.08	0.68	0.31
19	0.14	0.69	0.34
20	0.05	0.65	0.28

$\partial\ln(\text{LWP})/\partial\ln(N)$ relationship is relatively weak (values ranging from 0.03 to 0.16). While in all of the nodes the meteorological forcing does dominate over the aerosol forcing in the context of CRE_{SW}, we point out that the magnitude of the aerosol forcing on shortwave cloud radiative properties, represented by $\partial\ln(\text{CRE}_{SW})/\partial\ln(N)$, ranges from approximately 40 to 49% of the meteorological forcing depending on the large-scale circulation pattern.

In Fig. 8, we summarize the relationship been N, LWP, and CRE_{SW}. In the large-scale circulation patterns where the NPH is relatively suppressed and onshore flow dominates due to a land-falling cyclone (nodes 1, 6, 11, and 16), N is relatively small (ranging from 55.1 and 58.5 cm^{-3}) and LWP is relatively low (39.0 to 43.6 g m^{-2}), corresponding to relatively weak CRE_{SW} (~~first order effect)as opposed to meteorological factors (second order effect).~~ -120.7 to -115.5 W m^{-2}). For the circulation patterns where the NPH is controlling the synoptic-scale setup (nodes 5, 10, 15, 20), while N is relatively large, LWP is relatively high, and CRE_{SW} is relatively strong compared to the land-falling cyclone, these three variables exhibit much larger spread within the NPH regime compared to within the land-falling cyclone regime. For instance, median values for N, LWP, and CRE_{SW} are 103.7 cm^{-3}, 41.3 g m^{-2}, and -133.0 W m^{-2} for node 5 and 77.0 cm^{-3}, 56.9 g m^{-2}, and -146.8 W m^{-2} for node 20. Moreover, we again highlight the importance of both meteorology and aerosol on cloud properties: positive changes in LWP of ~19% and N of ~7% from node 1 to node 5 lead to +11% more reflective clouds, and positive changes in LWP

265

270

275

of ∼6% and N of ∼77% from node 1 to node 5 lead to +15% more reflective clouds. Our results suggest that both the meteorological and aerosol forcings are first-order effects that neither cannot be neglected when examining the influence of large-scale circulation patterns on MBL cloud properties over the NEP.

We emphasize that Fig. 8 elucidates the large range in LWP for the NPH regime, which is in stark contrast to the relatively
280 narrow range in LWP for the land-falling cyclone regime. As a result, for low LWP, the transition from the land-falling cyclone to NPH regime results in little to no change in LWP but a drastic increase in N and a commensurate increase in CRE_{SW}; this indicates that the meteorology exhibits little control on CRE_{SW} in the low-LWP NPH regime, and instead it is predominately driven by changes in N, which we relate to the offshore transport of continental aerosol. On the contrary, when moving from the land-falling cyclone regime to the high-LWP NPH regime, there is an increase in LWP as well as N, indicating the role of
285 both factors in CRE_{SW}, as discussed above.

Figure 8. Scatter plot of LWP versus N, colored by CRE_{SW} for each of the nodes. Each value represents the median, and values are taken from Table 2.

Overall, our SOM results elucidate the apparent coupling between NPH dynamics and mesoscale MBL cloud properties . That isthrough both meteorological and aerosol effects. In terms of the impact of large-scale circulation on cloud physics through aerosol forcing, 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)
290 of MBL clouds. Moreover, we hypothesize that a weaker regional pressure gradient allows for the transport of aerosol by the coastal jet 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. We note that other factors not explicitly accounted for here,

295 such as aerosol composition, turbulence, and sea surface temperature, may also play an important role in the linkage between large-scale meteorology and low-cloud properties.

4 Discussion and conclusions

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~~ micro- and macrophysical properties offshore

300 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~~NEP. Results from the SOM, using an 11-year period of NARR data, reveal several distinct synoptic patterns present ~~during the Northern Hemisphere~~ over the NEP during the warm season as well as their frequencies of occurrence; however, most notable is the high frequency of pronounced offshore continental flow and generally weak flow. Incorporating MODIS observations into

305 the analysis yields a connection between the synoptic-scale dynamics and mesoscale cloud ~~microphysics. Specifically, more instances of polluted clouds (as indicated by high~~ properties. That is, combining the SOM approach with the satellite retrievals reveals that synoptic-scale circulation patterns modify both the regional meteorology and aerosol transport. Specifically, the land-falling cyclone regime is characterized by relatively low 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),~~ small LWP, and weak

310 CRE$_{SW}$. In comparison, the NPH regime generally shows relatively high N, large LWP, and strong CRE$_{SW}$, albeit these variables exhibit much larger spread compared with those of the land-falling cyclone regime. In the middle of the SOM, the circulation patterns exhibit a smooth transition between these two regimes both in terms of the meteorological setup and MBL cloud properties.

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

315 because the synoptic-scale flow–cloud microphysics ~~relationship~~ relationships 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~~microphysical/radiative ~~forcing~~effects. One could then quantify the impact of using the radiative ~~forcing~~ effect from the observed SOM relationship with the modeled 850 hPa height field rather than the model-predicted radiative ~~forcing~~ effect over the semi-permanent marine

320 stratiform regions. Also, this analysis could be extended to evaluate in a statistical sense the ability of long-term simulations to replicate each large-scale ~~regime~~pattern.

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

325 novel method to link well-resolved synoptic-scale features to cloud microphysics and ~~shortwave radiative forcing~~the shortwave

radiative effect. Because the 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 prop-
330 erties. In general, using machine learning techniques to connect large-scale circulation patterns to cloud microphysics, which 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
335 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.

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 observa-
340 tions 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.

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.

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

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,
350 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.

Albrecht, B., Ghate, V., Mohrmann, J., Wood, R., Zuidema, P., Bretherton, C., Schwartz, C., Eloranta, E., Glienke, S., Donaher, S., Sarkar, M., McGibbon, J., Nugent, A. D., Shaw, R. A., Fugal, J., Minnis, P., Paliknoda, R., Lussier, L., Jensen, J., Vivekanandan, J., Ellis, S., Tsai, P., Rilling, R., Haggerty, J., Campos, T., Stell, M., Reeves, M., Beaton, S., Allison, J., Stossmeister, G., Hall, S., and Schmidt, S.: Following the Evolution of Boundary Layer Cloud Systems with the NSF–NCAR GV, *Bull. Amer. Meteor. Soc.*, 100, 93–121, 2019.

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.

Bender, F. A., Engström, A., and Karlsson, J.: Factors Controlling Cloud Albedo in Marine Subtropical Stratocumulus Regions in Climate Models and Satellite Observations, *J. Climate*, 29, 3559–3587, 2016.

Bender, F. A., Frey, L., McCoy, D. T., Grosvenor, D. P., and Mohrmann., J. K.: Assessment of aerosol–cloud–radiation correlations in satellite observations, climate models and reanalysis, *Climate Dynamics*, 52, 4371–34 392, 2018.

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.

Brient, F., Roehrig, R., and Voldoire, A.: Evaluating marine stratocumulus clouds in the CNRM-CM6-1 model using short-term hindcasts, *J. Adv. Model. Earth Syst.*, 11, 127–148, 2019.

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.

390 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., Coggon, M. M., Thompson, G., Rahn, D., Seinfeld, J. H., Sorooshian, A., and Lebo, Z. J.: Marine Boundary Layer Clouds

395 Associated with Coastally Trapped Disturbances: Observations and Model Simulations, *J. Atmos. Sci.*, 76, 2963–2993, 2019a.

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.*, 100, 631–651, 2019b.

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.

400 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.

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.

405 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. 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.,

410 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 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

415 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 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

420 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. 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 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. Soc.*, 81, 719–743, 2000.

Painemal, D., Minnis, P., and Nordeen, M.: Aerosol variability, synoptic-scale processes, and their link to the cloud microphysics over the northeast Pacific during MAGIC, *J. Geophys. Res. Atmos.*, 120, 5122–5139, 2015.

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.*, 136, 644–663, 2008.

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

Platnick, S. and Twomey, S.: Determining the Susceptibility of Cloud Albedo to Changes in Droplet Concentration with the Advanced Very High Resolution Radiometer, *J. Appl. Meteor.*, 33, 334–347, 1994.

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 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, 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 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.

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.

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.

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.

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 burning aerosols during the 2007 San Diego wildfires, *Environ. Sci. Technol.*, 47, 7633–7643, 2013.