



INSTITUTE FOR DEFENSE ANALYSES

Statistical Cluster Analysis of Global Aerosol Optical Depth for Simplified Atmospheric Modeling

N. T. Plymale

J. E. Szekely

A. H. Rubinstein

Distribution Statement A.

Approved for public release; distribution is unlimited.

Content of this paper is UNCLASSIFIED.

Final

April 2020

IDA Document NS D-13184

Log: H 2020-000163

Copy

INSTITUTE FOR DEFENSE
ANALYSES 4850 Mark Center Drive
Alexandria, Virginia 22311-1882



The Institute for Defense Analyses is a nonprofit corporation that operates three Federally Funded Research and Development Centers. Its mission is to answer the most challenging U.S. security and science policy questions with objective analysis, leveraging extraordinary scientific, technical, and analytic expertise.

About This Publication

This work was conducted by the IDA Systems and Analyses Center under contract HQ0034-14-D-0001, Project DC-1-3994, "Collaborative Nuclear Targeting Analyses (CNTA)," for the DC / DTRA / Defense Threat Reduction Agency. The views, opinions, and findings should not be construed as representing the official position of either the Department of Defense or the sponsoring organization.

Acknowledgements

The authors would like to thank IDA committee, Dr. Steve Warner (chair), Dr. Geoffrey M. Koretsky, Dr. James M. Gilmore, Dr. Jennifer L. Bewley, Dr. Nathan Platt, and Dr. Peter AM Dolph for providing technical review of this effort.

We would also like to thank Dr. Joanna Ingraham for providing helpful feedback over the course of this work.

For More Information

Peter AM Dolph, Project Leader
pdolph@ida.org, 703-845-6819

Steve Warner, Director, SED
swarner@ida.org, 703-845-2096

Copyright Notice

© 2020 Institute for Defense Analyses
4850 Mark Center Drive, Alexandria, Virginia 22311-1882 • (703) 845-2000.

This material may be reproduced by or for the U.S. Government pursuant to the copyright license under the clause at DFARS 252.227-7013 (a)(16) [Jun 2013].

Rigorous Analysis | Trusted Expertise | Service to the Nation

UNCLASSIFIED

INSTITUTE FOR DEFENSE ANALYSES

IDA Document NS D-13184

Statistical Cluster Analysis of Global Aerosol Optical Depth for Simplified Atmospheric Modeling

N. T. Plymale

J. E. Szekely

A. H. Rubinstein

UNCLASSIFIED

UNCLASSIFIED

UNCLASSIFIED

Executive Summary

Atmospheric aerosols originating from natural and anthropogenic sources have important implications for modeling atmospheric phenomena, but aerosols conditions can change significantly and rapidly because of their dependence on local geography and atmospheric conditions. Herein, we applied a computational k-means clustering algorithm to a global set of data obtained from the Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2) to yield a set of 25 clusters that discriminate based on land type, elevation, and atmospheric conditions to predict statistical aerosol optical depth (AOD) information. We analyzed annual 2016 MERRA-2 data in addition to data from four meteorological seasons for the five year period from 2012 through 2016 to arrive at five separate sets of 25 clusters. We make the clustered AOD information available with decision trees, qualitative cluster descriptions, and color-coded cluster maps to assist in identifying which cluster to use in retrieving AOD information. The results of this analysis have applications in atmospheric modeling where knowledge of approximate or typical aerosol conditions are needed in lookup table form without requiring access to large atmospheric databases or computationally intensive aerosol models.

(This page is intentionally blank.)

Contents

1. Article (formatted as submission to journal)	1-1
Appendix A. Supplemental Material (formatted as submission to journal).....	A-1

(This page is intentionally blank.)

1. Statistical Cluster Analysis of Global Aerosol Optical Depth for Simplified Atmospheric Modeling

(This page is intentionally blank.)

Statistical Cluster Analysis of Global Aerosol Optical Depth for Simplified Atmospheric Modeling

Noah T. Plymale,* Joshua E. Szekely, and Anna H. Rubinstein

System Evaluation Division, Institute for Defense Analyses, Alexandria, VA 22311, United States

Abstract

Atmospheric aerosols originating from natural and anthropogenic sources have important implications for modeling atmospheric phenomena, but aerosols conditions can change significantly and rapidly because of their dependence on local geography and atmospheric conditions. Herein, we applied a computational k-means clustering algorithm to a global set of data obtained from the Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2) to yield a set of 25 clusters that discriminate based on land type, elevation, and atmospheric conditions to predict statistical aerosol optical depth (AOD) information. We analyzed annual 2016 MERRA-2 data in addition to data from four meteorological seasons for the five year period from 2012 through 2016 to arrive at five separate sets of 25 clusters. We make the clustered AOD information available with decision trees, qualitative cluster descriptions, and color-coded cluster maps to assist in identifying which cluster to use in retrieving AOD information. The results of this analysis have applications in atmospheric modeling where knowledge of approximate or typical aerosol conditions are needed in lookup table form without requiring access to large atmospheric databases or computationally intensive aerosol models.

*Author to whom correspondence should be addressed, email: nplymale@ida.org

1. Introduction

Atmospheric aerosols, which are suspensions of solid or liquid particles in air, participate in Earth's surface energy budget by scattering or absorbing radiation as a function of their size and composition (Rizza et al. 2019). Additionally, aerosols can promote or inhibit cloud formation and can contribute to the mechanisms that result in poor breathable air quality (Sun and Ariya 2006; Prospero 1999). The properties of aerosols make them important to consider in developing models for atmospheric phenomena, such as weather models that can predict hazardous conditions (Xian et al. 2019) or atmospheric transmission models important to applications in remote sensing (Berk et al. 1998).

Aerosol concentrations and compositions are influenced by proximity to geographic features (such as deserts or bodies of water), atmospheric conditions (including amount of precipitation and wind speed), natural events (such as volcanic activity and forest fires), and human activity (including burning fossil fuels) (Viana et al. 2014). The dependence of aerosols on these factors causes their concentrations in the atmosphere to vary significantly and, often, rapidly along with their contributions to the atmospheric phenomena they influence (Satheesh et al. 2017). Atmospheric models that require knowledge of aerosol conditions are challenged by the difficulty in predicting aerosol concentrations accurately and quickly.

The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2) is a reanalysis of atmospheric conditions provided by the National Aeronautics and Space Administration (NASA) Global Modeling and Assimilation Office (GMAO) encompassing 1 January, 1980 to the present (Gelaro et al. 2017). MERRA-2 includes a substantial database of atmospheric and weather information, including daily average aerosol optical depth (AOD) data, on a grid with 0.5° latitude $\times$ 0.625° longitude resolution (Bosilovich et al. 2016). MERRA-2 uses

the Goddard Global Ozone Chemistry Aerosol Radiation and Transport (GOCART) model (Chin et al. 2002; Colarco et al. 2010), which accounts for organic carbon, black carbon, dust, sea salt, and sulfate aerosols, to describe atmospheric aerosol conditions by providing AOD data for each of these five categories in addition to a total AOD that sums these components. AOD data included in the MERRA-2 model are assimilated from satellite-based sensors, including the Moderate Resolution Imaging Spectroradiometer (MODIS), Multi-angle Imaging Spectroradiometer (MISR), and Advanced Very High-Resolution Radiometer (AVHRR), and from the ground-based Aerosol Robotics Network (AERONET) (Holben et al. 1998; Randles et al. 2017).

The total AOD, which quantifies the optical properties for the total aerosols present in the atmosphere, is defined as the natural logarithm of the ratio of incident to transmitted light through a vertical column of the atmosphere (Wenny et al. 1998). The AOD is wavelength dependent but is provided by MERRA-2 at 550 nm, which corresponds to the maximum sensitivity of the human eye. All references to AOD in this work are specific to the AOD at 550 nm (abbreviated AOD₅₅₀) unless otherwise specified. The AOD is a common input that describes aerosol conditions in atmospheric transmission models, including the Moderate-Resolution Atmospheric Transmission (MODTRAN) model (Berk et al, 2011), which is developed and maintained by Spectral Sciences, Inc. (Burlington, MA) and the Air Force Research Laboratory (Dayton, OH). MODTRAN relies on accurate AOD inputs to compute transmission spectra for atmospheric cross sections, and these inputs commonly come from sources such as MERRA-2 or computationally intensive models that can calculate AOD information as a function of atmospheric aerosol conditions.

We describe herein a process of clustering global geographic and atmospheric data from MERRA-2 using a computational machine learning technique to find geographic regions having distinct atmospheric conditions, each having accessible statistical AOD information in lookup

table form. In clustering the global data, we considered regions of the ocean separate from all other land types, which consist of lake, land, and ice-covered land. We consider annual and seasonal averages of the input dataset and provide 3 “ocean” clusters and 22 “land” clusters for a total of 25 global clusters. The “land” clusters are further categorized as “permafrost,” “inland,” “lake,” or “coast.” The results are presented in a series of tables that contain the cluster information and allow a user to identify an appropriate cluster and its associated AOD information. The result of the analysis is a detailed set of AOD lookup tables organized into clusters that do not require access to a database or computer model, providing data portability and accessibility.

2. Methods

The k-means algorithm is an unsupervised computational machine learning technique that groups complex datasets into a defined number of clusters, k , that have similarity across the input observation vectors (MacQueen 1967). The clusters consist of a set of centroids that closely approximate the mean value for the observations associated with a cluster. The k-means algorithm operates by accepting a set of N observation vectors and the desired number of clusters, k , and returns k clusters that each have N centroids. The algorithm begins with an initialized set of random centroids and iteratively adjusts the centroid positions to minimize the distortion, which is defined as the sum of the squared distance between each observation and its nearest centroid.

The result from the k-means algorithm is a code book that maps clusters to their respective centroid values and can be used to assign each set of original observations to the most appropriate cluster. The code book that is ultimately generated is sensitive to the random positions of the initial cluster centroids; local minima in the distortion can result in different code books each time the k-means algorithm is executed on the same set of observations. Moreover, the k-means algorithm accepts k as an input parameter, requiring the user to identify an optimal value for k prior to

beginning. While numerous quantitative methods exist for optimizing k , none of them unequivocally yields the most suitable value.

Section A of the supporting materials details the methodology and logic we used in optimizing the number of clusters for the MERRA-2 input variables (Satopää et al. 2011; Tolsa 2000; Zhao et al. 2008). An ideal value for k maximizes the information represented by the clusters while conserving computational resources. We compared four methods for optimizing k and considered values only where multiple methods were in agreement. Additionally, we restricted the range of k that we considered in order to group “ocean” (fraction of ocean = 1) and “land” (*i.e.*, not exclusively ocean, fraction of ocean < 1) data subsets into clusters according to the projected relative interest in identifying an AOD₅₅₀ for locations in these data subsets. We assumed lower specific interest in modeling the AOD₅₅₀ over the open ocean and consequently limited the number of ocean clusters, while accommodating a larger number of clusters over regions containing land, ice-covered land, or lake. Ultimately, we divided “ocean” regions into 3 clusters and “land” regions into 22 clusters for a total of 25 global clusters.

The k-means algorithm is well suited to finding correlations between atmospheric aerosols and natural factors that influence them, such as geography and ambient weather, because its distortion minimization algorithm is intended to identify correlations among equally weighted sets of observations (Govender and Sivakumar 2019; Liu and George 2005; Wegner et al. 2012). We accounted for geography using land type and elevation variables that we derived from the MERRA-2 database. We accounted for ambient weather in each MERRA-2 grid cell by including relative humidity, temperature, wind speed, and fractions of showers and cloud cover as inputs to the clustering algorithm. Detailed information on the variables derived from MERRA-2 is provided in the following Data Sources section.

We extracted MERRA-2 data for the years 2012 through 2016 and average the data either annually or according to the meteorological season desired. We stored the data as linear observation vectors and created separate datasets to represent “ocean” (fraction of ocean = 1) and “land” (fraction of ocean < 1) subsets of the data. We used the “whiten” function available in the SciPy package (v. 1.3.2) implemented in Python (v. 3.7) to give the data unit variance by dividing each data point by the standard deviation of the data in the corresponding vector. We then used the “kmeans” function of the SciPy package to generate a code book containing the cluster centroids using $k = 3$ clusters for the “ocean” subset and using $k = 22$ clusters for the “land” subset of the data. Using SciPy’s “vq” function, we mapped each set of observations to a centroid using the appropriate code book, removed the “whiten” processing on the code book by multiplying each entry by the standard deviation of the data in each observation vector, and made the results available in sections B, C, D, E, and F of the supplemental materials provided with this work.

3. Data Sources

Table 1 summarizes the variables considered in the k-means clustering analysis and gives the corresponding MERRA-2 name(s) (Bosilovich et al. 2016). Variables that were calculated from raw MERRA-2 data are the elevation, relative humidity, 10-m wind speed, fraction of showers, fraction of black carbon, fraction of dust, fraction of organic carbon, fraction of sea salt, and fraction of sulfate. The fraction of each of the five aerosol types were computed by finding the ratio of each aerosol component AOD to the total AOD₅₅₀ (TOTEXTTAU in MERRA-2).

The elevation (h) was calculated from the surface geopotential height (SGH) provided in MERRA-2 using equation 1 (Leroy 1997).

$$h = \frac{(6.356 \times 10^6)SGH/9.8}{6.356 \times 10^6 - (SGH/9.8)} \quad (1)$$

Here, we approximate Earth as a sphere with a radius of 6.356×10^6 m and assume the acceleration due to gravity is constant at 9.8 m s^{-2} . The input SGH is in $\text{m}^2 \text{ s}^{-2}$.

The relative humidity was calculated from the surface pressure (p), the 2-m specific humidity (q), and 2-m air temperature (T) using the following method. First, the measured mass mixing ratio (w) of water vapor to dry air was calculated using equation 2 (Wallace and Hobbs 2006).

$$w = \frac{q}{1-q} \quad (2)$$

Then, the saturation vapor pressure (e_s) was calculated above liquid water (where $T \geq 0$ °C) and above ice (where $T < 0$ °C) using Arden Buck equations 3 and 4, respectively (Buck 1981).

$$e_{s,T \geq 0 \text{ } ^\circ\text{C}} = 611.21 \exp \left[\left(18.678 - \frac{T}{234.5} \right) \left(\frac{T}{257.14 + T} \right) \right] \quad (3)$$

$$e_{s,T < 0 \text{ } ^\circ\text{C}} = 611.15 \exp \left[\left(23.036 - \frac{T}{333.7} \right) \left(\frac{T}{279.82 + T} \right) \right] \quad (4)$$

Here T is in °C. The saturation vapor pressure was converted to the saturation mass mixing ratio of water to dry air at equilibrium (w_s) using equation 5 (Wallace and Hobbs 2006) assuming that surface pressure (p) is approximated by the 2-m surface pressure variable in MERRA-2.

$$w_s = 0.622 \frac{e_s}{p - e_s} \quad (5)$$

The relative humidity (RH) was then calculated using equation 6 (Wallace and Hobbs 2006).

$$RH = \frac{w}{w_s} \times 100\% \quad (6)$$

The 10-m wind speed (W_T) was calculated from the 10-m eastward and northward wind speeds (W_E and W_N , respectively) using equation 7.

$$W_T = \sqrt{W_E^2 + W_N^2} \quad (7)$$

The areal fraction of showers (S_F) was calculated for each MERRA-2 grid cell by finding the maximum of the variables representing the fraction of anvil showers (S_A), the fraction of convective showers (S_C), and the fraction of non-anvil large-scale showers (S_{LS}). Equation 8 depicts the methodology used to find S_F .

$$S_F = \text{MAX}(S_A, S_C, S_{LS}) \quad (8)$$

In this analysis, we considered MERRA-2 data for all of 2016 in addition to four meteorological seasons from 2012 through 2016: (1) Dec/Jan/Feb, (2) Mar/Apr/May, (3) Jun/Jul/Aug, and (4) Sep/Oct/Nov. The data corresponding to these time periods were extracted from the MERRA-2 database and averaged over all days in the applicable date range before clustering the data. To provide an example of the data we extracted from MERRA-2, Figure 1 visually depicts the 50th percentile of the global total AOD₅₅₀ for the years 2012 through 2016.

In performing a limited scope verification and validation of the results from this analysis, we obtained AOD data from the AERONET system of sun photometers (Giles et al. 2019) and extracted the data from 996 candidate sites relevant to compare with the AOD₅₅₀ that we obtained from MERRA-2. The AERONET sites we analyzed had AOD data collected daily at 555, 551, and/or 532 nm, which were the three closest standard wavelengths to 550 nm for which AOD data is available from AERONET. We used daily average Level 2.0 AERONET Version 2 data

(accessed 21 October 2017 from https://aeronet.gsfc.nasa.gov/new_web/aerosols.html), which is quality assured, cloud cleared, and manually inspected, in this comparison.

4. Results

The result of performing the k-means algorithm on the MERRA-2 variables summarized in Table 1 is a code book that contains the centroids for each of the clusters identified by the algorithm. The code book is then used to retroactively assign the observations at each location in the original global dataset to the cluster that best represents the data. The result is an array of integers that assign each of the global MERRA-2 grid cells to a cluster.

Figure 2 shows the geographic location of the 3 clusters belonging to the “ocean” subset of the processed annual 2016 MERRA-2 data. In the geographic cluster plots, the clusters are organized by their centroid for temperature, where blue represents cooler temperatures and yellow represents warmer temperatures. In the clustered “ocean” data, cool regions of the ocean near the poles tended to cluster together with the exception of an area off the western coast of Africa, while warm regions closer to the equator also clustered together. The green cluster in Figure 2, which has an average daily temperature between the blue and yellow clusters, is discontinuous and contains cool regions in the mid-latitudes and equatorial regions of the Pacific and Indian Oceans. The discontinuity in the green and, to a lesser extent, blue cluster regions is a result of the k-means algorithm clustering regions of the ocean where variables other than the temperature show correlation, including regions with similar humidity, wind speed, precipitation, and cloud cover.

Figure 3 shows the 22 “land” clusters identified by the k-means algorithm for the global 2016 averaged MERRA-2 dataset. These clusters are also organized by their temperature centroid, where blue clusters are coolest and yellow clusters are warmest. The k-means algorithm was able to cluster a number of regions around the globe that intuitively belong to the same group. For

example, the pale green tropical rainforest regions of equatorial South America and Africa are grouped together in a cluster with warm temperatures, high cloud coverage, and very low wind speed. The algorithm also identified a cluster for eastern China, Korea, and much of India that coincides with locations having high AOD₅₅₀, indicated by dark regions in Figure 1. Moreover, the algorithm separated coastal regions, areas with inland lakes, and regions of permafrost from inland areas without these features.

Sections B, C, D, E, and F of the supplemental materials contain the code books and global cluster assignments for MERRA-2 data averaged over the specified periods of time. Table 2 details which supplemental section corresponds to which time period of MERRA-2 data used in the clustering process. Each of these supplemental sections presents the results in the same format. The code books contain the centroid data for each cluster, but identifying a cluster of interest directly from a code book is onerous because of the level of specificity contained in the cluster centroids.

To assist users in identifying an appropriate AOD₅₅₀ for their application, we have provided several tools that help identify a suitable cluster. First, each of the MERRA-2 cluster code books has a unique decision tree that assists users in narrowing the set of clusters under consideration according to specific criteria, such as land type and elevation. Second, we provide a qualitative description that categorizes the most intuitive cluster variables and describes associated characteristic regions for each cluster. Lastly, we provide global color-coded maps for each of the individual clusters to assist in visual identification of the region of interest.

The first table in sections B, C, D, E, and F of the supplemental materials (*e.g.*, Table B1) is the code book, providing the number associated with each cluster in addition to the cluster centroid values. We used the first four variables (fraction of lake, land, land-ice, and ocean) to categorize

the cluster according to land type, where the predominant land type is color-coded in each table. The first three clusters always correspond to the ocean subset of the data, which is clustered separately from the land subset. The centroid for the total AOD₅₅₀ is given in the last column of each table.

In all cases, the centroid for fraction of showers is greater than the centroid for fraction of cloud coverage. This result is not intuitive, assuming cloud coverage is required to produce showers. We observed this phenomenon in the raw MERRA-2 dataset for individual grid cells as well. Consequently, we recommend that these two variables be considered relative to the dataset as a whole and encourage use of the cluster qualitative descriptors to understand these variables and relate them to measurements of precipitation and cloud cover.

The next two tables (*e.g.*, Tables B2 and B3) provide the 10th, 25th, 50th, 75th, and 90th percentiles of the global ocean and land data, respectively, for each of the clustered variables. We used these percentiles to assign qualitative descriptors that categorize the centroids of the following five variables: humidity, temperature, wind speed, fraction of showers, and fraction of cloud coverage. These descriptors effectively bin each of the cluster centroids for these five categories into a percentile range to assist in describing the clusters relative to each other. Qualitative descriptions of these variables further provide a means of identifying a cluster with a set of desired properties. Table 3 maps each qualitative descriptor to a percentile range.

The descriptors defined in Table 3 are provided for the five relevant variables for each cluster in the fourth table in sections B, C, D, E, and F of the supplemental materials (*e.g.*, Table B4) next to a summary of the characteristic regions encompassed by each cluster. The statistical information for the AOD₅₅₀ in each cluster is available in this table. The AOD centroid and AOD average are typically very similar values; however, the centroid is calculated through an iterative process by

the k-means algorithm and sometimes deviates slightly from the average. We provide the standard deviation to give a sense of the spread of the AOD₅₅₀ within a cluster. We additionally provide the 10th, 50th, and 90th percentile AOD₅₅₀ to represent low aerosol (high visibility), typical aerosol (typical visibility), and high aerosol (low visibility) conditions, respectively.

The first figure presented in sections B, C, D, E, and F of the supplemental materials (*e.g.*, Figure B1) consists of a decision tree that graphically breaks apart the clusters according to their unique cluster centroids. The seven variables considered in the decision trees are land type (an interpretation of fraction of lake, land, land-ice, and ocean), elevation, humidity, temperature, wind speed, fraction of showers, and fraction of cloud coverage. We used quantitative metrics to divide clusters in the decision trees using these variables; we grouped clusters that had similar centroids for a particular variable. We generally arrived at these quantitative delineations by calculating the midpoint between adjacent cluster centroids. Distinguishing between different clusters is sensitive to the order in which the quantitative delineations are made between clusters (*e.g.*, whether a cluster is separated by its humidity before temperature or *vice versa*), particularly for similar inland clusters. We recommend using the decision trees sparingly and in conjunction with the other available representations of the cluster data to arrive at the best cluster for the intended application.

The list of characteristic regions provided for each cluster in the fourth table of sections B, C, D, E, and F of the supplemental materials is not exhaustive. In Figures 2 through 7 of these sections, (*e.g.*, Figures B2 through B7), we provide color-coded maps of the individual clusters for the “ocean” and “land” regions of the globe. These figures are particularly useful for identifying the geographic extent of a region and can help with developing an intuitive picture of what a region represents on a global scale. Using the graphical representation of the clusters in conjunction with

the decision tree and the qualitative description of the cluster centroids provides the user with a comprehensive means of identifying an appropriate cluster for a particular application.

We recommend the following steps for using sections B, C, D, E, and F of the supplemental materials to find the most appropriate cluster for the intended application. We use the letter 'X' as a stand-in for B, C, D, E, and F to illustrate which supplemental figure or table to refer to.

1. Use Table 2 to decide which section of the supplemental materials to refer to according to the time of year appropriate for the intended application.
2. Determine which cluster best represents the conditions and region of interest.
 - a. Use the decision tree in Figure X1 to identify which clusters correspond to the land type of interest. The decision tree can be consulted to further narrow the subset of clusters under consideration, but we caution against using the decision tree as the only method for cluster identification.
 - b. Consult Figures X2 through X7 to identify candidate clusters according to their graphical representations on a map.
 - c. Consult the qualitative description of the candidate clusters in Table X4 and use these descriptions to narrow the clusters under consideration by their characteristics.
 - d. If a specific cluster is not yet identified, refer back to the decision tree in Figure X1, the characteristic regions in Table X4, and the graphical cluster representations in Figures X2 through X7 to further narrow the clusters under consideration.

3. Find the cluster number in Table X4 and record the last three columns, which are the 10th, 50th, and 90th percentile AOD. These represent low, typical, and high aerosol conditions, respectively, for that cluster.

5. Discussion

The AOD is an important but difficult to predict parameter in modeling atmospheric phenomena. The AOD₅₅₀ data tables and cluster maps in sections B, C, D, E, and F of the supplemental materials are intended to provide a fast means of estimating the AOD₅₅₀ without requiring access to a large database like MERRA-2 or requiring the time to run complex computer models to make estimates for AOD₅₅₀. However, the k-means algorithm, which is an unsupervised machine learning technique, used for identifying the clusters does not assess the accuracy of the results and cannot determine whether using AOD₅₅₀ statistics given for each cluster is appropriate.

In a limited assessment of the accuracy of the AOD₅₅₀ data available for each cluster, we compared experimental AOD data from AERONET sites with the statistical AOD₅₅₀ data reported herein for the cluster in which each site is located. This assessment is limited because of 996 candidate sites only 30 had AOD data available for any of the desired wavelengths, and many sites were located in the same cluster; a total of 7 unique clusters contained AERONET sites for each time of year. We compared the average AERONET AOD at 555, 551, and/or 532 nm for all days contained within the respective time of year with the average AOD₅₅₀ for the cluster in which each AERONET site is located. We also compared the 10th, 50th, and 90th percentile AOD at each candidate AERONET site with the percentiles for the corresponding cluster. We found the magnitude of the difference (absolute value) between the AOD statistic computed from AERONET data and the respective cluster AOD₅₅₀ statistic and divided by the AERONET statistic to compute the fractional difference between the actual (AERONET) and predicted (cluster) AOD

result. Table 4 presents the average of these fractional differences across all statistical metrics compared. Section G of the supporting materials contains plots that show all of the AERONET AOD data versus the corresponding cluster AOD₅₅₀ data considered in this analysis.

The results of this limited assessment of the accuracy of the clustered AOD data shown in Table 4 demonstrate that the statistical AOD₅₅₀ data provided for the clusters can vary substantially from experimentally measured AERONET AOD data. However, these results are far from comprehensive, as we only compared a small subset of cluster data with AERONET data. Moreover, much of the AERONET AOD data were collected during years prior to the 2012 through 2016 years from which the clustered data were extracted. To give further context, Table 4 also presents the average spread of the AERONET data, which is a representation of the standard deviation of the AOD as a fraction of the average AOD, for each respective time of year. The spread in the AERONET AOD data can be significant and, with the exception of the Dec/Jan/Feb data, is greater than the average error observed from the cluster AOD₅₅₀ prediction.

The results of the comparison in Table 4 are far from definitive, but they do emphasize that, given the unpredictable nature of atmospheric aerosols, atmospheric models that account for aerosols should consider the statistical spread of the AOD information available. Perhaps the best solution is to anticipate a realistic range of conditions to model that consider extreme and typical conditions together. We included the 10th, 50th, and 90th percentile AOD₅₅₀ data for each cluster in sections B, C, D, E, and F of the supplemental materials to provide a way of accounting for the spread in AOD. A more comprehensive validation of the clustered AOD₅₅₀ data is warranted in the future using experimental AOD data only from 2012 through 2016 and from multiple other data sources (*e.g.*, satellites).

An additional consideration in interpreting the results is that each of the observation vectors derived from MERRA-2 included in the k-means clustering process was weighted equally by the algorithm. However, the natural correlation between some variables could be stronger than for others, and this could be accounted for in principle by applying weights to variables that are known to have greater influence on the resulting AOD. Where such relationships are known *a priori*, weights could be readily applied to the variables in the clustering process, but an analysis of this nature is outside the scope of this initial work on identifying unique global regions with distinct atmospheric and aerosol conditions.

6. Conclusions

The k-means clustering algorithm can help identify relationships between variables where no obvious correlation is directly observable, but the algorithm cannot provide an assessment of its performance in retrospect. The results from clustering global variables derived from MERRA-2 reported herein represent approximations of statistical AOD values that are representative of geographic regions with distinct atmospheric properties. The results allow for low (10th percentile), typical (50th percentile), and high (90th percentile) AOD values to be identified easily from lookup tables for applications that can tolerate lower fidelity AOD information in exchange for simplicity and speed. Moreover, the code books provided for each cluster contain centroid information for all variables included in the clustering algorithm and offer their own insight into the geographic and atmospheric properties of various global regions. In an assessment of the accuracy of the clustered AOD results relative to AERONET data, we determined that there can be a substantial difference between predicted and measured AOD, but the results emphasized that atmospheric models should account for the high degree of variability in aerosol conditions. The

statistical results provided for each of the clusters allow users of the lookup tables to account for AOD variability as required for their application.

Acknowledgements

This work was supported by the Defense Threat Reduction Agency (DTRA) under contract HQ0034-19-D-0001 and the central research program of the Institute for Defense Analyses. We thank Drs. Joanna Ingraham, Jennifer Bewley, Peter Dolph, Mike Gilmore, Geoffrey Koretsky, Nathan Platt, and Steve Warner for providing helpful feedback over the course of this work.

Data Availability Statement: Data from MERRA-2 used in this work are available from the NASA website at <https://disc.gsfc.nasa.gov/datasets> and described in the MERRA-2 file specification (Bosilovich et al. 2016). Daily average Level 2.0 AERONET Version 2 data used in this work are available from the AERONET website at <https://aeronet.gsfc.nasa.gov/> (Giles et al. 2019).

References

- Berk, A., G. P. Anderson, P. K. Acharya, and E. P. Shettle, 2011: *MODTRAN 5.2.1 User's Manual*. Spectral Sciences, 99 pp.
- Berk, A., L. S. Bernstein, G. P. Anderson, P. K. Acharya, D. C. Robertson, J. H. Chetwynd, and S. M. Alder-Golden, 1998: MODTRAN cloud and multiple scattering upgrades with application to AVIRIS. *Remote Sens. Environ.*, **63**, 3, 367–375, [https://doi.org/10.1016/S0034-4257\(98\)00045-5](https://doi.org/10.1016/S0034-4257(98)00045-5)
- Bosilovich, M. G., R. Lucchesi, and M. Suarez, 2016: MERRA-2: File specification. GMAO office note no. 9 (Version 1.1), 73 pp, available from <https://gmao.gsfc.nasa.gov/pubs/docs/Bosilovich785.pdf>

Buck, A. L., 1981: New equations for computing vapor pressure and enhancement factor. *J.*

Appl. Meteor., **20**, 1527–1532,

[https://doi.org/10.1175/1520-0450\(1981\)020<1527:NEFCVP>2.0.CO;2](https://doi.org/10.1175/1520-0450(1981)020<1527:NEFCVP>2.0.CO;2)

Chin, M., P. Ginoux, S. Kinne, O. Torres, B. N. Holben, B. N. Duncan, R. V. Martin, J. A.

Logan, H. Higurashi, and T. Nakajima, 2002: Tropospheric aerosol optical thickness

from the GOCART model and comparisons with satellite and sun photometer

measurements. *J. Atmos. Sci.*, **59**, 461–483,

[https://doi.org/10.1175/1520-0469\(2002\)059<0461:TAOTFT>2.0.CO;2](https://doi.org/10.1175/1520-0469(2002)059<0461:TAOTFT>2.0.CO;2)

Colarco, P. R., A. da Silva, M. Chin, and T. Diehl, 2010: Online simulations of global aerosol

distributions in the NASA GEOS-4 model and comparisons to satellite and ground-based

aerosol optical depth. *J. Geophys. Res.: Atmos.*, **115**, D14207,

<https://doi.org/10.1029/2009JD012820>

Gelaro, R., W. McCarty, M. J. Suárez, R. Todling, A. Molod, L. Takacs, C. A. Randles, A.

Darmenov, M. G. Bosilovich, R. Reichle et al., 2017: The modern-era retrospective

analysis for research and applications, version 2 (MERRA-2). *J. Climate*, **30**, 5419–5454,

<https://doi.org/10.1175/JCLI-D-16-0758.1>

Giles, D. M., A. Sinyuk, M. G. Sorokin, J. S. Schafer, A. Smirnov, I. Slutsker, T. F. Eck, B. N.

Holben, J. R. Lewis, J. R. Campbell et al., 2019: Advancements in the aerosol robotic

network (AERONET) version 3 database – automated near-real-time quality control

algorithm with improved cloud screening for sun photometer aerosol optical depth

(AOD) measurements. *Atmos. Meas. Tech.*, **12**, 169–209,

<https://doi.org/10.5194/amt-12-169-2019>

- Govender, P. and V. Sivakumar, 2019: Application of k-means and hierarchical clustering techniques for analysis of air pollution: A review (1980–2019). *Atmos. Pollut. Res.*, **11**, 1, 40–56, <https://doi.org/10.1016/j.apr.2019.09.009>
- Holben, B.N.; T. F. Eck, I. Slutsker, D. Tanre, J. P. Buis, A. Setzer, E. Vermote, J. A. Reagan, J. Kaufman, T. Nakajima et al., AERONET–A federated instrument network and data archive for aerosol characterization. *Remote Sens. of Environ.* **1998**, 66, 1–16, [https://doi.org/10.1016/S0034-4257\(98\)00031-5](https://doi.org/10.1016/S0034-4257(98)00031-5)
- Leroy, S. S., 1997: Measurement of geopotential heights by GPS radio occultation. *J. Geophys. Res.: Atmos.*, **102**, D6, 6971–6986, <https://doi.org/10.1029/96JD03083>
- Liu, Z. and R. George, 2005: Mining Weather Data Using Fuzzy Cluster Analysis. *Fuzzy Modeling with Spatial Information for Geographic Problems*, F. E. Petry, V. B. Robinson, and M. A. Cobb, Eds., Springer Berlin Heidelberg, 105–119, https://doi.org/10.1007/3-540-26886-3_5
- MacQueen, J., 1967: Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*. L. M. Le Cam and J. Neyman, Eds., University of California Press, 281–297.
- Prospero, J. M., 1999: Long-term measurements of the transport of African mineral dust to the southeastern United States: Implications for regional air quality. *J. Geophys. Res.: Atmos.*, **104**, D13, 15917–15927, <https://doi.org/10.1029/1999JD900072>
- Randles, C. A., A. M. da Silva, V. Buchard, P. R. Colarco, A. Darmenov, R. Govindaraju, A. Smirnov, B. Holben, R. Ferrare, J. Hair et al., 2017: The MERRA-2 aerosol reanalysis,

- 1980 onward. Part I: System description and data assimilation evaluation. *J. Climate*, **30**, 6823–6850, <https://doi.org/10.1175/JCLI-D-16-0609.1>
- Rizza, U., E. Mancinelli, M. Morichetti, G. Passerini, and S. Virgilli, 2019: Aerosol optical depth of the main aerosol species over Italian cities based on the NASA/MERRA-2 model reanalysis. *Atmosphere*, **2019**, 10, 709–723, <https://doi.org/10.3390/atmos10110709>
- Satheesh S.K., S. S. Babu, B. Padmakumari, G. Pandithurai, V. K. Soni, 2017: Variability of atmospheric aerosols over India. *Observed climate variability and change over the Indian region*, M. Rajeevan, and S. Nayak, Eds., Springer Geology, 221–248, https://doi.org/10.1007/978-981-10-2531-0_13
- Satopää, V., J. Albrecht, D. Irwin, and B. Raghavan, 2011: Finding a ‘kneedle’ in a haystack: Detecting knee points in system behavior. *2011 31st International Conference on Distributed Computing Systems Workshops*, 166–171, <https://doi.org/10.1109/ICDCSW.2011.20>
- Sun, J., and P. R. Ariya, 2006: Atmospheric organic and bio-aerosols as cloud condensation nuclei (CCN): A review. *Atmos. Environ.*, **40**, 5, 795–820, <https://doi.org/10.1016/j.atmosenv.2005.05.052>
- Tolsa X., 2000: Principal values for the Cauchy integral and rectifiability. *Proceedings of the American Mathematical Society*, 128, 7, 2111–2119, <https://doi.org/10.1090/S0002-9939-00-05264-3>
- Viana, M., J. Pey, X. Querol, A. Alastuey, F. de Leeuw, and A. Lükewille, 2014: Natural sources of atmospheric aerosols influencing air quality across Europe. *Sci. Total Environ.*, **472**, 825–833, <https://doi.org/10.1016/j.scitotenv.2013.11.140>

Wallace, J. M. and P. V. Hobbs, 2006: Atmospheric Thermodynamics. *Atmospheric Science (Second Edition)*, J. M. Wallace and P. V. Hobbs, Eds., Academic Press, 63–111, <https://doi.org/10.1016/B978-0-12-732951-2.50008-9>

Wegner, T., T. Hussein, K. Hämeri, T. Vesala, M. Kulmala, and S. Weber, 2012: Properties of aerosol signature size distributions in the urban environment as derived by cluster analysis. *Atmos. Environ.*, **61**, 350–360, <https://doi.org/10.1016/j.atmosenv.2012.07.048>

Wenny, B. N., J. S. Schafer, J. J. DeLuisi, V. K. Saxena, W. F. Barnard, I. V. Petropavlovskik, and A. J. Vergamini, 1998: A study of regional aerosol radiative properties and effects on ultraviolet-B radiation. *J. Geophys. Res.: Atmos.*, **103**, D4, 17083–17097, <https://doi.org/10.1029/98JD01481>

Xian, P., J. S. Reid, E. J. Hyer, C. R. Sampson, J. I. Rubin, M. Ades, N. Asencio, S. Basart, A. Benedetti, P. S. Bhattacharjee et al., 2019: Current state of the global operational aerosol multi-model ensemble: An update from the International Cooperative for Aerosol Prediction (ICAP). *Quart. J. Roy. Meteor. Soc.*, **145**, Suppl. 1, 176–209, <https://doi.org/10.1002/qj.3497>

Zhao, Q., V. Hautamaki, and P. Fränti, 2008: Knee point detection in BIC for detecting the number of clusters. *Advanced Concepts for Intelligent Vision Systems*, J. Blanc-Talon, S. Bourennane, W. Philips, D. Popescu, and P. Scheunders, Eds., Springer-Verlag, 664–673.

Tables

Table 1. MERRA-2 variables used for k-means clustering.

Variable Name	Units	Description	MERRA-2 Name
Fraction of Lake	unitless	Fraction of the MERRA-2 grid cell that is categorized as lake.	FRLAKE
Fraction of Land	unitless	Fraction of the MERRA-2 grid cell that is categorized as land.	FRLAND
Fraction of Land-Ice	unitless	Fraction of the MERRA-2 grid cell that is categorized as ice-covered land. Commonly permafrost regions.	FRLANDICE
Fraction of Ocean	unitless	Fraction of the MERRA-2 grid cell that is categorized as ocean.	FROCEAN
Elevation	m	The elevation is derived from the surface geopotential height ($\text{m}^2 \text{s}^{-2}$) by assuming the Earth's radius as 6356 km and the acceleration due to gravity as 9.8 m s^{-2} .	PHIS
Relative Humidity	%	The relative humidity is derived from the surface pressure (Pa), 2-m specific humidity (unitless), and 2-m air temperature (K).	PS, QV2M, T2M
2-m Air Temperature	K	Air temperature 2 m above the surface.	T2M
10-m Wind Speed	m s^{-1}	Wind speed 10 m above the surface. The wind speed is derived from the 10-m eastward and northward wind speeds (m s^{-1}).	U10M, V10M
Fraction Showers	unitless	The areal fraction of showers is defined as the maximum value across the variables representing the areal fraction of anvil showers, convective showers, and non-anvil large-scale showers.	FRCAN, FRCCN, FRCLS
Fraction Cloud Coverage	unitless	Areal fraction of clouds.	CLDTOT
Fraction Black Carbon	unitless	Fraction of black carbon aerosols contributing to the AOD at 550 nm.	BCEXTTAU
Fraction Dust	unitless	Fraction of dust aerosols contributing to the AOD at 550 nm.	DUEXTTAU
Fraction Organic Carbon	unitless	Fraction of organic carbon aerosols contributing to the AOD at 550 nm.	OCEXTTAU
Fraction Sea Salt	unitless	Fraction of sea salt aerosols contributing to the AOD at 550 nm.	SSEXTTAU
Fraction Sulfate	unitless	Fraction of sulfate aerosols contributing to the AOD at 550 nm.	SUEXTTAU
Total AOD	unitless	Total aerosol optical depth at 550 nm.	TOTEXTTAU

Table 2. Time of year corresponding to data in sections B, C, and F of the supplemental materials.

Supplemental Materials Section	Time of Year
B	Annual (2016)
C	Dec/Jan/Feb (2012 – 2016)
D	Mar/Apr/May (2012 – 2016)
E	Jun/Jul/Aug (2012 – 2016)
F	Sep/Oct/Nov (2012 – 2016)

Table 3. Cluster qualitative descriptor definitions.

Cluster Centroid Descriptor	Percentile Range
Very low	0 – 10
Low	10 – 25
Medium low	25 – 50
Medium high	50 – 75
High	75 – 90
Very high	90 – 100

Table 4. Comparison of AERONET and clustered AOD data.

Time of Year	Average Fractional Error in Clustered AOD Data versus AERONET ^a	Average Spread of AERONET AOD Data ^b	Spread of Clustered AOD ₅₅₀ Data ^c
Annual	0.58	0.68	0.34
Dec/Jan/Feb	1.0	0.63	0.34
Mar/Apr/May	0.58	0.60	0.42
Jun/Jul/Aug	0.49	0.61	0.35
Sep/Oct/Nov	0.76	0.70	0.33

^aThe fractional error in clustered AOD data is the average across all available AERONET wavelengths of the absolute difference between the AERONET AOD statistic and the respective cluster AOD statistic divided by the AERONET AOD statistic. ^bThe average spread of AERONET AOD data is the average for all three wavelengths of the standard deviation for the daily AERONET AOD data available for the corresponding time of year divided by the average AOD for that time period. ^cThe spread of the clustered AOD data is the average of the standard deviation of the AOD divided by the average AOD within all clusters.

Figures

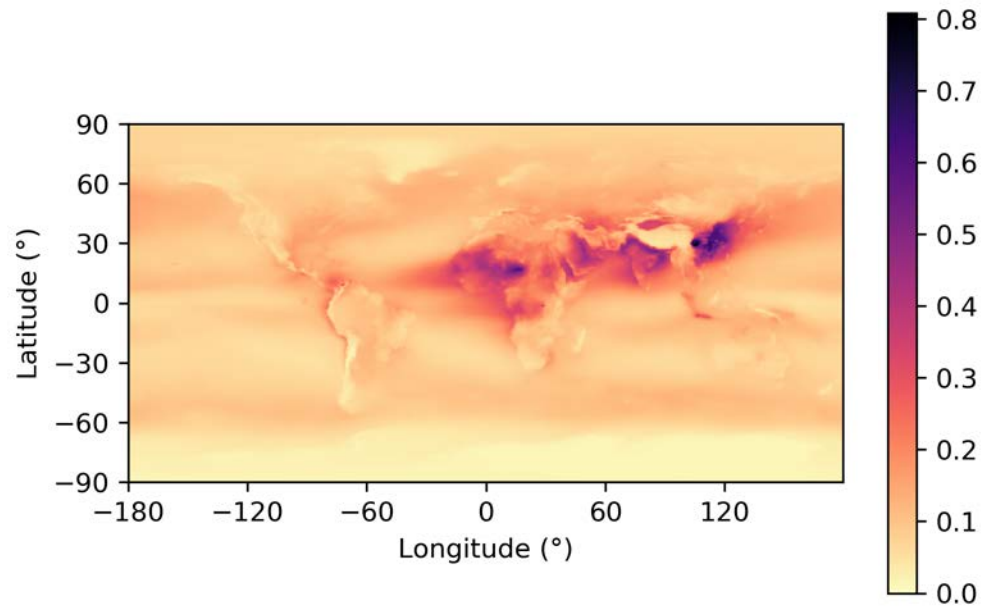


Figure 1. 50th percentile global AOD₅₅₀ for full 2012–2016 MERRA-2 dataset.

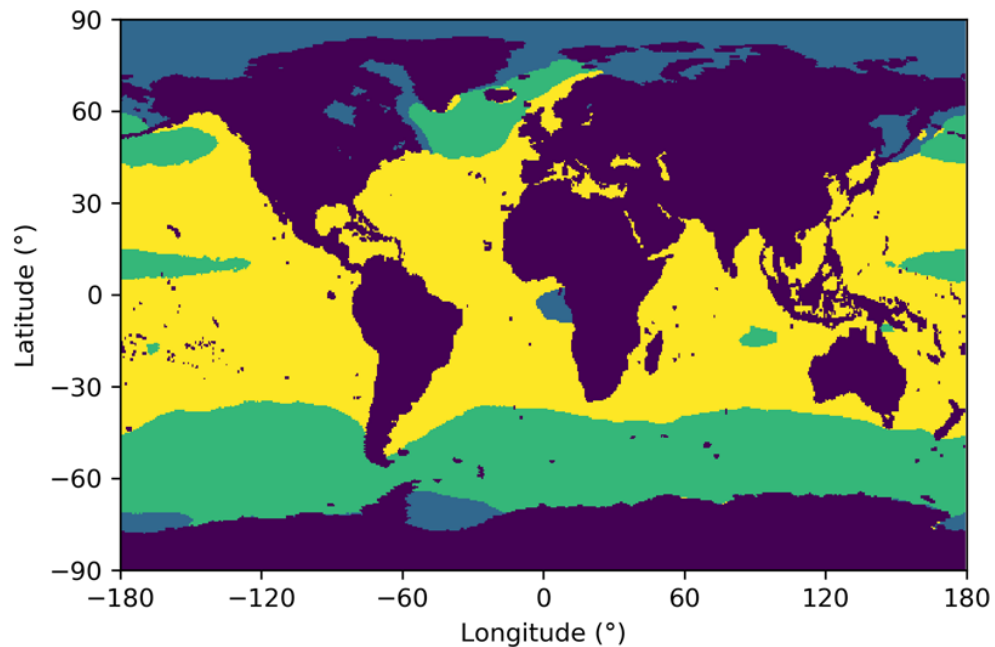


Figure 2. Geographic location of the clusters identified from the “ocean” subset of the full 2016 global MERRA-2 dataset. The clusters are arranged by average temperature, where the coolest region is blue and warmest region is yellow.

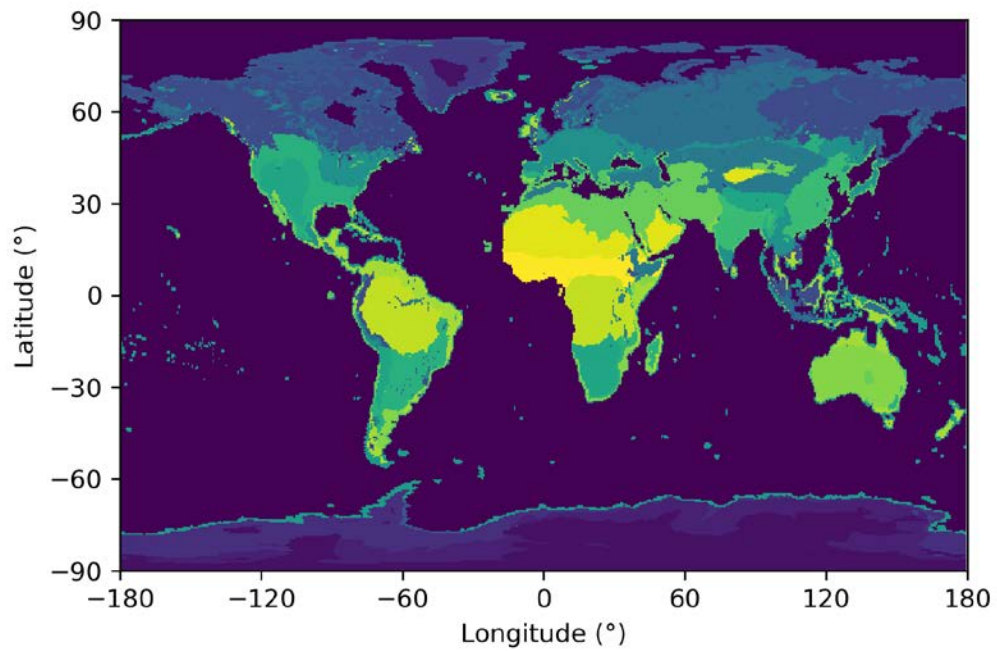


Figure 3. Geographic location of the clusters identified from the “land” subset of the full 2016 global MERRA-2 dataset. The clusters are arranged by average temperature, where the coolest regions are blue and warmest regions are yellow.

UNCLASSIFIED

(This page is intentionally blank.)

UNCLASSIFIED

Appendix A Supplemental Materials

(This page is intentionally blank.)

Supplemental Materials for

Statistical Cluster Analysis of Global Aerosol Optical Depth for Simplified Atmospheric Modeling

Noah T. Plymale,* Joshua E. Szekely, and Anna H. Rubinstein

System Evaluation Division, Institute for Defense Analyses, Alexandria, VA 22311, United States

Contents

A. Optimizing the Number of Clusters	S2
B. Full 2016 Global Clusters	S11
C. 2012 to 2016 December through February Global Clusters	S22
D. 2012 to 2016 March through May Global Clusters.....	S33
E. 2012 to 2016 June through August Global Clusters.....	S44
F. 2012 to 2016 September through November Global Clusters.....	S55
G. Comparison of Cluster AOD Predictions with AERONET AOD	S66
H. References.....	S72

*Author to whom correspondence should be addressed, email: nplymale@ida.org

A. Optimizing the Number of Clusters

Organizing large amounts of data into a small number of clusters (k) that represent the dataset as a whole generally involves a tradeoff between the computational time to organize the data into clusters and the number of clusters. Ideally, k should be minimized to reduce computation time while maximizing the information contained in the cluster centroids. A way of observing how well the cluster centroids fit the dataset is by calculating the squared distance between each data value in a cluster and its associated centroid, yielding a quantity known as distortion. Distortion decreases and approaches 0 as k approaches the number of data points in the original dataset. In principle, there is an ideal amount of distortion for k that is much smaller than the number of points in the dataset.

Multiple methods have been proposed to optimize k by analyzing the distortion as a function of k . These include the “Kneedle” method, a geometric angle-based method, and the Menger curvature method. Additionally, we quantified the Gaussian overlap between the centroids in each cluster as a function of k in order to identify values for k that yield unique cluster centroids.

Optimizing k Using the “Kneedle” Method

The k -means clustering algorithm is an unsupervised machine learning technique that requires the user to input the desired number of clusters, k . It is common practice to optimize k by identifying a “knee” in the distortion versus k curve, shown in Figure A1 for the 2016 land and ocean MERRA-2 datasets used in this work. The “knee” or “elbow” of a plot is generally defined as a point of maximum curvature. In some cases, the “knee” can be identified by visual inspection, but the lack of an obvious “knee” in this case requires a more quantitative and rigorous analysis.

The “Kneedle” method (Satopää et al. 2011) for identifying the “knee” of a curve quantifies the distance between each point in the dataset and the closest point on a line connecting the two extreme points on the curve. The result is depicted in Figure A1. The distortion versus k curves presented are averaged for 25 iterations of the k -means algorithm performed on the 2016 dataset. The point on the curve that falls farthest from the indicated line is $k = 14$ for the land dataset and $k = 15$ for the ocean dataset. Analysis of the data for fewer iterations of the k -means algorithm showed that the optimum k fluctuated between $k = 14$ and 15 for both subsets of the MERRA-2 data. Note that the steep negative slope of the line connecting the two extreme points results in the line forming the shortest distance to each data point having a nearly horizontal slope.

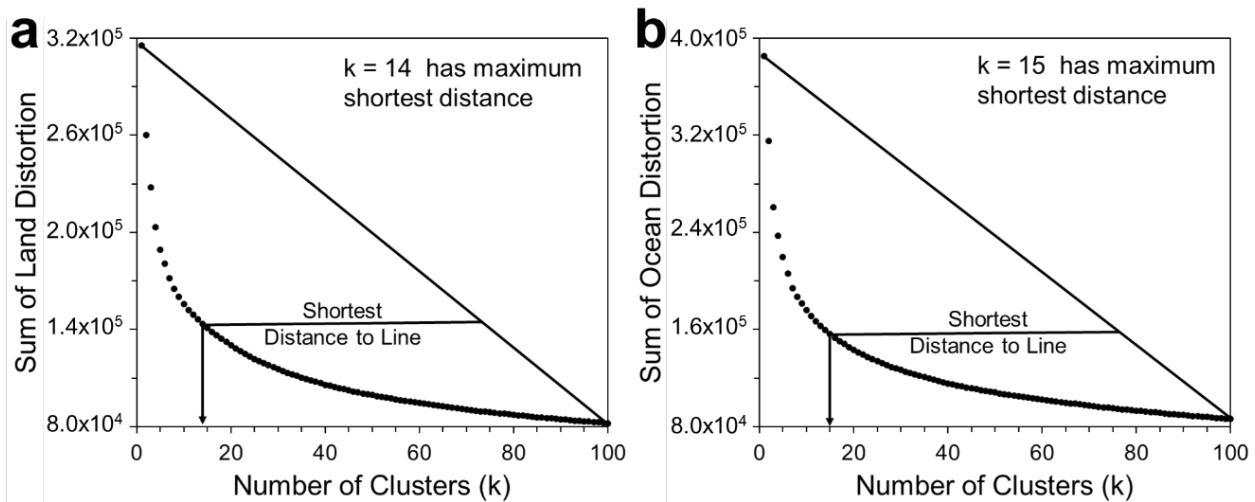


Figure A1. The sum of the distortion produced by the k -means algorithm for all points in all clusters for a given k . The k -means algorithm was performed on the subset of land and ocean data from the full 2016 MERRA-2 dataset. Panel a shows the distortion for the land dataset, and panel b shows the distortion for the ocean dataset. The point on each plot that falls farthest from the line connecting the first and last point in the dataset is indicated.

We found the “Kneedle” method to be sensitive to the point selected as the maximum number of clusters. For example, if a line is drawn between points at $k = 1$ and $k = 50$ (instead of $k = 100$), the “Kneedle” algorithm determines $k = 9$ is optimal for both land and ocean datasets. Therefore,

we evaluated several other methods to identify local points of maximum curvature in the distortion versus k plots.

Optimizing k Using the Geometric Angle-Based Method

Another method for identifying the point of maximum curvature for a set of data is the geometric angle-based method (Zhao et al. 2008), which compares the y -coordinates for point k and the two adjacent points $k-1$ and $k+1$ on the curve. The first step in the geometric angle-based method is to evaluate the formula

$$y_{k-1} + y_{k+1} - 2y_k \quad (A1)$$

In this application, the y -coordinates are the distortion values from the plots given in Figure A1. The resulting plot of equation A1 evaluated as a function of k is given in Figure A2a for the 2016 land subset of the MERRA-2 dataset. At local points of maximum curvature, equation A1 evaluates to be greater than the adjacent points, appearing as peaks in Figure A2a. To determine which peaks represent the most significant curvature in the data as a whole, the angle formed between each point in Figure A2a and its two adjacent points is evaluated and presented in Figure A2b. The angle data in Figure A2b is only given where a local maximum is present in Figure A2a for $k \geq 7$. Where the angle between adjacent points is smallest, the curvature in the distortion dataset is greatest.

Analyzing the local maxima in Figure A2a in conjunction with the minimum angles found in Figure A2b indicates that $k = 7, 14$, and 40 are points where particularly strong curvature is observed in the distortion versus k plot for the land subset of the 2016 MERRA-2 data. A broader sample of points with significant curvature include $k = 16, 20, 22, 24, 32, 38$, and 42 . This method

generally agrees with the “Kneedle” method for the land dataset, which indicated $k = 14$ as a point of maximum curvature.

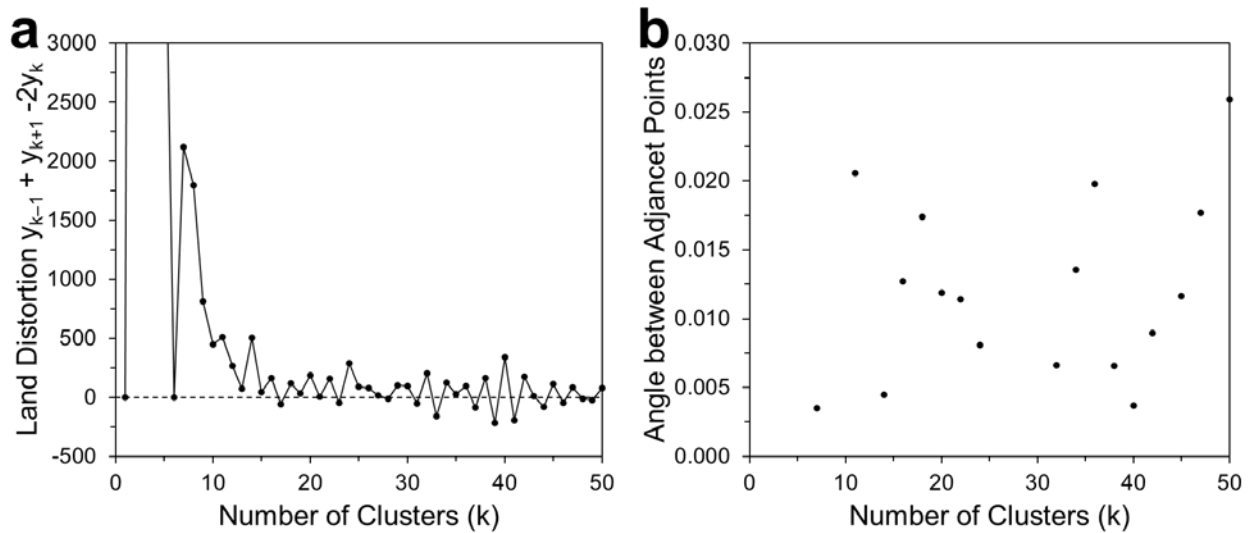


Figure A2. Geometric angle-based method for land dataset. Panel a shows equation A1 evaluated for the distortion data provided by the k-means algorithm performed on the land subset of the 2016 MERRA-2 dataset. Local maxima represent points of maximum curvature in the distortion versus k plot. Panel b gives the angle (radians) formed by the local maxima in panel a and their two adjacent points.

The results from applying the geometric angle-based method to the ocean subset of the 2016 MERRA-2 dataset are presented in Figure A3. Similar to the land dataset, points of maximum curvature appear as local maxima in the plot in Figure A3a. The angle formed by the local maxima and the two adjacent points is given in Figure A3b for the points having local maximum curvature.

The results presented in Figure A3 for the ocean dataset indicate that the point exhibiting the highest overall curvature occurs at $k = 7$. Other points where local curvature is high occur at $k = 13, 15, 17, 19, 26, 36, 38,$ and 40 . Possibly by coincidence, the results for the ocean dataset agree with the results for the land dataset for where $k = 7$ and 40 , but many of the points identified as having high local curvature fall between those identified for the land dataset. Comparing the results

of the geometric angle-based method with the results from the “Kneedle” algorithm performed on the ocean dataset indicates that both methods agree for $k = 15$.

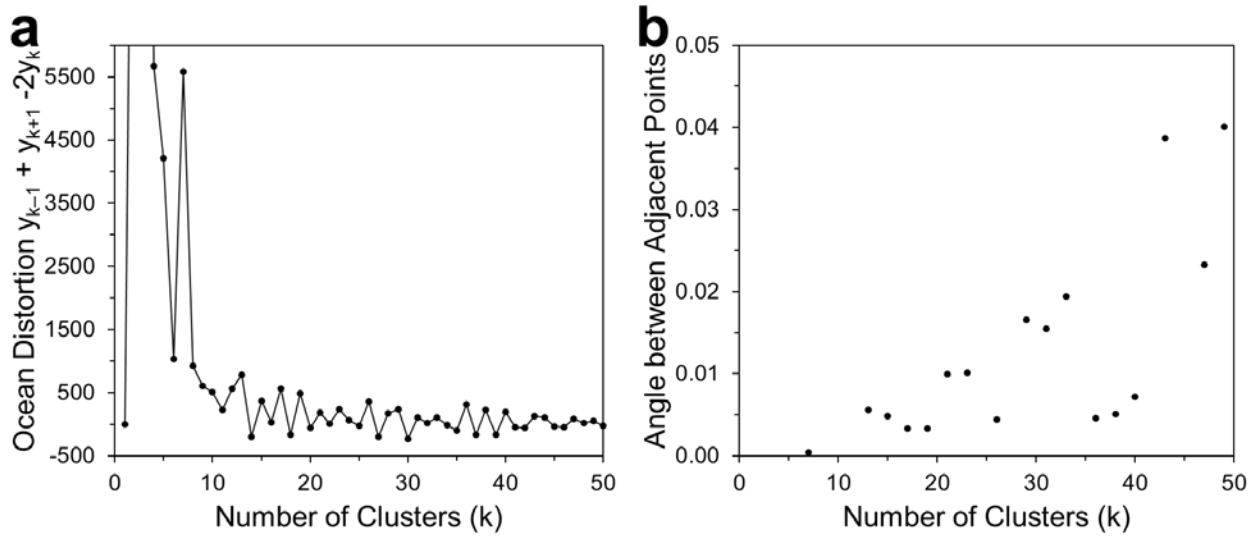


Figure A3. Geometric angle-based method for ocean dataset. Panel a shows equation A1 evaluated for the distortion data provided by the k-means algorithm performed on the ocean subset of the 2016 MERRA-2 dataset. Local maxima represent points of maximum curvature in the distortion versus k plot. Panel b gives the angle (radians) formed by the local maxima in panel a and their two adjacent points.

Optimizing k Using the Menger Curvature Method

The local curvature for a set of data can further be analyzed using the Menger curvature method (Tolsa 2000). The Menger curvature for three points is the reciprocal of the radius for a circle that passes through the three points. The Menger curvature was evaluated for the distortion versus k data from Figure A1 in order to identify local areas of maximum curvature. For each value of k , the angle, θ , formed by the points at $k-1$, k , and $k+1$ was calculated by evaluating the lengths of the sides of the triangle formed by the three points, as shown in Figure A4, where the length, L , is the distance between points at $k-1$ and $k+1$. The radius, r , of a circle circumscribed by the points that define the triangle in Figure A4 is given by equation A2.

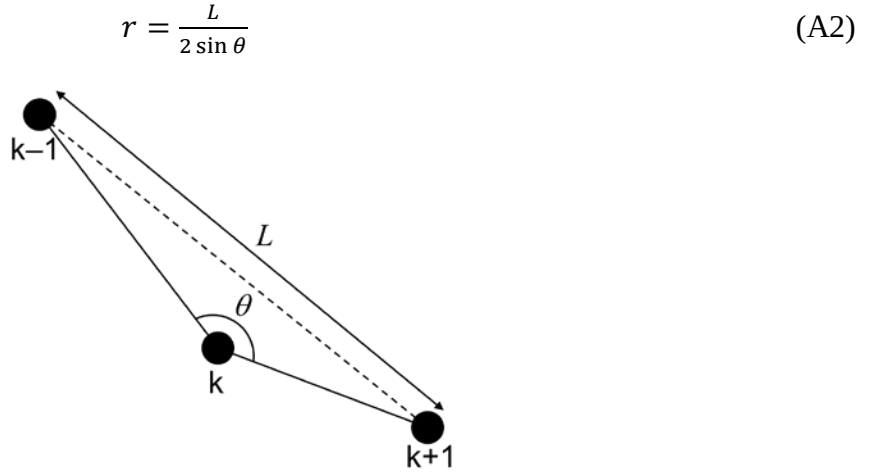


Figure A4. Depiction of the angle (θ) created by three points.

The Menger curvature is calculated by taking the reciprocal of the radius calculated by equation A2. Figure A5 presents the results for the calculation of Menger curvature as a function of k for the land distortion (Figure A5a) and ocean distortion (Figure A5b) plots presented in Figure A1. The steep negative slope of the distortion data results in θ approaching 180° (π radians) for many values of k , yielding large r and thus small Menger curvature.

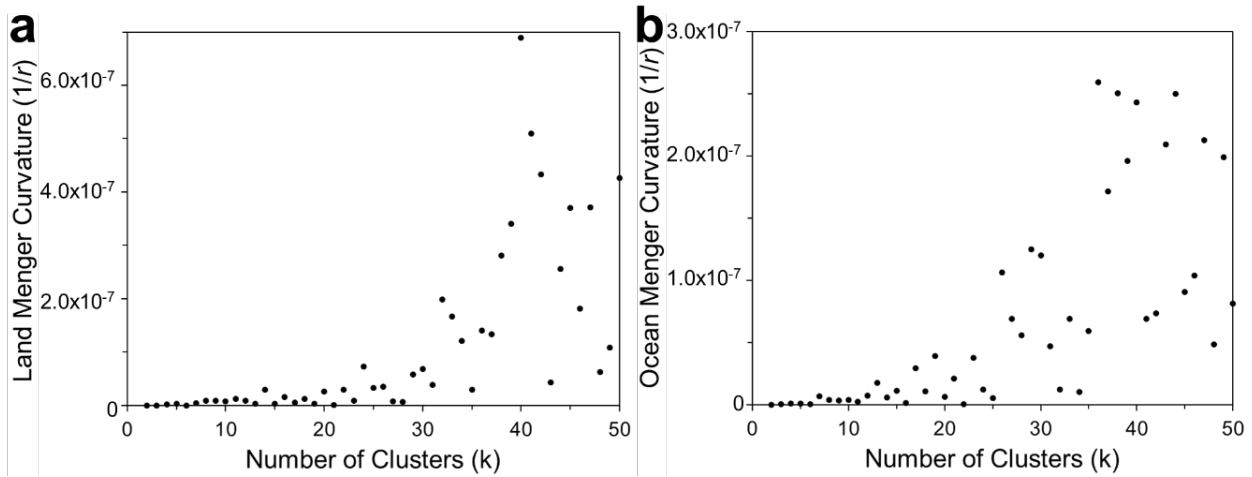


Figure A5. The Menger curvature is presented for (a) the land distortion versus k data in Figure A1a and (b) the ocean distortion versus k data in Figure A1b as a function of k . Points with high Menger curvature have high local curvature in the distortion versus k plots.

The results presented in Figure A5 indicate values of k for which the local curvature of the distortion versus k plots in Figure A1 is high. The Menger curvature tends to grow significantly greater as k increases, but local maxima in curvature are clearly visible throughout the data. For the land subset, values of k having high Menger curvature include $k = 14, 20, 22, 24, 32$, and 40 . These results agree well with the points of maximum curvature determined for the land dataset using the geometric angle-based method presented in Figure A2. The Menger curvature for the ocean subset shows high curvature at $k = 13, 17, 19, 23, 26, 29, 36, 38, 40$, and 44 . Again, there is significant overlap between the results from the Menger curvature method and the geometric angle-based method for finding points of maximum curvature for the ocean dataset.

Optimizing k for Minimal Cluster Overlap

The methods that identify local areas of high curvature in the k -means distortion versus k plots do not consider whether redundant clusters with closely overlapping centroids are produced by the k -means algorithm. To identify values of k that produce clusters with minimal overlap, we assumed that the cluster centroids represented the average value for each cluster and that the standard deviation of the data associated with each centroid could be used to represent a Gaussian population. Some centroids were associated with data that had 0 standard deviation. In these cases, we assigned a small standard deviation of 10^{-9} . The integral overlap between each of the Gaussian cluster centroids was computed for each variable as a function of k and the results are presented in Figure A6. The results are averaged for 25 iterations of the k -means algorithm from $k = 1$ to 50 .

The data in Figure A6 show relatively noisy data, particularly for $k < 10$, although the overlap observed for both the land and ocean subsets exhibited a local minimum at $k = 7$ clusters. For the land subset of the data, the cluster centroid overlap was found to exhibit other local minima at $k = 17, 22, 24, 29, 32, 33, 35, 41, 45, 47$, and 49 . A slow trend toward lower overlap was observed

for the land data for $k > 14$, which was the point of maximum overlap. The data from the ocean subset resulted in clusters that had local minimum overlap at $k = 23, 30, 35, 37, 42, 46$, and 49 , and the overlap trended toward a horizontal asymptote as k increased.

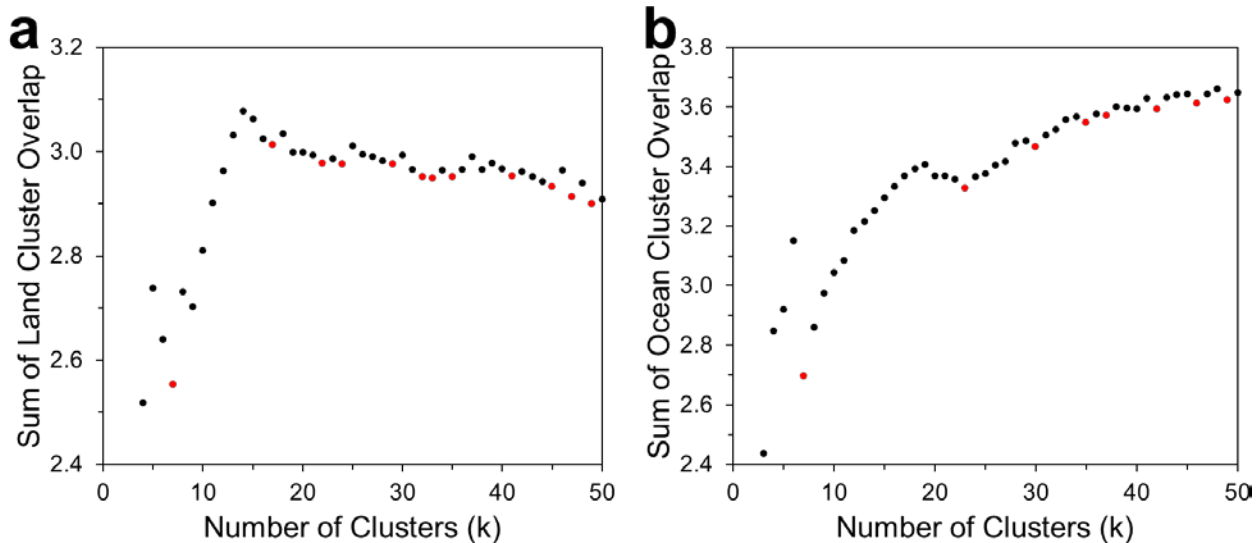


Figure A6. The sum of the integrated overlap for the Gaussian cluster centroids for (a) the land subset and (b) the ocean subset of the 2016 MERRA-2 dataset as a function of number of clusters. Local minima are indicated in red.

Comparison of Results

The “Kneedle” method is convenient because it yields a single answer that it identifies as an optimized value for k . However, the observed dependence of this result on the range of the data being analyzed demonstrates that this method has some subjective dependence on user input parameters. Comparison of the results from the “Kneedle” method with the geometric angle-based method and Menger curvature method allows for a more complete picture of which values of k fall on local “knees” in the distortion versus k curves. These three methods applied to the land subset of the data agree that $k = 14$ exhibits high curvature. However, the centroids generated by the k -means algorithm for $k = 14$ showed the greatest degree of overlap, suggesting that the 14 clusters are least unique.

Of the four techniques used to quantify curvature and cluster overlap, three methods agreed at $k = 14, 22, 24$, and 32 for the land dataset. The methods used to evaluate curvature tended to agree most often, given that they were performed on the same dataset and evaluated the same parameters. The high cluster overlap observed for $k = 14$ suggests that this result is not ideal for the purposes of identifying unique and representative clusters. Considering that $k = 22, 24$, and 32 exhibited similar “knee” behavior and have favorable centroid overlap, we were interested in selecting a number for k that would produce lookup tables containing a manageable amount of meaningful information. Therefore, we selected $k = 22$ clusters to represent the land dataset in this work.

The analysis performed to identify optimal values for k for the ocean dataset yielded no consensus where three of the four methods agreed. The ocean data were clustered using the same variables as the land data, but the ocean data consisted of considerably lower variation for the variables that defined the terrain (all points had fraction of ocean = 1) and the elevation. The ocean dataset is, therefore, more uniform and does not require as many clusters as the land dataset. Further, the purpose of defining representative clusters in this work is to allow for the rapid identification of an appropriate AOD_{550} to use for atmospheric modeling purposes, which we anticipate is of lower interest over the open ocean. The results for optimizing k for the ocean dataset suggest that $k = 7$ clusters could be most appropriate. We noted, however, that assigning 7 ocean clusters yielded numerous discontinuous regions that were difficult to describe geographically and may add unnecessary bulk to the results. Therefore, we determined that fewer ocean clusters were warranted and settled on using $k = 3$ clusters to represent the ocean dataset.

B. Full 2016 Global Clusters

Table B1. Cluster centroids for the full 2016 global MERRA-2 dataset.

Cluster	Fraction of Lake	Fraction of Land	Fraction of Land-Ice	Fraction of Ocean	Elevation (meters)	Relative Humidity (%)	2-m Air Temperature (K)	10-m Wind Speed (m s ⁻¹)	Fraction Showers	Fraction Cloud Coverage	Fraction Black Carbon	Fraction Dust	Fraction Organic Carbon	Fraction Sea Salt	Fraction Sulfate	Total AOD
1	3.79x10 ⁻⁹	3.02x10 ⁻⁷	2.16x10 ⁻⁷	1.00	0.0357	93.1	265	5.63	0.925	0.817	0.0567	0.133	0.241	0.156	0.413	0.0850
2	0	3.69x10 ⁻⁸	1.89x10 ⁻⁸	1.00	7.09x10 ⁻⁵	86.8	278	8.41	0.946	0.776	0.0244	0.0571	0.0982	0.603	0.218	0.102
3	7.93x10 ⁻¹⁰	2.03x10 ⁻⁷	0	1.00	-0.0286	79.5	296	6.05	0.857	0.578	0.0383	0.0969	0.123	0.490	0.252	0.120
4	0	5.06x10 ⁻⁵	1.00	0	2965	96.9	226	6.70	0.863	0.624	0.0630	0.110	0.241	0.139	0.447	0.0171
5	0	5.05x10 ⁻⁸	0.998	2.15x10 ⁻³	1944	86.1	237	10.5	0.731	0.551	0.0600	0.119	0.217	0.182	0.423	0.0196
6	0	2.64x10 ⁻³	0.954	0.0430	511	88.6	250	5.85	0.813	0.668	0.0459	0.100	0.182	0.269	0.404	0.0292
7	3.45x10 ⁻³	0.0642	0.865	0.0676	1435	87.1	256	5.99	0.734	0.555	0.0757	0.155	0.284	0.0737	0.412	0.0470
8	0.0131	0.941	0.0148	0.0314	536	90.0	268	2.87	0.880	0.680	0.0607	0.140	0.283	0.0598	0.456	0.106
9	5.41x10 ⁻³	0.152	0.0480	0.795	36.3	90.9	270	5.16	0.884	0.738	0.0570	0.128	0.227	0.154	0.435	0.0960
10	0.144	0.823	6.47x10 ⁻³	0.0269	277	87.6	271	3.17	0.866	0.660	0.0577	0.141	0.256	0.0771	0.468	0.112
11	0.0146	0.962	9.58x10 ⁻⁴	0.0224	241	88.0	273	3.15	0.871	0.677	0.0486	0.150	0.203	0.0575	0.541	0.146
12	6.99x10 ⁻³	0.985	3.81x10 ⁻³	3.75x10 ⁻³	1766	54.0	280	3.28	0.589	0.429	0.0478	0.417	0.128	0.0280	0.379	0.152
13	0.535	0.462	5.59x10 ⁻⁶	3.12x10 ⁻³	460	74.5	282	3.46	0.765	0.577	0.0608	0.204	0.276	0.0666	0.393	0.146
14	7.24x10 ⁻³	0.952	3.71x10 ⁻⁴	0.0402	499	75.7	283	3.04	0.734	0.542	0.0646	0.140	0.185	0.0715	0.538	0.166
15	2.78x10 ⁻³	0.0787	0.0404	0.878	21.2	82.9	286	5.84	0.870	0.683	0.0338	0.0699	0.125	0.461	0.310	0.102
16	4.82x10 ⁻³	0.991	2.89x10 ⁻³	1.71x10 ⁻³	2120	45.3	287	3.15	0.512	0.366	0.0830	0.164	0.278	0.0646	0.411	0.0914
17	7.75x10 ⁻³	0.983	2.13x10 ⁻³	7.30x10 ⁻³	615	66.5	289	3.01	0.651	0.474	0.0675	0.135	0.290	0.103	0.405	0.120
18	9.87x10 ⁻³	0.941	1.90x10 ⁻⁵	0.0489	404	67.3	292	2.67	0.635	0.474	0.0808	0.151	0.164	0.0406	0.563	0.439
19	2.44x10 ⁻³	0.150	0	0.847	29.5	76.6	293	4.77	0.599	0.426	0.0516	0.212	0.144	0.206	0.386	0.201
20	2.60x10 ⁻³	0.963	0	0.0347	611	34.9	294	3.42	0.355	0.241	0.0366	0.613	0.0755	0.0443	0.230	0.255
21	1.54x10 ⁻³	0.970	0	0.0287	285	47.1	295	3.86	0.538	0.352	0.0509	0.243	0.180	0.231	0.295	0.0924
22	0.0142	0.869	2.73x10 ⁻³	0.115	260	74.4	296	2.91	0.830	0.635	0.0453	0.108	0.204	0.348	0.296	0.131
23	0.0131	0.974	0	0.0129	528	71.8	299	1.48	0.819	0.648	0.0777	0.127	0.491	0.0858	0.218	0.206
24	2.21x10 ⁻³	0.986	0	0.0119	473	27.5	299	3.75	0.518	0.364	0.0211	0.786	0.0593	0.0234	0.110	0.388
25	9.90x10 ⁻³	0.927	0	0.0632	410	61.5	300	2.69	0.791	0.619	0.0482	0.538	0.222	0.0497	0.142	0.349

Blue = Ocean; Orange = Permafrost; Red = Inland; Green = Lake; Purple = Coast

Table B2. Category percentiles for the ocean subset of the 2016 MERRA-2 dataset.

Percentile	Fraction of Lake	Fraction of Land	Fraction of Land-Ice	Fraction of Ocean	Elevation (meters)	Relative Humidity (%)	2-m Air Temperature (K)	10-m Wind Speed (m s^{-1})	Fraction Showers	Fraction Cloud Coverage	Fraction Black Carbon	Fraction Dust	Fraction Organic Carbon	Fraction Sea Salt	Fraction Sulfate	Total AOD
10	0	0	0	1.00	0	77.4	263	5.09	0.812	0.481	0.0187	0.0295	0.0712	0.122	0.163	0.0676
25	0	0	0	1.00	0	79.2	274	5.57	0.874	0.581	0.0245	0.0466	0.0901	0.359	0.192	0.0765
50	0	0	0	1.00	0	82.4	288	6.49	0.914	0.696	0.0355	0.0699	0.118	0.509	0.237	0.0985
75	0	0	0	1.00	0	88.1	298	7.69	0.946	0.797	0.0467	0.122	0.154	0.630	0.331	0.120
90	0	0	0	1.00	0	93.4	301	9.03	0.966	0.848	0.0600	0.149	0.254	0.694	0.429	0.158

Table B3. Category percentiles for the land subset of the 2016 MERRA-2 dataset.

Percentile	Fraction of Lake	Fraction of Land	Fraction of Land-Ice	Fraction of Ocean	Elevation (meters)	Relative Humidity (%)	2-m Air Temperature (K)	10-m Wind Speed (m s^{-1})	Fraction Showers	Fraction Cloud Coverage	Fraction Black Carbon	Fraction Dust	Fraction Organic Carbon	Fraction Sea Salt	Fraction Sulfate	Total AOD
10	0	0	0	0	6.67	45.0	226	1.96	0.511	0.351	0.0349	0.0831	0.109	0.0200	0.209	0.0169
25	0	0	0	0	105	69.3	252	2.85	0.685	0.501	0.0474	0.104	0.163	0.0447	0.348	0.0318
50	0	0.831	0	0	412	85.0	273	4.07	0.810	0.612	0.0581	0.129	0.222	0.115	0.428	0.0976
75	3.77×10^{-3}	1.00	1.00	0.0160	1476	93.4	293	5.87	0.882	0.685	0.0649	0.165	0.259	0.183	0.459	0.157
90	0.0261	1.00	1.00	0.820	2850	96.7	300	8.21	0.922	0.739	0.0747	0.428	0.305	0.303	0.521	0.269

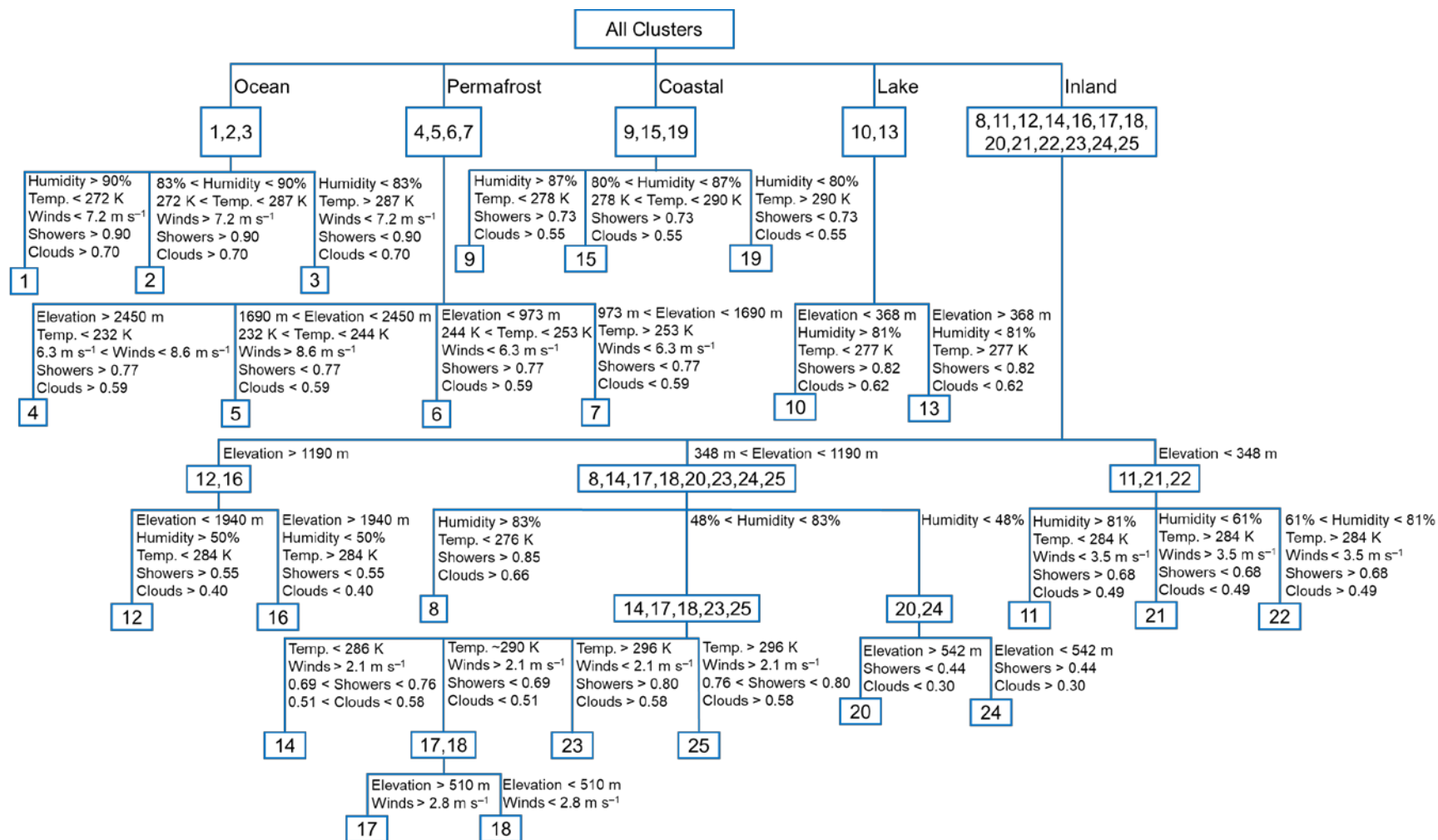


Figure B1. Decision tree for clusters describing the full 2016 global MERRA-2 dataset.

Table B4. Cluster descriptions and AOD₅₅₀ statistics for the full 2016 global MERRA-2 dataset.

Cluster	Qualitative Description	Characteristic Regions	AOD Centroid	AOD Average	AOD Standard Deviation	AOD 10th Percentile	AOD 50th Percentile	AOD 90th Percentile
1	Regions of the oceans: Humidity: high; Temperature: low; Wind speed: medium low; Showers: medium high; Cloud coverage: high	Arctic Ocean, Hudson Bay, Sea of Okhotsk, northern Bearing Sea, southernmost portion of Southern Sea.	0.0850	0.0847	0.0461	0.0624	0.0741	0.110
2	Regions of the oceans: Humidity: medium high; Temperature: medium low; Wind speed: high ;Showers: high ;Clouds coverage: medium high	Median temperature regions of the Southern and Atlantic Oceans, including a central portion of the Pacific.	0.102	0.102	0.0261	0.0629	0.105	0.132
3	Regions of the oceans: Humidity: medium low; Temperature: medium high; Wind speed: medium low; Showers: low; Clouds coverage: low	Majority of oceans and seas in mid-latitudes. Extends in an arc past the British Isles and north of Norway.	0.120	0.120	0.0620	0.0700	0.103	0.192
4	Inland permafrost regions: Elevation: very high; Humidity: very high; Temperature: low; Wind speed: high; Showers: medium high; Cloud coverage: medium high	Coldest regions of central Greenland and inland Antarctica with highest elevation.	0.0171	0.0171	3.95x10 ⁻³	0.0150	0.0164	0.0185
5	Inland permafrost regions: Elevation: high; Humidity: medium high; Temperature: low; Wind speed: very high; Showers: medium low; Cloud coverage: medium low	Slightly warmer and windy regions of inland Antarctica at lower elevation.	0.0196	0.0196	3.61x10 ⁻³	0.0157	0.0189	0.0240
6	Inland permafrost regions: Elevation: medium high; Humidity: medium high; Temperature: low; Wind speed: medium high; Showers: medium high; Cloud coverage: medium high	Slightly warmer, less windy regions of inland Antarctica nearest the coast at lower elevation.	0.0292	0.0292	8.70x10 ⁻³	0.0232	0.0272	0.0373
7	Inland permafrost regions: Elevation: medium high; Humidity: medium high; Temperature: medium low; Wind speed: high; Showers: medium low; Cloud coverage: medium low	Inland Greenland (except the center) and far northern Canada.	0.0470	0.0470	7.16x10 ⁻³	0.0398	0.0468	0.0553
8	Inland regions: Elevation: medium high; Humidity: medium high; Temperature: medium low; Wind speed: medium low; Showers: medium high; Cloud coverage: medium high	Inland Alaska and Norther Canada, eastern Russia (Siberia), Norway, Malaysia and Indonesia, and a strip of equatorial South America east of the Andes mountains.	0.106	0.105	0.0411	0.0649	0.0971	0.164
9	Coastal regions: Elevation: low; Humidity: medium high; Temperature: medium low; Wind speed: medium high; Showers: high; Cloud coverage: high	Northern coastal regions of Canada, Alaska, northern Europe including the Nordic countries, and northern and eastern Russia. Includes some coastal regions in Malaysia and Indonesia and a small portion of coastal Antarctica.	0.0960	0.0965	0.0396	0.0631	0.0891	0.133
10	Inland regions near lakes: Elevation: medium low; Humidity: medium high; Temperature: medium low; Wind speed: medium low; Showers: medium high; Cloud coverage: medium high	Scattered global areas near lakes concentrated in northern Canada and northern Eurasia.	0.112	0.115	0.0426	0.0732	0.107	0.168
11	Inland regions: Elevation: medium low; Humidity: medium high; Temperature: medium high; Wind speed: medium low; Showers: medium high; Cloud coverage: medium high	Significant portions of inland central Russia, including some inland parts of Indonesia and southwestern Alaska.	0.146	0.146	0.0366	0.0995	0.146	0.188
12	Inland regions: Elevation: high; Humidity: low; Temperature: medium high; Wind speed: medium low; Showers: low; Cloud coverage: low	Regions of western China excluding Tibet and the Taklamakan and Gobi deserts extending west through Uzbekistan and southern Kazakhstan to the Caspian Sea. Includes inland regions west of the Caspian sea through Turkey, inland but near coastal north Africa, and portions of eastern Ethiopia.	0.152	0.152	0.0499	0.0870	0.149	0.221

UNCLASSIFIED

13	Inland regions near lakes: Elevation: medium high; Humidity: medium low; Temperature: medium high; Wind speed: medium low; Showers: medium low; Cloud coverage: medium low	Scattered global areas near lakes, concentrated in northwestern Canada, the Amazon river, equatorial eastern Africa, northeastern Europe, and central Asia.	0.146	0.145	0.0610	0.0805	0.133	0.199
14	Inland regions: Elevation: medium high; Humidity: medium low; Temperature: medium high; Wind speed: medium low; Showers: medium low; Cloud coverage: medium low	Inland regions of northeastern and Midwestern United States, France, England, and northern Spain, east just past the Black sea and including some Mediterranean regions, northeastern China, Thailand, Laos, and southern India.	0.166	0.167	0.0557	0.111	0.153	0.256
15	Coastal regions: Elevation: low; Humidity: medium low; Temperature: medium high; Wind speed: medium high; Showers: medium high; Cloud coverage: medium high	Coastal areas near the equator and islands, also includes coastal Iceland, southern Greenland, the northern British Isles, Norway, the Alaskan archipelago, southern South America, and most of coastal Antarctica.	0.102	0.102	0.0466	0.0435	0.100	0.162
16	Inland regions: Elevation: high; Humidity: low; Temperature: medium high; Wind speed: medium low; Showers: low; Cloud coverage: low	Inland areas of the southwestern United States and northern Mexico, eastern Chile and western Argentina, southern Africa (Namibia to South Africa), and Tibet.	0.0914	0.0915	0.0349	0.0573	0.0820	0.140
17	Inland regions: Elevation: medium high; Humidity: low; Temperature: medium high; Wind speed: medium low; Showers: low; Cloud coverage: low	Southeastern United States to eastern Texas, western plains states into southern Canada and Pacific Northwest, central South America from the southern half of Brazil to central Argentina, central Madagascar and near coastal but inland strip of southeast Africa, central Spain, eastern Mongolia, portions of Myanmar.	0.120	0.120	0.0359	0.0770	0.117	0.158
18	Inland regions: Elevation: medium low; Humidity: low; Temperature: medium high; Wind speed: low; Showers: low; Cloud coverage: low	Eastern China and the Korean peninsula, northern and eastern India.	0.439	0.440	0.103	0.3167	0.428	0.571
19	Coastal regions: Elevation: low; Humidity: medium low; Temperature: high; Wind speed: medium high; Showers: low; Cloud coverage: low	Coastal areas in the mid-latitudes. Includes coastal Mediterranean, coastal Black and Caspian Seas, coastal Arabian peninsula, coastal India, China, Korea, and Japan, coastal Australia, coastal Madagascar and southern Africa, coastal southern South America and western South America to Ecuador, and coastal United States & western Mexico.	0.201	0.201	0.0989	0.0983	0.176	0.357
20	Inland regions: Elevation: medium high; Humidity: very low; Temperature: high; Wind speed: medium low; Showers: very low; Cloud coverage: very low	Strip of northern Africa growing broader to the east including all of Egypt, northern half of Arabian peninsula north to Syria and east to Pakistan and northwest India, includes Iran north to Turkmenistan. Includes some of the Gobi desert in China and a small portion of central Australia.	0.255	0.254	0.0723	0.173	0.246	0.348
21	Inland regions: Elevation: medium low; Humidity: low; Temperature: high; Wind speed: medium low; Showers: low; Cloud coverage: low	Inland central Australia, inland southern Spain and southern Argentina, land around the Gulf of California, some thin inland but near coastal regions of southern Africa.	0.0924	0.0924	0.0362	0.0618	0.0847	0.129
22	Inland regions: Elevation: medium low; Humidity: medium low; Temperature: high; Wind speed: medium low; Showers: medium high; Cloud coverage: medium high	Inland central America and inland but near coastal South America from Colombia to southern Brazil, inland southern Chile, strip of inland but near coastal Africa south of Ethiopia, inland Ireland, Scotland, and New Zealand, inland areas of the Philippines and near coastal regions of Indonesia, and near coastal regions of northern and southern Australia.	0.131	0.131	0.0485	0.0777	0.124	0.191
23	Inland regions: Elevation: medium high; Humidity: medium low; Temperature: high; Wind speed: very low; Showers: medium high; Cloud coverage: medium high	Equatorial inland South America and Africa, including a small central area in southeast Asia. Regions that include jungle and tropical rainforests.	0.206	0.206	0.101	0.117	0.162	0.383
24	Inland regions: Elevation: medium high; Humidity: very low; Temperature: high; Wind speed: medium low; Showers: low; Cloud coverage: low	Bulk of the Sahara desert from Mauritania east to Sudan. Includes southern half of the Arabian peninsula and the Taklamakan desert in China.	0.388	0.389	0.0728	0.293	0.388	0.485
25	Inland regions: Elevation: medium low; Humidity: low; Temperature: high; Wind speed: low; Showers: medium low; Cloud coverage: medium high	Portion of Africa just south of the Sahara but north of the equator.	0.349	0.349	0.0919	0.248	0.338	0.480

Blue = Ocean; Orange = Permafrost; Red = Inland; Green = Lake; Purple = Coast

UNCLASSIFIED

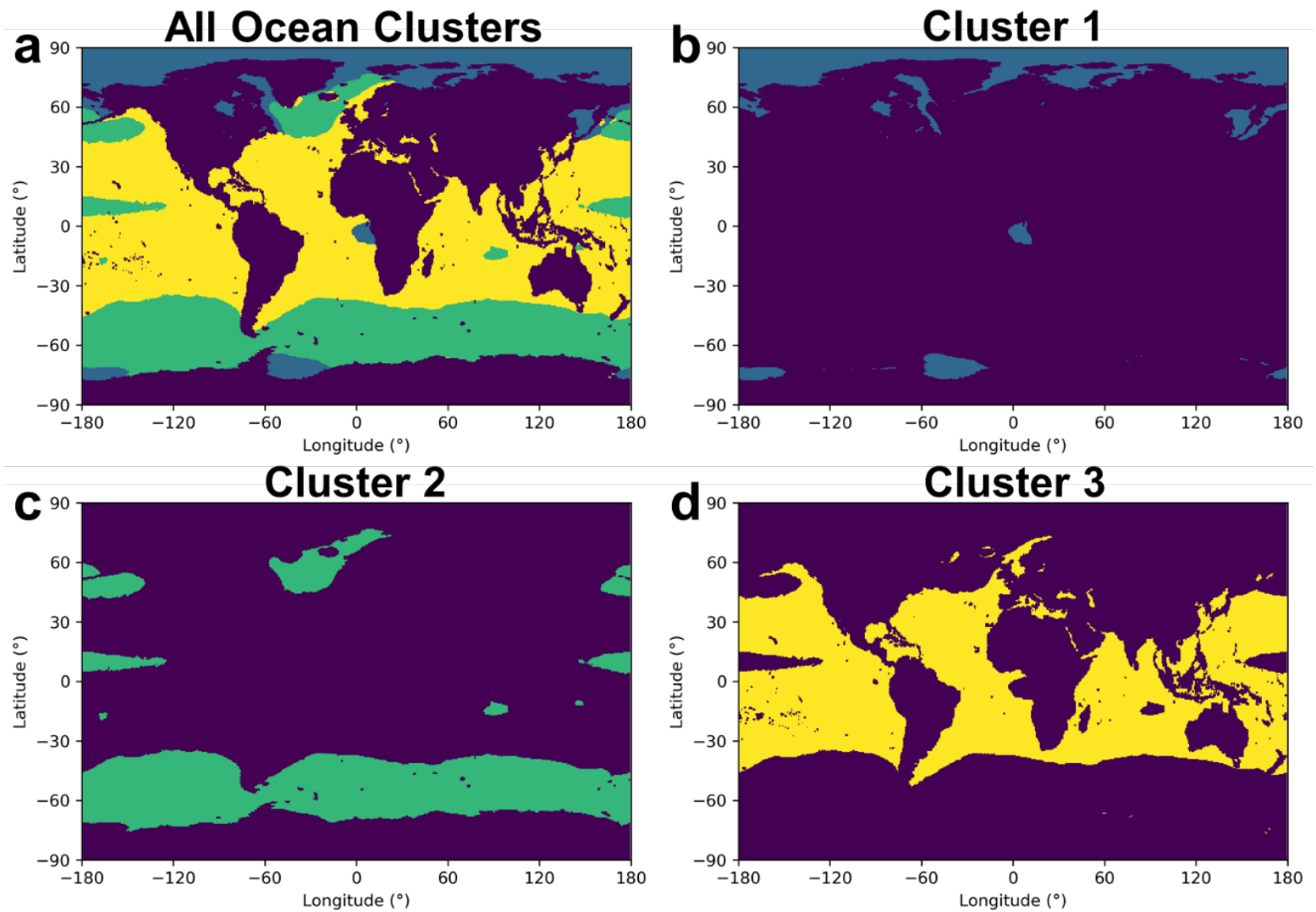


Figure B2. Ocean clusters 1 – 3 from the full 2016 global MERRA-2 dataset.

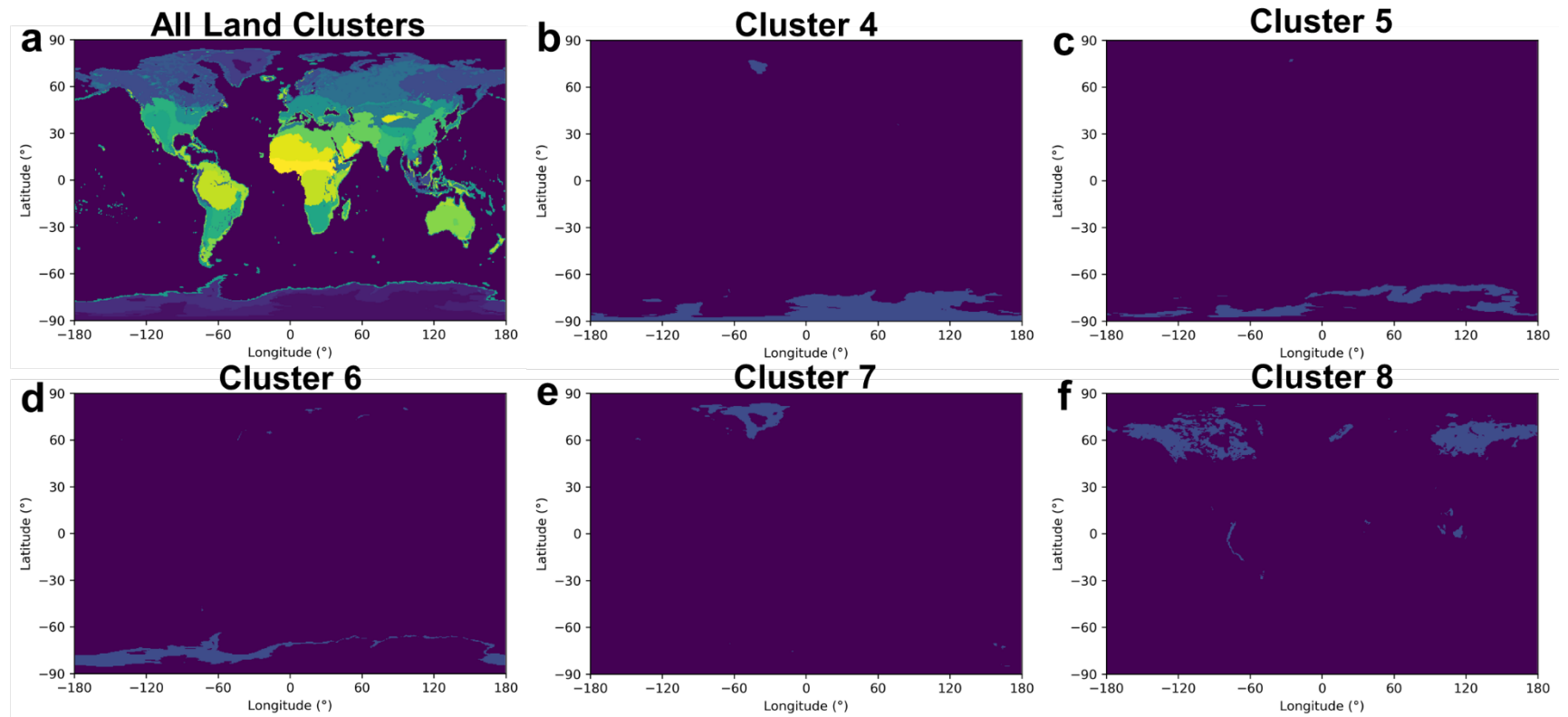


Figure B3. Land clusters 4 – 8 from the full 2016 global MERRA-2 dataset.

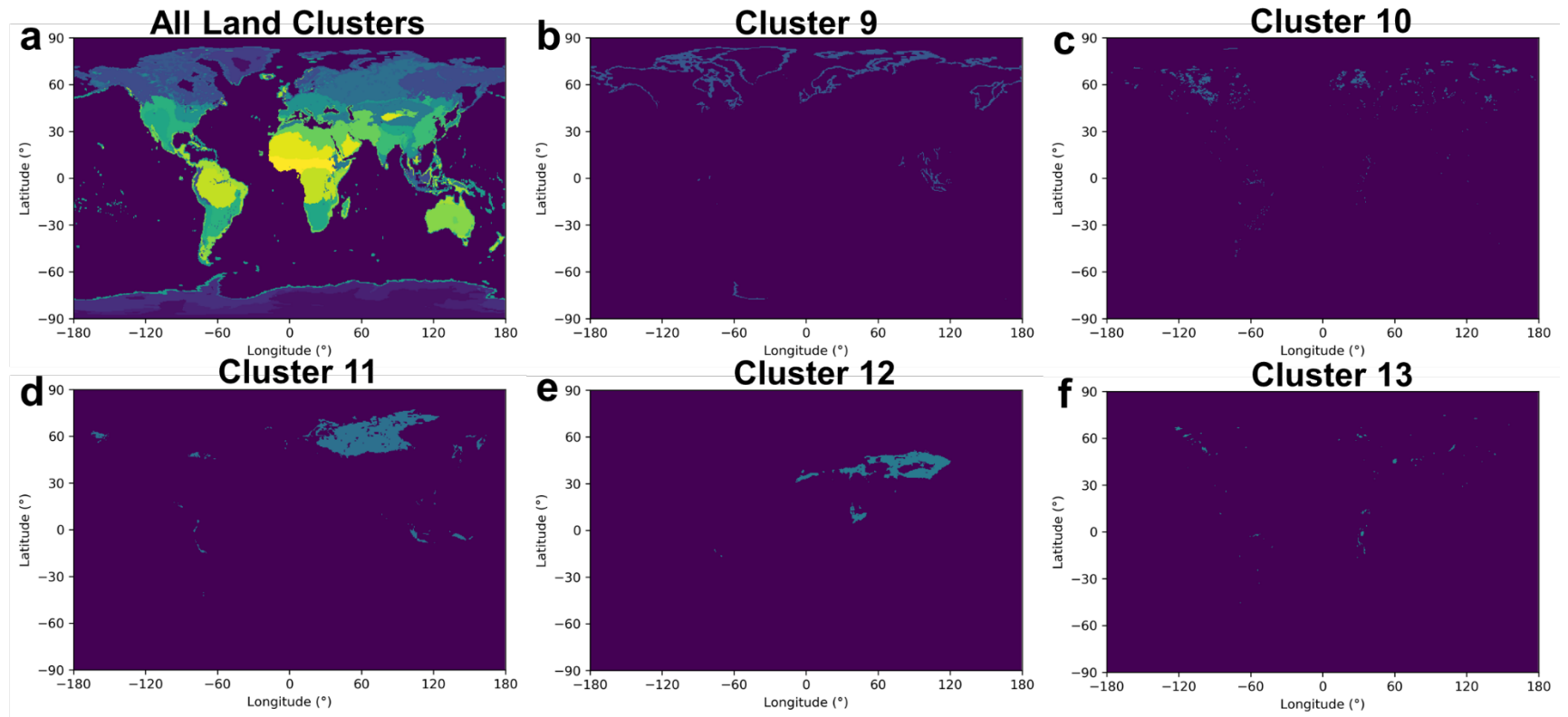


Figure B4. Land clusters 9 – 13 from the full 2016 global MERRA-2 dataset.

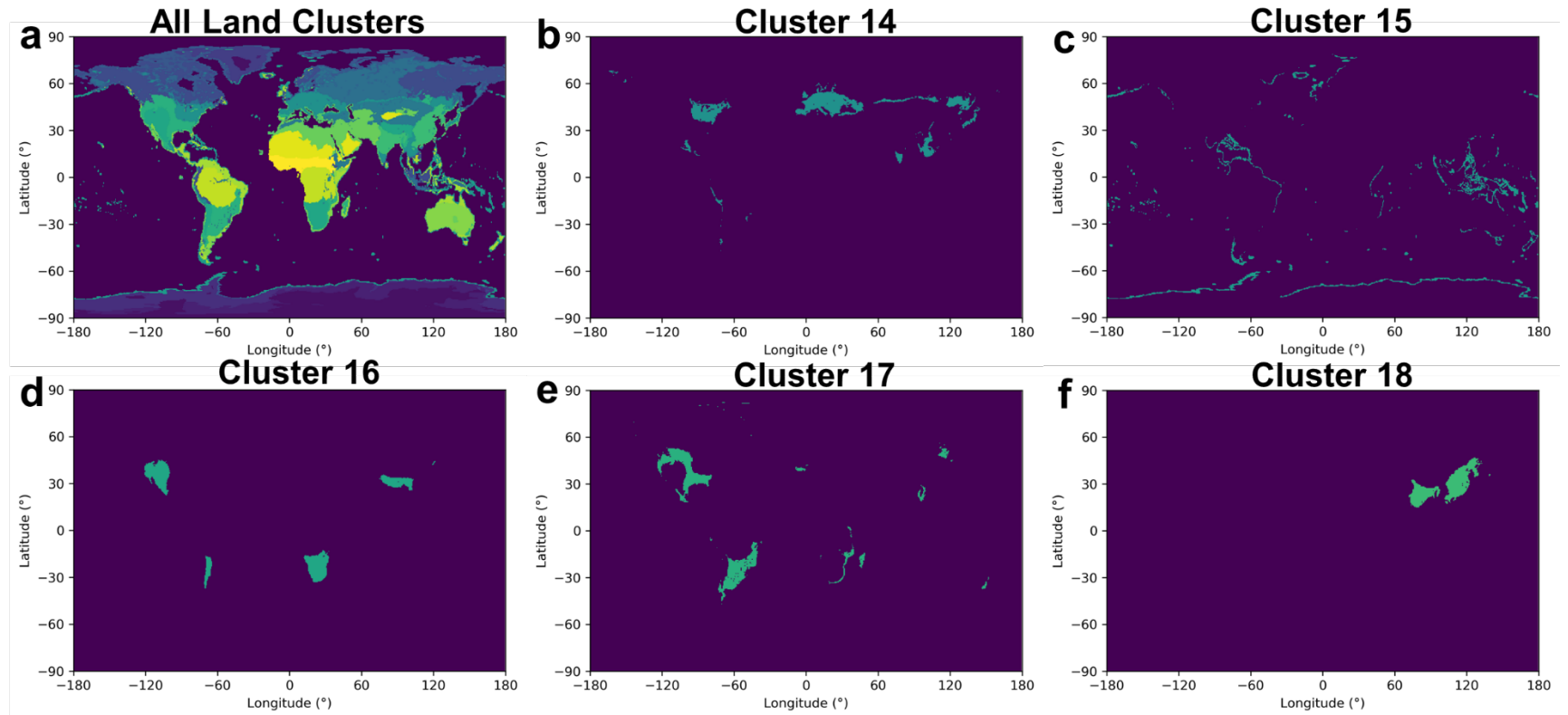


Figure B5. Land clusters 14 – 18 from the full 2016 global MERRA-2 dataset.

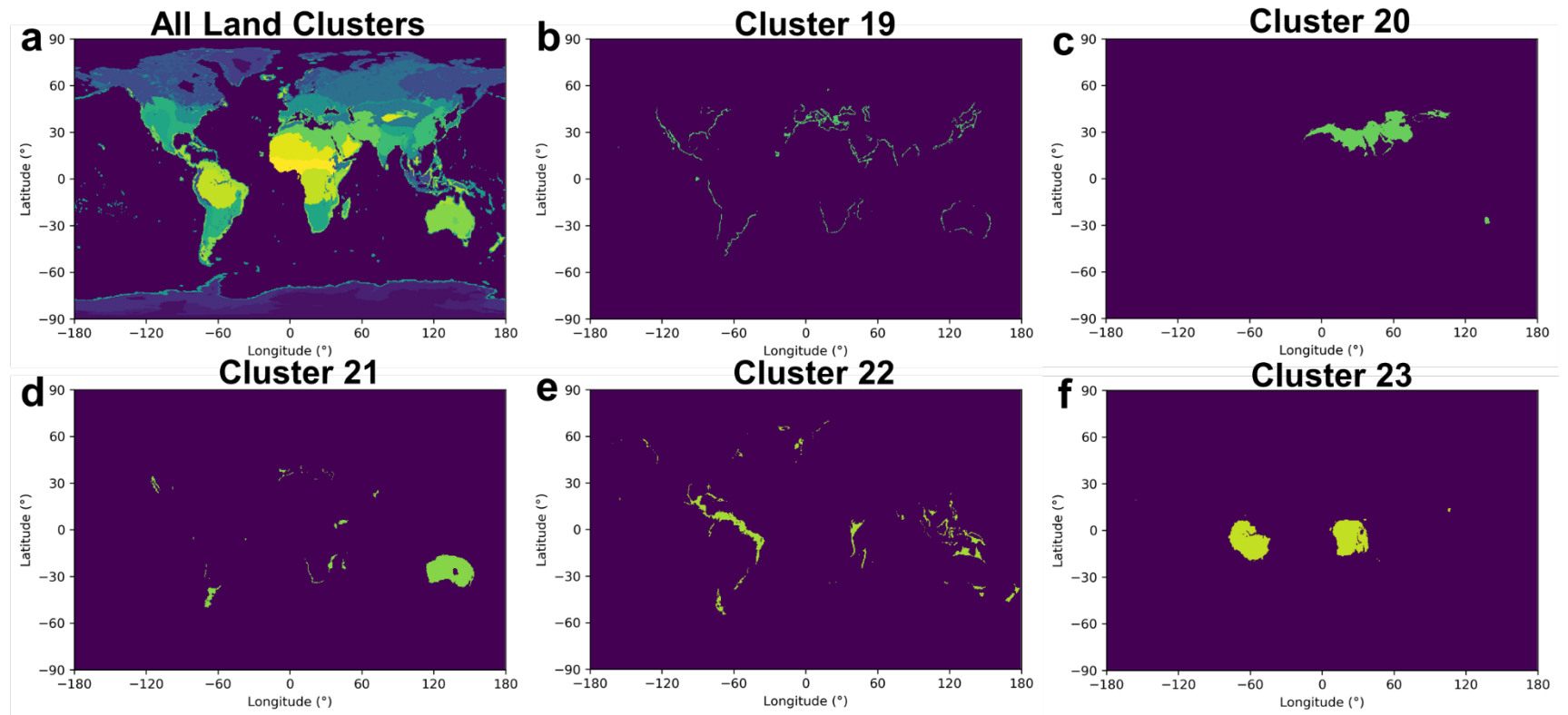


Figure B6. Land clusters 19 – 23 from the full 2016 global MERRA-2 dataset.

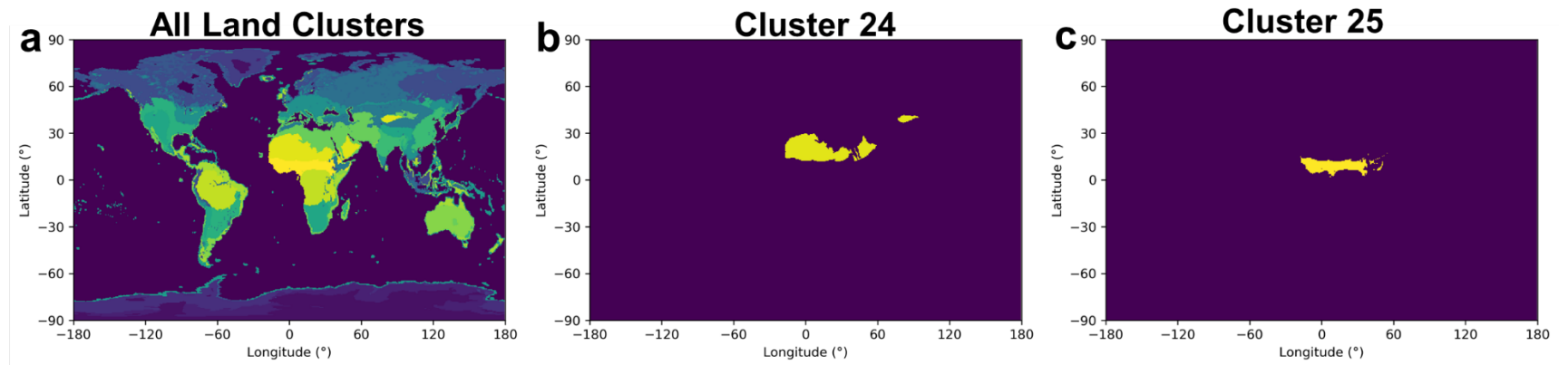


Figure B7. Land clusters 24 – 25 from the full 2016 global MERRA-2 dataset.

C. 2012 to 2016 December through February Global Clusters

Table C1. Cluster centroids for the 2012 – 2016 Dec/Jan/Feb global MERRA-2 dataset.

Cluster	Fraction of Lake	Fraction of Land	Fraction of Land-Ice	Fraction of Ocean	Elevation (meters)	Relative Humidity (%)	2-m Air Temperature (K)	10-m Wind Speed (m s ⁻¹)	Fraction Showers	Fraction Cloud Coverage	Fraction Black Carbon	Fraction Dust	Fraction Organic Carbon	Fraction Sea Salt	Fraction Sulfate	Total AOD
1	4.28x10 ⁻⁹	3.08x10 ⁻⁷	1.79x10 ⁻⁷	1.00	0.0300	95.3	253	5.93	0.961	0.878	0.0538	0.124	0.148	0.150	0.527	0.0707
2	5.42x10 ⁻¹⁰	1.47x10 ⁻⁷	2.69x10 ⁻⁸	1.00	1.57x10 ⁻⁴	84.9	282	7.82	0.939	0.746	0.0257	0.0582	0.0854	0.562	0.270	0.109
3	4.84x10 ⁻¹⁰	1.43x10 ⁻⁷	1.03x10 ⁻⁸	1.00	-0.0321	79.0	294	6.01	0.843	0.546	0.0438	0.100	0.135	0.466	0.257	0.119
4	1.46x10 ⁻⁵	1.33x10 ⁻³	0.998	5.92x10 ⁻⁵	2984	94.8	241	4.48	0.775	0.475	0.0706	0.141	0.220	0.0625	0.507	0.0205
5	1.47x10 ⁻⁷	1.04x10 ⁻⁴	1.00	9.55x10 ⁻⁶	2737	91.7	243	6.41	0.648	0.404	0.0704	0.149	0.211	0.0628	0.508	0.0201
6	1.06x10 ⁻⁴	5.90x10 ⁻³	0.991	3.18x10 ⁻³	2036	86.1	248	9.14	0.726	0.549	0.0698	0.135	0.210	0.111	0.476	0.0293
7	0.0216	0.909	0.0199	0.0493	424	97.9	252	3.63	0.910	0.712	0.0622	0.118	0.173	0.112	0.536	0.0717
8	5.62x10 ⁻³	0.157	0.0538	0.784	37.9	96.8	253	5.52	0.939	0.834	0.0557	0.123	0.151	0.141	0.532	0.0782
9	0.0130	0.975	9.59x10 ⁻⁶	0.0115	332	97.9	253	3.44	0.938	0.751	0.0440	0.162	0.111	0.0634	0.621	0.127
10	0.186	0.786	5.45x10 ⁻³	0.0226	293	95.4	256	3.32	0.910	0.710	0.0550	0.134	0.169	0.103	0.540	0.0985
11	3.87x10 ⁻⁴	8.07x10 ⁻³	0.974	0.0177	1085	84.4	257	6.64	0.694	0.513	0.0550	0.136	0.171	0.130	0.510	0.0311
12	1.39x10 ⁻³	0.0161	0.928	0.0543	380	89.5	260	3.92	0.762	0.594	0.0475	0.118	0.147	0.169	0.520	0.0389
13	0.612	0.385	7.95x10 ⁻⁶	3.89x10 ⁻³	490	84.2	270	3.21	0.845	0.664	0.0599	0.208	0.222	0.0863	0.426	0.129
14	3.30x10 ⁻³	0.988	7.52x10 ⁻⁶	8.45x10 ⁻³	1106	60.4	272	3.13	0.536	0.395	0.0493	0.470	0.0940	0.0279	0.361	0.147
15	6.12x10 ⁻³	0.969	6.05x10 ⁻³	0.0189	777	82.3	272	3.24	0.730	0.561	0.0713	0.133	0.184	0.0747	0.538	0.102
16	6.72x10 ⁻³	0.986	5.82x10 ⁻³	1.46x10 ⁻³	2749	52.5	274	3.27	0.536	0.403	0.0862	0.203	0.225	0.0374	0.449	0.0734
17	3.13x10 ⁻³	0.0948	0.0813	0.821	38.0	85.5	277	5.95	0.878	0.732	0.0399	0.0742	0.119	0.331	0.438	0.100
18	9.50x10 ⁻³	0.929	1.59x10 ⁻⁵	0.0613	369	59.4	285	2.45	0.492	0.331	0.105	0.137	0.193	0.0223	0.545	0.430
19	2.64x10 ⁻³	0.159	8.57x10 ⁻⁴	0.838	29.4	74.5	291	5.21	0.617	0.421	0.0573	0.196	0.137	0.218	0.393	0.173
20	3.03x10 ⁻³	0.976	0	0.0207	434	32.8	291	4.00	0.434	0.300	0.0339	0.723	0.0904	0.0296	0.125	0.244
21	4.55x10 ⁻³	0.128	3.29x10 ⁻³	0.864	23.0	80.9	293	6.44	0.861	0.642	0.0326	0.0551	0.110	0.550	0.255	0.121
22	0.0122	0.894	3.93x10 ⁻³	0.0899	294	74.9	297	2.90	0.848	0.669	0.0473	0.0866	0.209	0.360	0.299	0.132
23	0.0116	0.980	0	7.97x10 ⁻³	606	77.5	298	1.62	0.891	0.743	0.0690	0.110	0.421	0.117	0.284	0.146
24	0.0133	0.922	0	0.0643	433	63.0	299	1.88	0.804	0.643	0.0882	0.405	0.400	0.0234	0.0847	0.493
25	4.10x10 ⁻³	0.982	0	0.0139	474	42.3	299	3.91	0.560	0.382	0.0511	0.277	0.195	0.183	0.296	0.108

Blue = Ocean; Orange = Permafrost; Red = Inland; Green = Lake; Purple = Coast

Table C2. Category percentiles for the ocean subset of the 2012 – 2016 Dec/Jan/Feb MERRA-2 dataset.

Percentile	Fraction of Lake	Fraction of Land	Fraction of Land-Ice	Fraction of Ocean	Elevation (meters)	Relative Humidity (%)	2-m Air Temperature (K)	10-m Wind Speed (m s^{-1})	Fraction Showers	Fraction Cloud Coverage	Fraction Black Carbon	Fraction Dust	Fraction Organic Carbon	Fraction Sea Salt	Fraction Sulfate	Total AOD
10	0	0	0	1.00	0	76.7	254	4.92	0.804	0.452	0.0207	0.0192	0.0712	0.143	0.174	0.0675
25	0	0	0	1.00	0	79.0	274	5.75	0.879	0.581	0.0254	0.0481	0.0865	0.388	0.222	0.0776
50	0	0	0	1.00	0	82.2	288	6.52	0.925	0.712	0.0341	0.0724	0.104	0.505	0.262	0.0971
75	0	0	0	1.00	0	88.6	298	8.01	0.959	0.794	0.0458	0.103	0.136	0.577	0.361	0.120
90	0	0	0	1.00	0	95.4	300	9.05	0.974	0.876	0.0572	0.136	0.158	0.663	0.523	0.154

Table C3. Category percentiles for the land subset of the 2012 – 2016 Dec/Jan/Feb MERRA-2 dataset.

Percentile	Fraction of Lake	Fraction of Land	Fraction of Land-Ice	Fraction of Ocean	Elevation (meters)	Relative Humidity (%)	2-m Air Temperature (K)	10-m Wind Speed (m s^{-1})	Fraction Showers	Fraction Cloud Coverage	Fraction Black Carbon	Fraction Dust	Fraction Organic Carbon	Fraction Sea Salt	Fraction Sulfate	Total AOD
10	0	0	0	0	6.67	46.5	241	2.05	0.498	0.328	0.0361	0.0704	0.0955	0.0216	0.208	0.0201
25	0	0	0	0	105	74.1	247	2.87	0.653	0.433	0.0443	0.105	0.119	0.0520	0.373	0.0370
50	0	0.831	0	0	412	88.8	262	4.17	0.784	0.581	0.0589	0.139	0.172	0.0814	0.500	0.0871
75	3.77×10^{-3}	1.00	1.00	0.0160	1476	95.9	288	5.53	0.910	0.736	0.0701	0.164	0.216	0.159	0.537	0.135
90	0.0261	1.00	1.00	0.820	2850	99.0	299	7.00	0.964	0.797	0.0780	0.402	0.258	0.307	0.602	0.207

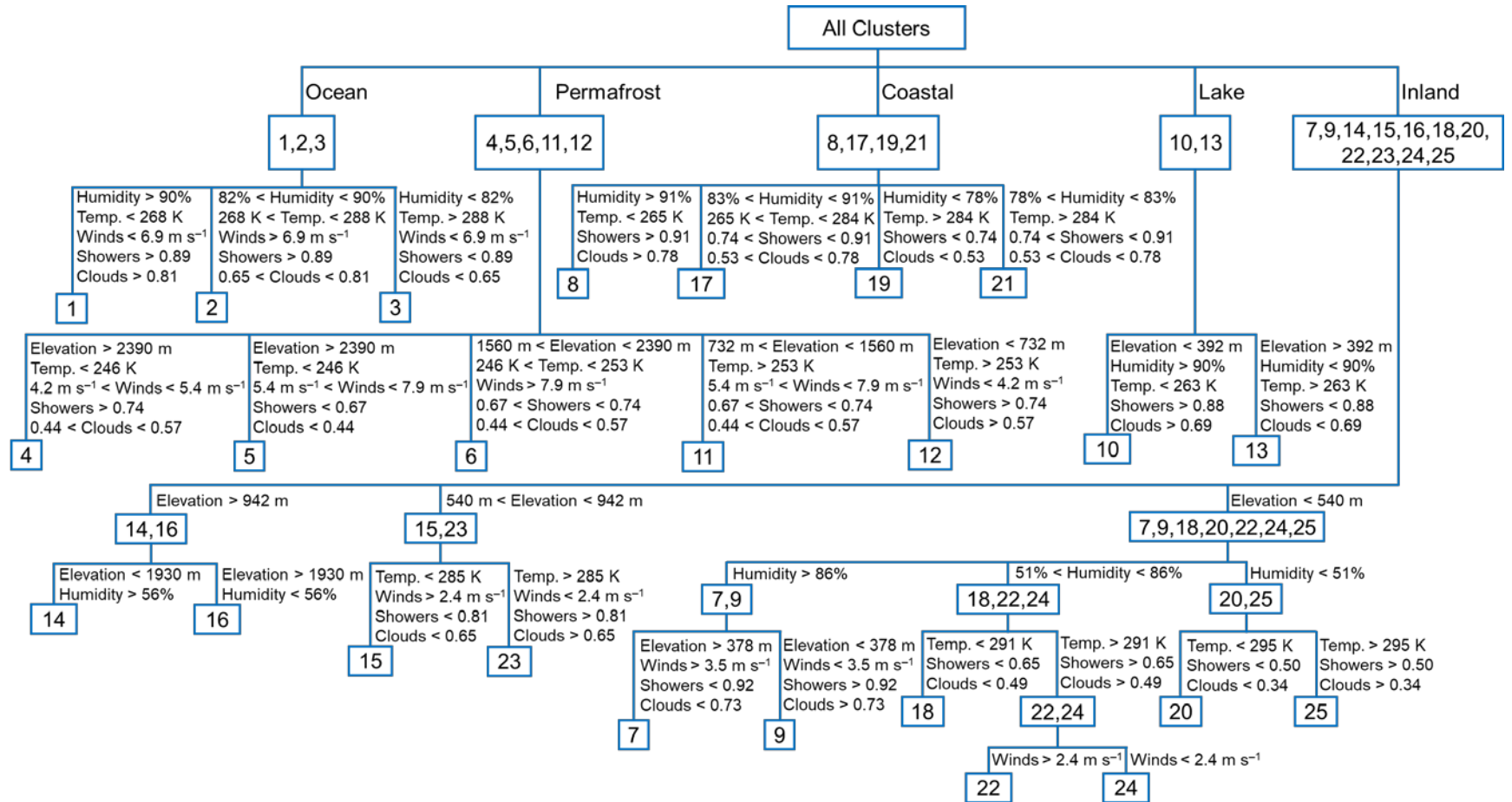


Figure C1. Decision tree for clusters describing the 2012 – 2016 Dec/Jan/Feb global MERRA-2 dataset.

Table C4. Cluster descriptions and AOD₅₅₀ statistics for the 2012 – 2016 Dec/Jan/Feb global MERRA-2 dataset.

Cluster	Qualitative Description	Characteristic Regions	AOD Centroid	AOD Average	AOD Standard Deviation	AOD 10th Percentile	AOD 50th Percentile	AOD 90th Percentile
1	Regions of the oceans: Humidity: high; Temperature: very low; Wind speed: medium low; Showers: high; Clouds coverage: very high	Arctic Ocean, Hudson Bay north to Davis Straight, region of the Southern Ocean closest to Antarctica.	0.0707	0.0707	0.0139	0.0563	0.0706	0.0836
2	Regions of the oceans: Humidity: medium high; Temperature: medium low; Wind speed: medium high; Showers: medium high; Clouds coverage: medium high	Northern Atlantic Ocean in an arc north of Norway, northern and central Pacific Ocean and Bearing Sea, Sea of Okhotsk, central Indian Ocean, Southern Ocean.	0.109	0.108	0.0277	0.0747	0.107	0.145
3	Regions of the oceans: Humidity: low; Temperature: medium high; Wind speed: medium low; Showers: low; Clouds coverage: low	Parts of the Atlantic Ocean, eastern Pacific Ocean and stretch of Pacific Ocean from the China east to North America, Pacific Ocean east of Australia, Indian Ocean excluding central region.	0.119	0.121	0.0756	0.0694	0.0966	0.201
4	Inland permafrost regions: Elevation: very high; Humidity: medium high; Temperature: very low; Wind speed: medium high; Showers: medium low; Cloud coverage: medium low	Coldest regions of central Antarctica and central Greenland with highest elevation.	0.0205	0.0205	4.02x10 ⁻³	0.0181	0.0197	0.0238
5	Inland permafrost regions: Elevation: high; Humidity: medium high; Temperature: low; Wind speed: high; Showers: low; Cloud coverage: low	Inland regions of Antarctica at high elevation. Least cloudy and rainy of the permafrost regions.	0.0201	0.0201	1.65x10 ⁻³	0.0182	0.0200	0.0224
6	Inland permafrost regions: Elevation: high; Humidity: medium low; Temperature: medium low; Wind speed: very high; Showers: medium low; Cloud coverage: medium low	Inland regions of Antarctica and Greenland at high elevation and with very high winds.	0.0293	0.0293	7.08x10 ⁻³	0.0216	0.0284	0.0380
7	Inland regions: Elevation: medium high; Humidity: high; Temperature: medium low; Wind speed: medium low; Showers: medium high; Cloud coverage: medium high	Inland regions of Alaska and Canada, eastern Russia north of the Sea of Okhotsk, the Scandinavian Peninsula, and northern France east to Poland and the Baltic states.	0.0717	0.0717	0.0238	0.0496	0.0655	0.103
8	Coastal regions: Elevation: low; Humidity: high; Temperature: medium low; Wind speed: medium high; Showers: high; Cloud coverage: very high	Northern coastal Canada, Greenland, and Russia. Includes eastern coast of Russia along the Sea of Okhotsk and the Great Lakes of North America.	0.0782	0.0781	0.0251	0.0559	0.0686	0.113
9	Inland regions: Elevation: medium low; Humidity: high; Temperature: medium low; Wind speed: medium low; Showers: high; Cloud coverage: high	Majority of inland Russia, excluding easternmost portion north of the Sea of Okhotsk. Includes eastern Ukraine and Belarus. Includes central portion of Ontario and a portion of the Andes Mountains in northern Bolivia and southern Peru.	0.127	0.127	0.0262	0.0939	0.128	0.156
10	Inland regions near lakes: Elevation: medium low; Humidity: medium high; Temperature: medium low; Wind speed: medium low; Showers: medium high; Cloud coverage: medium high	Scattered global inland regions near small lakes concentrated in Canada and Russia, but including eastern Africa, and western South America including the Amazon River.	0.0985	0.0986	0.0436	0.0577	0.0872	0.146
11	Inland permafrost regions: Elevation: medium high; Humidity: medium low; Temperature: medium low; Wind speed: high; Showers: medium low; Cloud coverage: medium low	Inland but near coastal regions of Antarctica and Greenland at medium high elevation.	0.0311	0.0310	7.14x10 ⁻³	0.0245	0.0287	0.0415
12	Inland permafrost regions: Elevation: medium low; Humidity: medium high; Temperature: medium low; Wind speed: medium low; Showers: medium low; Cloud coverage: medium high	Inland but near coastal regions of Antarctica and Greenland. Warmest and lowest elevation of the permafrost regions.	0.0389	0.0389	9.26x10 ⁻³	0.0292	0.0378	0.0485
13	Inland regions near lakes: Elevation: medium high; Humidity: medium low; Temperature: medium high; Wind speed: medium low; Showers: medium high; Cloud coverage: medium high	Scattered global inland regions near large lakes concentrated in western and northern Canada, western and southern Russia, and eastern Africa.	0.129	0.129	0.0559	0.0614	0.115	0.203

UNCLASSIFIED

14	Inland regions: Elevation: medium high; Humidity: low; Temperature: medium high; Wind speed: medium low; Showers: low; Cloud coverage: low	Inland northern Africa east through Israel, Jordan, Syria, and Iraq north to Azerbaijan. Extends east through most of Iran and north to Uzbekistan, then further East through northwestern China, including the Taklamakan and Gobi Deserts, Inner Mongolia in China, and southern Mongolia.	0.147	0.148	0.0525	0.0807	0.143	0.223
15	Inland regions: Elevation: medium high; Humidity: medium low; Temperature: medium high; Wind speed: medium low; Showers: medium low; Cloud coverage: medium low	Majority of the inland eastern United States and northeastern Mexico as far west as central Kansas, north western United States and south central Canada, parts of Alaska, southern and central Europe east to the Black Sea, Turkey, southern India, southeastern Africa, and portions of the Andes mountains.	0.102	0.103	0.0476	0.0517	0.0990	0.154
16	Inland regions: Elevation: high; Humidity: low; Temperature: medium high; Wind speed: medium low; Showers: low; Cloud coverage: low	Arid inland regions of the southwestern United States and north central Mexico, Tibet and portions of southern and central China, central South Africa, portion of the Andes mountains.	0.0734	0.0727	0.0435	0.0371	0.0568	0.136
17	Coastal regions: Elevation: low; Humidity: medium low; Temperature: medium high; Wind speed: high; Showers: medium high; Cloud coverage: medium high	Coastal western Canada and southern Alaska, coastal eastern Canada and southern Greenland, northern coastal Europe, including around the Baltic Sea and north of Norway and western Russia, coastal Malaysia, Indonesia, and Papua New Guinea, coastal western equatorial South America, coastal Antarctica.	0.100	0.100	0.0445	0.0506	0.0990	0.152
18	Inland regions: Elevation: medium low; Humidity: low; Temperature: medium high; Wind speed: low; Showers: very low; Cloud coverage: low	Inland eastern China, Korea, and most of India excluding the southern tip.	0.430	0.429	0.160	0.247	0.402	0.631
19	Coastal regions: Elevation: low; Humidity: medium low; Temperature: high; Wind speed: medium high; Showers: low; Cloud coverage: low	Coastal regions in the mid-latitudes. Coastal United States and Mexico, Mediterranean coast, coasts of the Black and Caspian Seas, coastal Arabian peninsula and India, coastal China, Korea, and Japan, coastal southern South America, Africa, and Australia.	0.173	0.173	0.0948	0.0873	0.144	0.289
20	Inland regions: Elevation: medium high; Humidity: very low; Temperature: high; Wind speed: medium low; Showers: very low; Cloud coverage: very low	Majority of the Sahara desert and Arabian Peninsula. Includes some portions of Southern Iran and center of Australia.	0.244	0.244	0.102	0.146	0.215	0.377
21	Coastal regions: Elevation: low; Humidity: medium low; Temperature: high; Wind speed: high; Showers: medium high; Cloud coverage: medium high	Coastal and island regions concentrated in the tropics. Caribbean islands and coastal northern and eastern South America, coastal Madagascar and east Africa, majority of the Philippine islands and northern coastal Indonesia. Includes coastal British Isles and Iceland, Alaskan archipelago, and southern tip of South America.	0.121	0.121	0.0332	0.0854	0.115	0.165
22	Inland regions: Elevation: medium low; Humidity: medium low; Temperature: high; Wind speed: medium low; Showers: medium high; Cloud coverage: medium high	Inland central America and southern Mexico, inland British Isles and Iceland, inland but near coastal regions of Colombia south through Brazil, inland but near coastal eastern Africa, inland Madagascar, inland northern and eastern Australia, inland New Zealand, inland Papua New Guinea, and portions of inland Indonesia and Malaysia.	0.132	0.131	0.0455	0.0802	0.124	0.192
23	Inland regions: Elevation: medium high; Humidity: medium low; Temperature: high; Wind speed: very low; Showers: medium high; Cloud coverage: high	Equatorial portions of South America, Africa, and southeast Asia that include jungle and tropical rainforests.	0.146	0.146	0.0515	0.0971	0.133	0.213
24	Inland regions: Elevation: medium high; Humidity: low; Temperature: high; Wind speed: very low; Showers: medium high; Cloud coverage: medium high	Inland regions of Africa south of the Sahara to the center of the Democratic Republic of the Congo	0.493	0.492	0.159	0.317	0.464	0.746
25	Inland regions: Elevation: medium high; Humidity: very low; Temperature: high; Wind speed: medium low; Showers: low; Cloud coverage: low	Inland Australia, inland portion of the Horn of Africa, inland southwestern Africa, inland Argentina and southern Chile, land around the Gulf of California.	0.108	0.108	0.0303	0.0761	0.102	0.140

Blue = Ocean; Orange = Permafrost; Red = Inland; Green = Lake; Purple = Coast

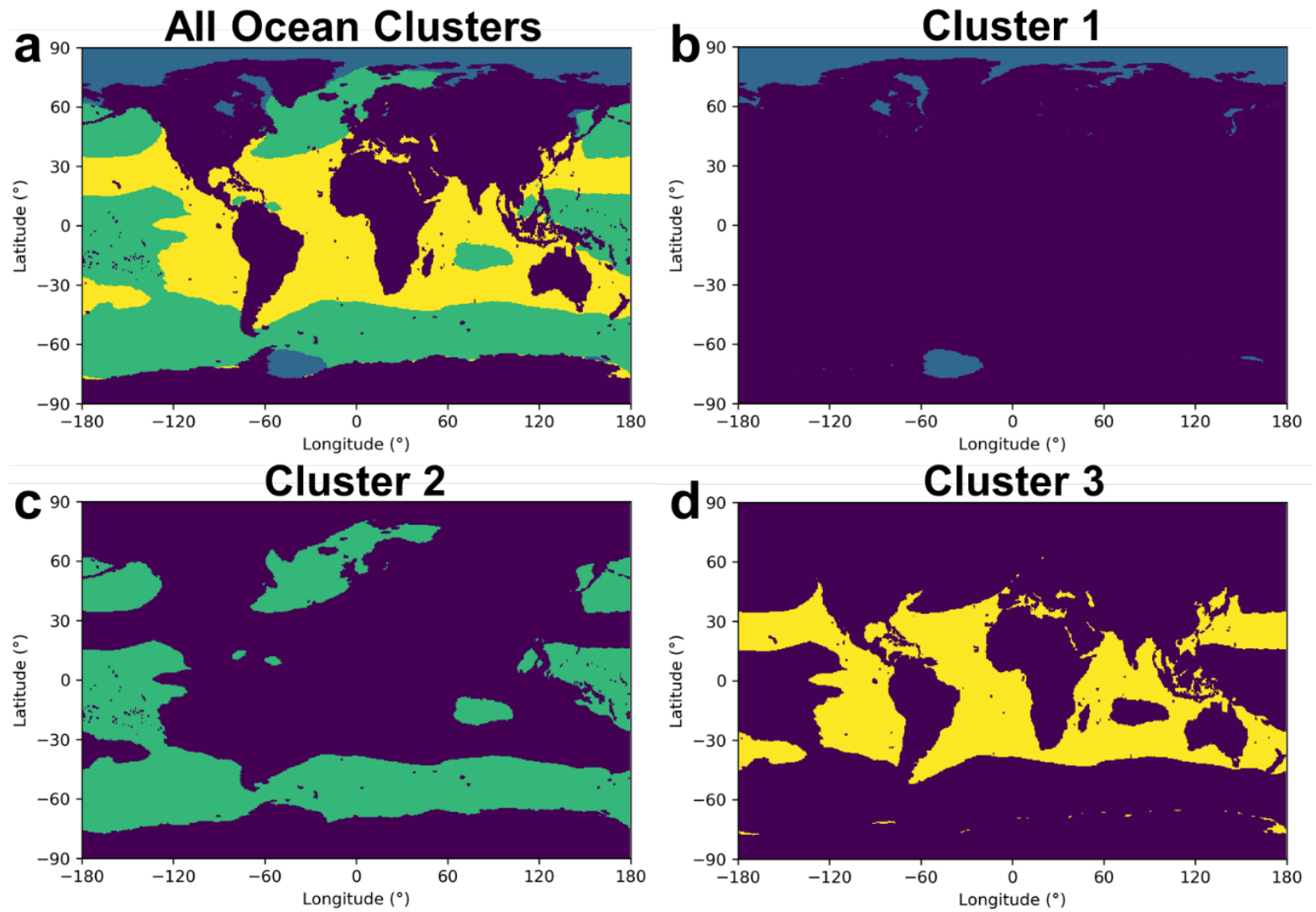


Figure C2. Ocean clusters 1 – 3 from the 2012 – 2016 Dec/Jan/Feb global MERRA-2 dataset.

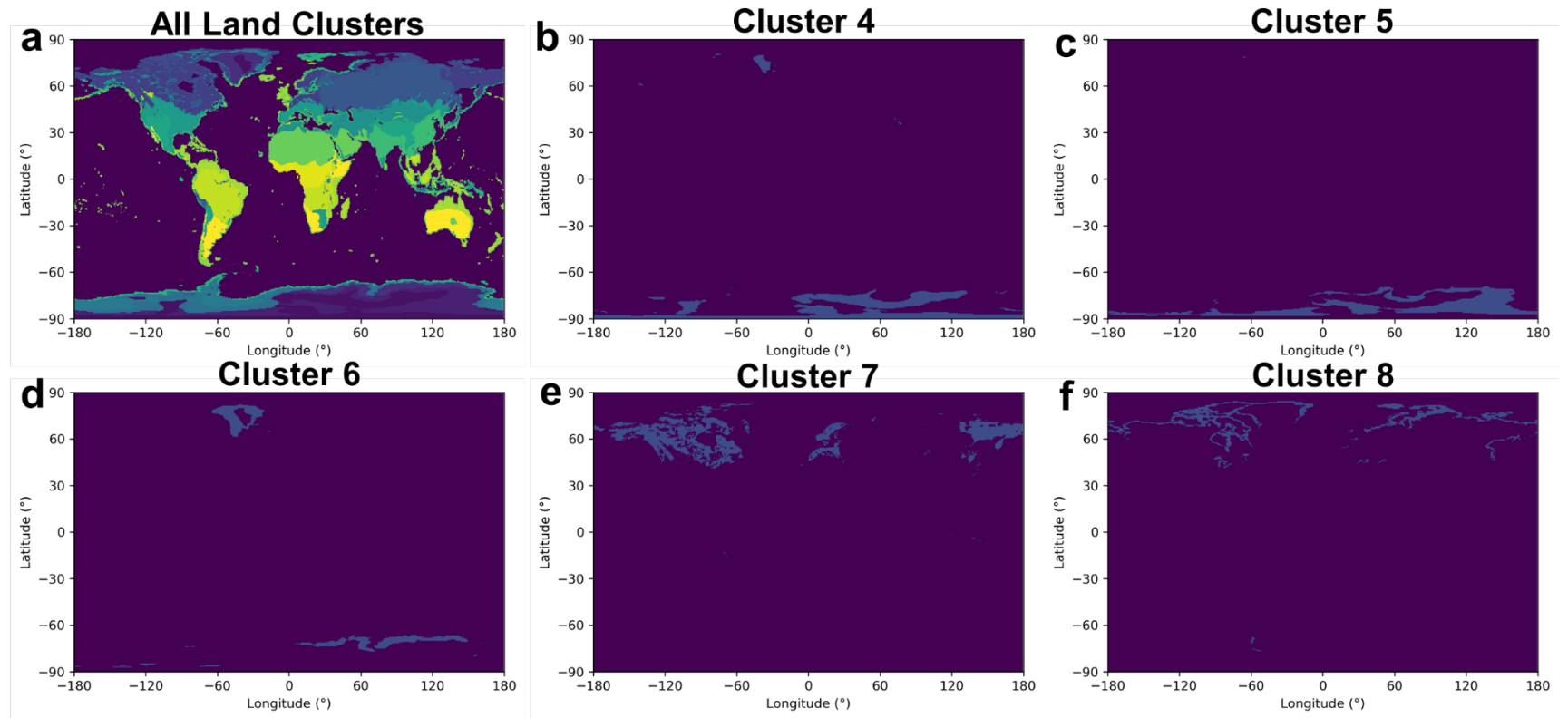


Figure C3. Land clusters 4 – 8 from the 2012 – 2016 Dec/Jan/Feb global MERRA-2 dataset.

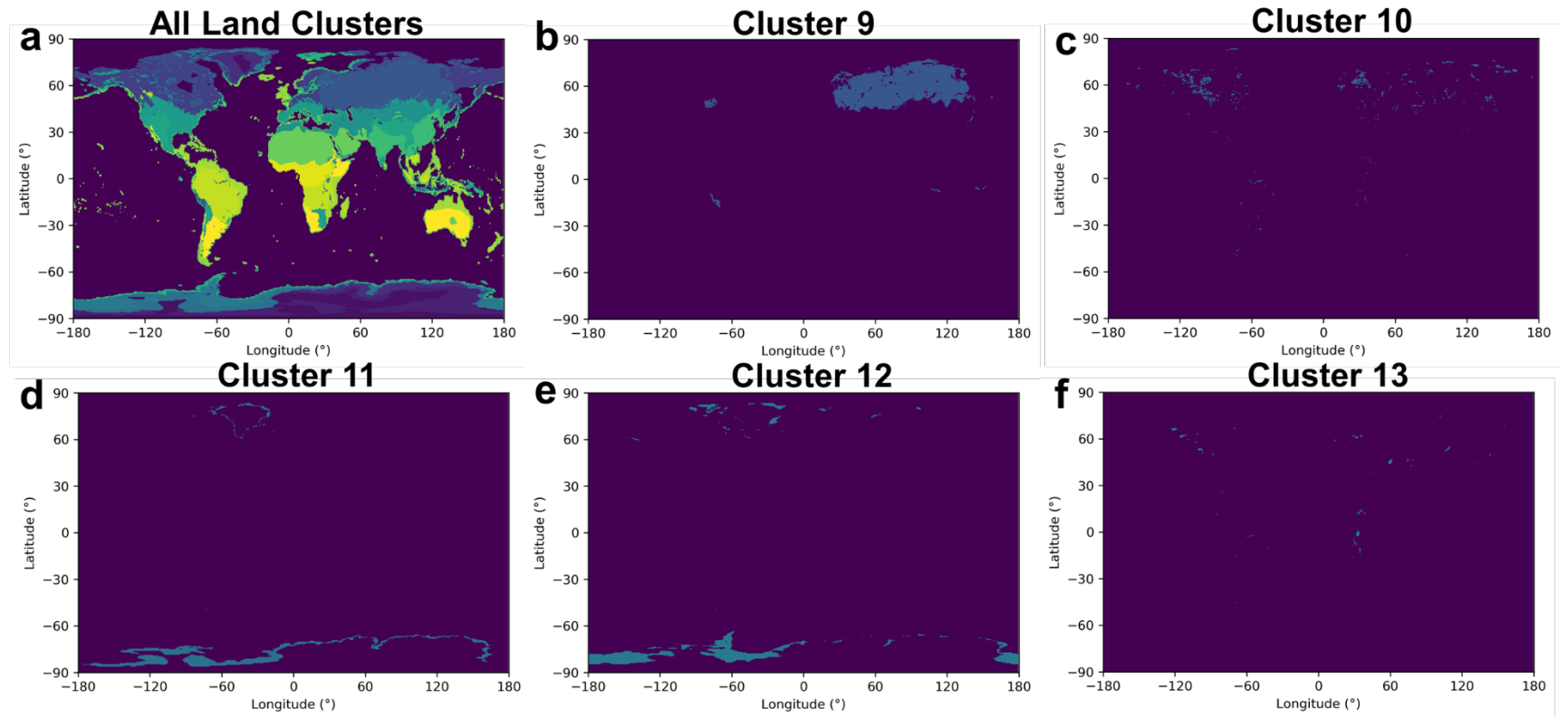


Figure C4. Land clusters 9 – 13 from the 2012 – 2016 Dec/Jan/Feb global MERRA-2 dataset.

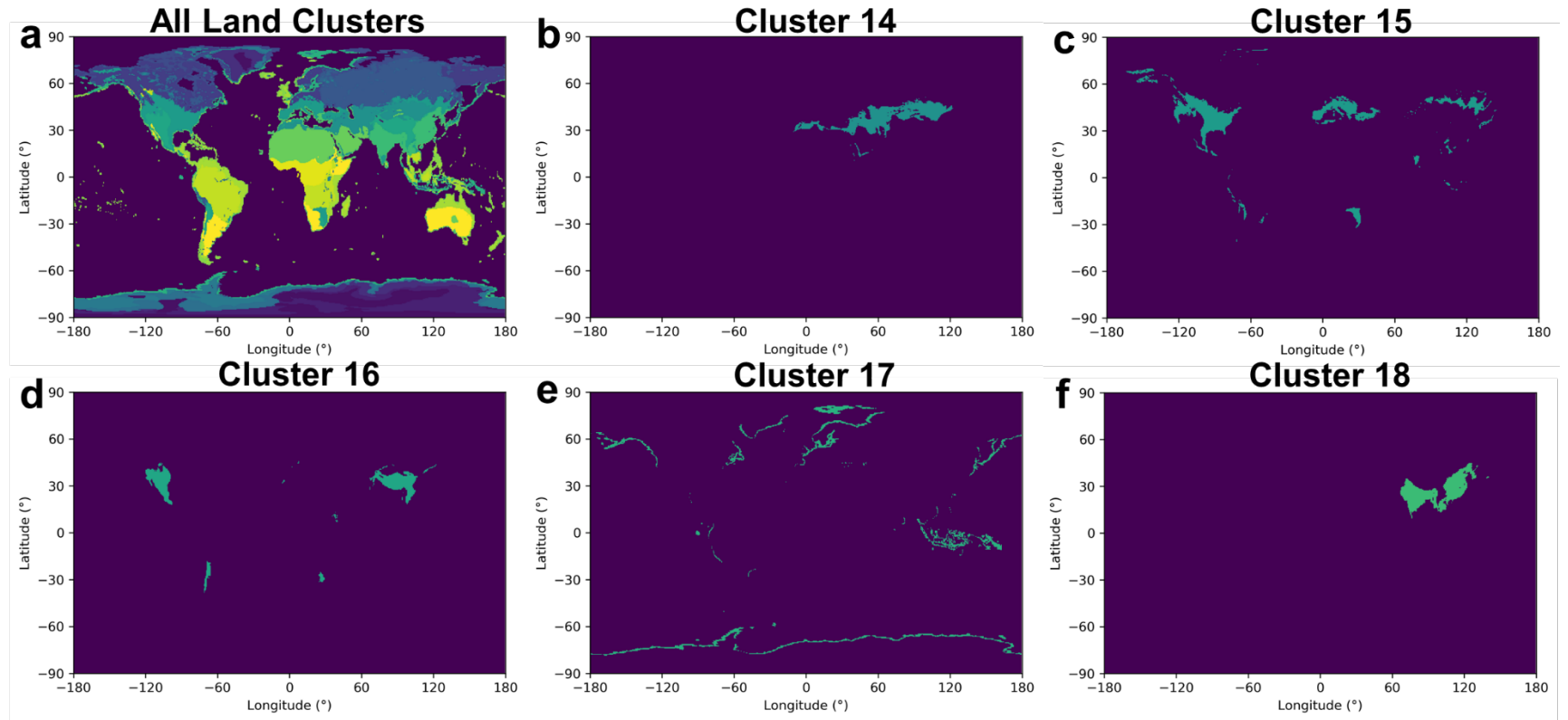


Figure C5. Land clusters 14 – 18 from the 2012 – 2016 Dec/Jan/Feb global MERRA-2 dataset.

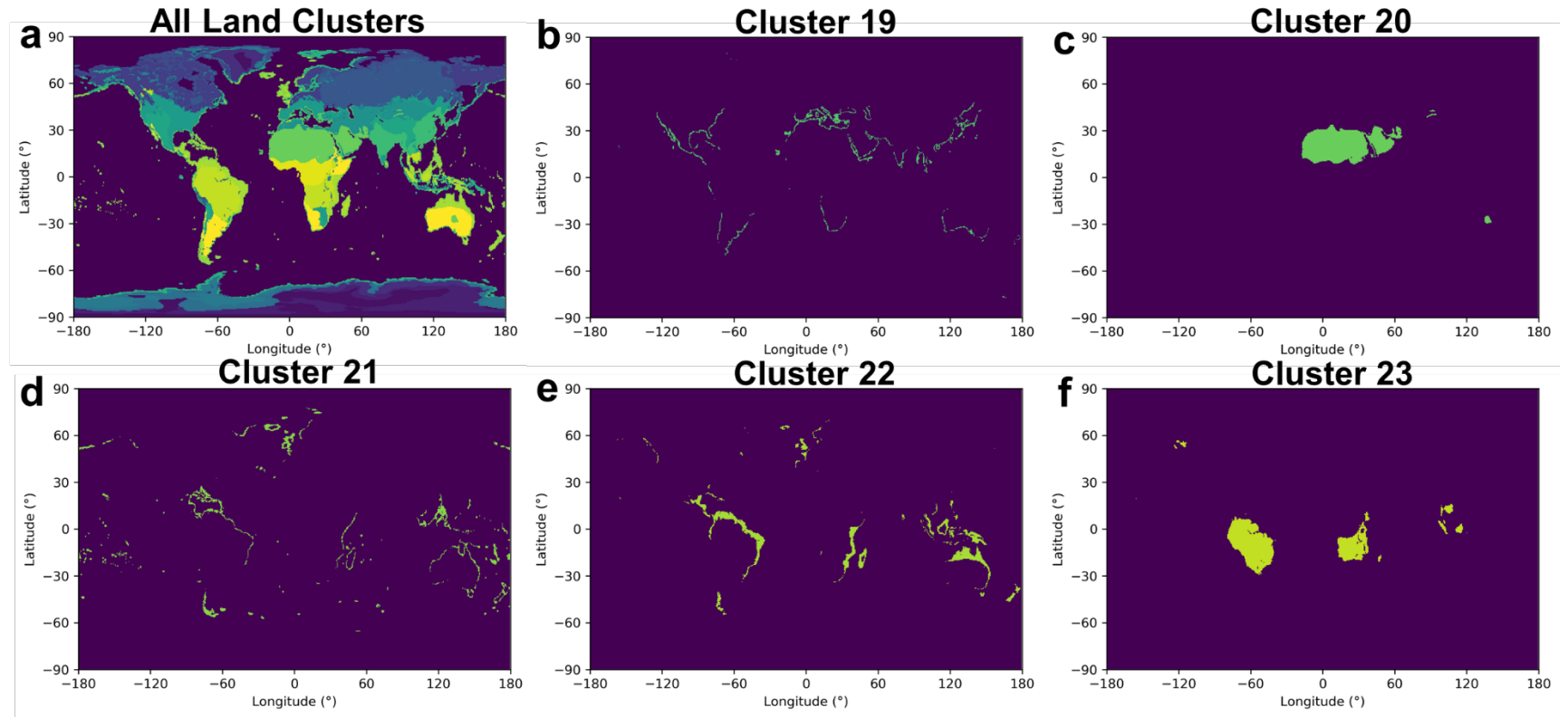


Figure C6. Land clusters 19 – 23 from the 2012 – 2016 Dec/Jan/Feb global MERRA-2 dataset.

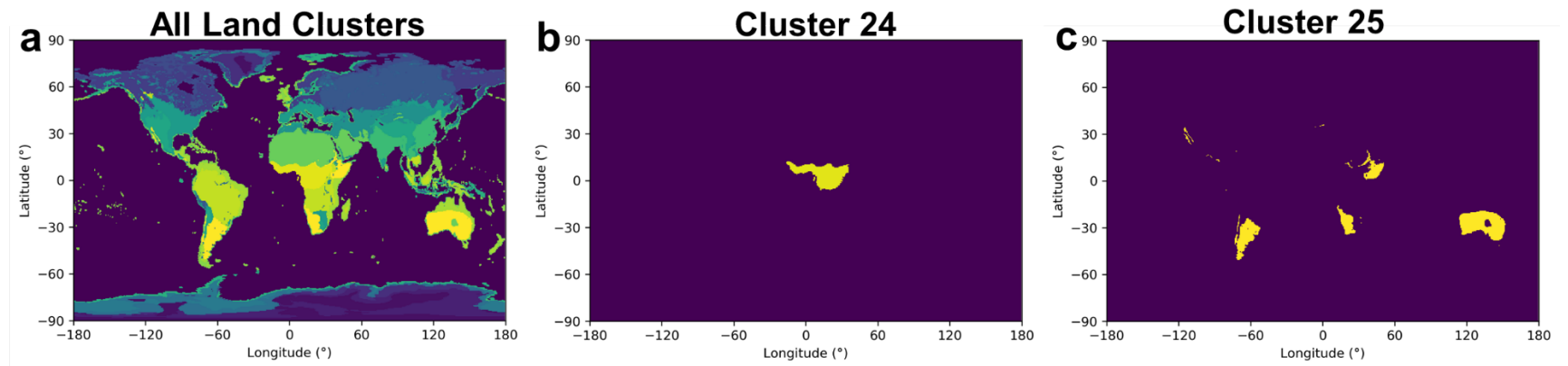


Figure C7. Land clusters 24 – 25 from the 2012 – 2016 Dec/Jan/Feb global MERRA-2 dataset.

D. 2012 to 2016 March through May Global Clusters

Table D1. Cluster centroids for the 2012 – 2016 Mar/Apr/May global MERRA-2 dataset.

Cluster	Fraction of Lake	Fraction of Land	Fraction of Land-Ice	Fraction of Ocean	Elevation (meters)	Relative Humidity (%)	2-m Air Temperature (K)	10-m Wind Speed (m s ⁻¹)	Fraction Showers	Fraction Cloud Coverage	Fraction Black Carbon	Fraction Dust	Fraction Organic Carbon	Fraction Sea Salt	Fraction Sulfate	Total AOD
1	3.62x10 ⁻⁹	2.87x10 ⁻⁷	2.06x10 ⁻⁷	1.00	0.0340	93.4	260	5.63	0.935	0.834	0.0493	0.202	0.217	0.103	0.429	0.116
2	0	9.01x10 ⁻⁸	1.14x10 ⁻⁸	1.00	1.25x10 ⁻⁴	83.3	287	7.51	0.926	0.703	0.0206	0.0555	0.0707	0.664	0.189	0.0864
3	1.24x10 ⁻⁹	2.19x10 ⁻⁷	1.18x10 ⁻⁹	1.00	-0.0445	80.1	293	5.97	0.836	0.597	0.0442	0.191	0.165	0.332	0.267	0.161
4	0	0	1.00	0	3082	99.2	220	6.20	0.948	0.686	0.0660	0.127	0.231	0.108	0.468	0.0143
5	0	0	1.00	0	2582	95.7	224	9.05	0.846	0.587	0.0668	0.137	0.215	0.125	0.457	0.0127
6	1.04x10 ⁻⁷	8.46x10 ⁻⁸	0.994	5.98x10 ⁻³	1494	80.9	239	12.0	0.701	0.538	0.0608	0.145	0.190	0.180	0.425	0.0156
7	3.54x10 ⁻⁴	3.08x10 ⁻³	0.967	0.0290	506	89.9	246	5.90	0.837	0.687	0.0496	0.125	0.172	0.234	0.419	0.0212
8	3.24x10 ⁻³	0.0596	0.879	0.0584	1417	87.6	251	5.61	0.713	0.525	0.0616	0.236	0.239	0.0611	0.403	0.0671
9	5.05x10 ⁻³	0.141	0.0516	0.803	33.4	93.0	263	5.03	0.909	0.761	0.0510	0.212	0.219	0.0757	0.443	0.126
10	7.68x10 ⁻⁴	0.0188	0.176	0.805	38.5	89.4	264	7.01	0.877	0.760	0.0379	0.112	0.131	0.385	0.333	0.059
11	0.0186	0.934	0.0114	0.0353	408	92.1	267	3.18	0.893	0.690	0.0508	0.224	0.231	0.0402	0.454	0.135
12	0.0120	0.973	0.0152	0	3943	49.3	273	3.34	0.711	0.560	0.0383	0.492	0.159	0.0170	0.295	0.140
13	0.182	0.790	5.33x10 ⁻³	0.0226	307	85.1	273	3.30	0.850	0.647	0.0511	0.238	0.222	0.0577	0.431	0.140
14	9.25x10 ⁻³	0.955	5.11x10 ⁻⁴	0.0353	391	77.0	281	3.39	0.779	0.591	0.0483	0.242	0.157	0.0512	0.502	0.193
15	0.615	0.381	8.08x10 ⁻⁶	3.95x10 ⁻³	485	72.3	282	3.85	0.768	0.584	0.0512	0.315	0.221	0.0599	0.354	0.157
16	7.16x10 ⁻³	0.981	2.42x10 ⁻³	8.68x10 ⁻³	703	68.5	284	3.09	0.700	0.519	0.0559	0.215	0.262	0.0588	0.408	0.154
17	4.12x10 ⁻³	0.142	0.0116	0.842	31.9	79.9	289	4.63	0.727	0.548	0.0512	0.197	0.185	0.190	0.377	0.195
18	3.88x10 ⁻³	0.962	3.64x10 ⁻⁵	0.0336	788	43.9	289	3.55	0.540	0.387	0.0373	0.525	0.112	0.0498	0.277	0.236
19	2.50x10 ⁻³	0.977	1.46x10 ⁻⁴	0.0200	1263	43.0	291	3.15	0.475	0.323	0.0683	0.208	0.241	0.0841	0.399	0.113
20	9.83x10 ⁻³	0.921	9.38x10 ⁻⁶	0.0690	423	63.2	292	2.70	0.682	0.501	0.0683	0.176	0.254	0.0382	0.464	0.568
21	5.33x10 ⁻³	0.167	3.18x10 ⁻³	0.825	30.3	80.7	295	5.38	0.879	0.667	0.0245	0.0744	0.0977	0.535	0.267	0.103
22	6.08x10 ⁻³	0.947	7.04x10 ⁻⁴	0.0458	307	63.9	296	3.34	0.719	0.518	0.0410	0.191	0.180	0.308	0.279	0.0945
23	0.0120	0.975	2.84x10 ⁻⁶	0.0123	513	80.3	298	1.31	0.885	0.723	0.0578	0.182	0.381	0.118	0.261	0.139
24	2.34x10 ⁻³	0.957	0	0.0404	488	23.9	299	3.74	0.431	0.295	0.0212	0.793	0.0553	0.0224	0.109	0.394
25	8.76x10 ⁻³	0.959	1.29x10 ⁻⁵	0.0323	454	50.6	302	2.69	0.857	0.670	0.0311	0.704	0.147	0.0338	0.085	0.425

Blue = Ocean; Orange = Permafrost; Red = Inland; Green = Lake; Purple = Coast

Table D2. Category percentiles for the ocean subset of the 2012 – 2016 Mar/Apr/May MERRA-2 dataset.

Percentile	Fraction of Lake	Fraction of Land	Fraction of Land-Ice	Fraction of Ocean	Elevation (meters)	Relative Humidity (%)	2-m Air Temperature (K)	10-m Wind Speed (m s ⁻¹)	Fraction Showers	Fraction Cloud Coverage	Fraction Black Carbon	Fraction Dust	Fraction Organic Carbon	Fraction Sea Salt	Fraction Sulfate	Total AOD
10	0	0	0	1.00	0	77.0	259	4.75	0.811	0.480	0.0149	0.0341	0.0543	0.0508	0.142	0.0646
25	0	0	0	1.00	0	79.0	274	5.46	0.877	0.594	0.0197	0.0501	0.0681	0.256	0.172	0.0792
50	0	0	0	1.00	0	82.6	288	6.51	0.918	0.712	0.0313	0.0861	0.106	0.529	0.222	0.105
75	0	0	0	1.00	0	87.3	298	7.73	0.949	0.793	0.0469	0.201	0.180	0.675	0.322	0.127
90	0	0	0	1.00	0	94.0	301	9.22	0.967	0.858	0.0537	0.240	0.237	0.732	0.442	0.187

Table D3. Category percentiles for the land subset of the 2012 – 2016 Mar/Apr/May MERRA-2 dataset.

Percentile	Fraction of Lake	Fraction of Land	Fraction of Land-Ice	Fraction of Ocean	Elevation (meters)	Relative Humidity (%)	2-m Air Temperature (K)	10-m Wind Speed (m s ⁻¹)	Fraction Showers	Fraction Cloud Coverage	Fraction Black Carbon	Fraction Dust	Fraction Organic Carbon	Fraction Sea Salt	Fraction Sulfate	Total AOD
10	0	0	0	0	6.67	40.2	220	2.05	0.530	0.366	0.0288	0.108	0.095	0.0150	0.193	0.0127
25	0	0	0	0	105	68.6	249	2.89	0.695	0.511	0.0436	0.128	0.158	0.0287	0.322	0.0217
50	0	0.831	0	0	412	85.5	272	4.19	0.828	0.629	0.0529	0.202	0.208	0.0872	0.429	0.108
75	3.77×10 ⁻³	1.00	1.00	0.0160	1476	95.0	292	5.79	0.906	0.708	0.0625	0.254	0.235	0.153	0.465	0.182
90	0.0261	1.00	1.00	0.820	2850	98.1	300	8.53	0.947	0.758	0.0683	0.553	0.278	0.277	0.483	0.334

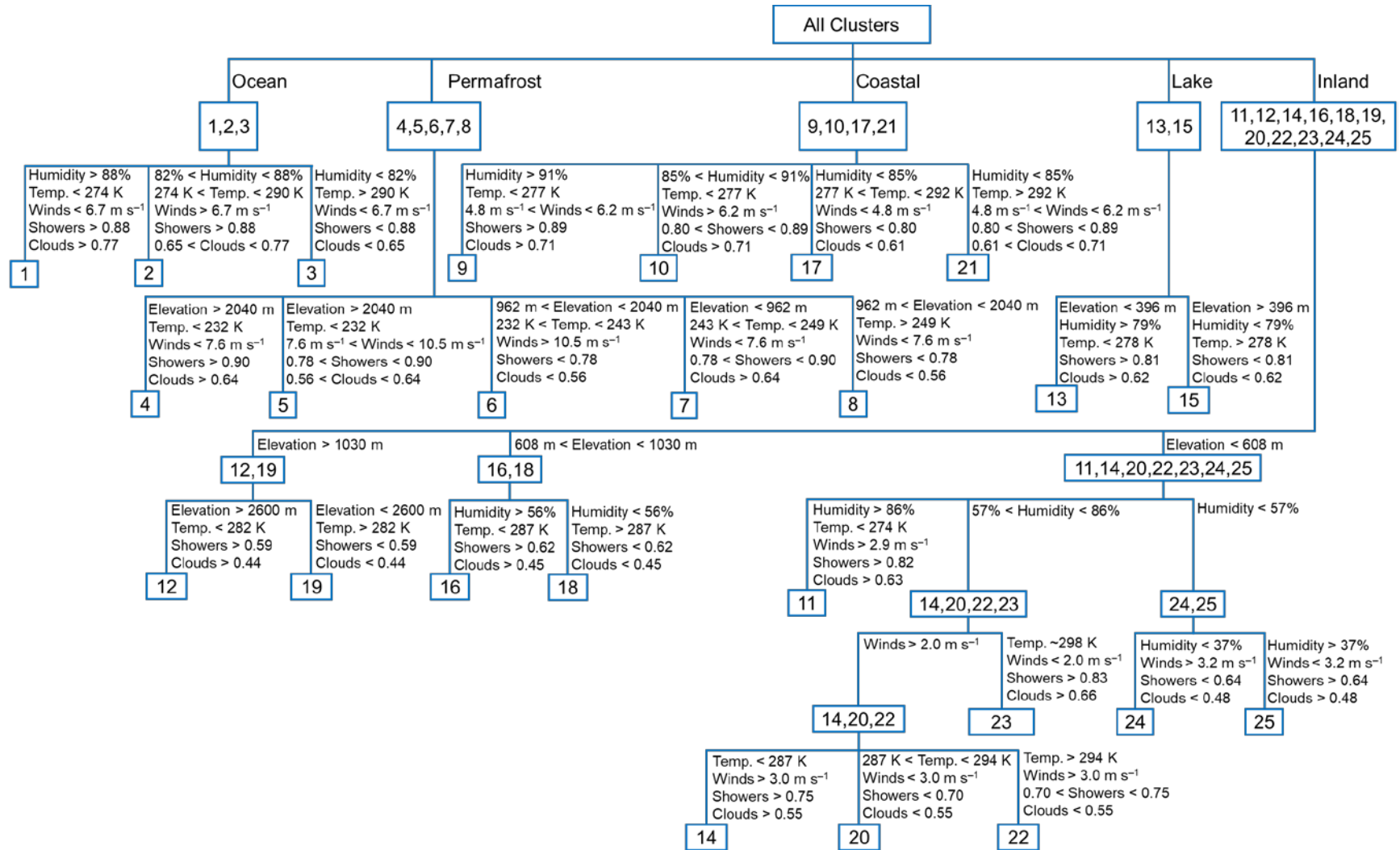


Figure D1. Decision tree for clusters describing the 2012 – 2016 Mar/Apr/May global MERRA-2 dataset.

Table D4. Cluster descriptions and AOD₅₅₀ statistics for the 2012 – 2016 Mar/Apr/May global MERRA-2 dataset.

Cluster	Qualitative Description	Characteristic Regions	AOD Centroid	AOD Average	AOD Standard Deviation	AOD 10th Percentile	AOD 50th Percentile	AOD 90th Percentile
1	Regions of the oceans: Humidity: high; Temperature: low; Wind speed: medium low; Showers: medium high; Clouds coverage: high	The Arctic Ocean including Hudson Bay north through the Davis Strait, the Bearing Sea, and the Sea of Okhotsk. Portions of the Southern Ocean nearest Antarctica.	0.116	0.119	0.0543	0.0349	0.115	0.180
2	Regions of the oceans: Humidity: medium high; Temperature: medium low; Wind speed: medium high; Showers: medium high; Clouds coverage: medium low	Southern Pacific Ocean, southern Atlantic Ocean, southern Indian Ocean, and the majority of the Southern Ocean.	0.0864	0.0868	0.0216	0.0617	0.0854	0.113
3	Regions of the oceans: Humidity: medium low; Temperature: medium high; Wind speed: medium low; Showers: low; Clouds coverage: medium low	Northern Pacific Ocean, central and northern Atlantic Ocean including the Norwegian Sea, Mediterranean, Black, Caspian, and Red Seas, and northern Indian Ocean.	0.161	0.162	0.0806	0.0824	0.142	0.265
4	Inland permafrost region: Elevation: very high; Humidity: very high; Temperature: very low; Wind speed: high; Showers: very high; Cloud coverage: medium high	Inland permafrost regions of Antarctica and Greenland at very high elevation and with very cold temperatures. Rainiest of the permafrost regions.	0.0143	0.0144	9.68×10^{-3}	0.0110	0.0125	0.0140
5	Inland permafrost region: Elevation: high; Humidity: high; Temperature: low; Wind speed: very high; Showers: medium high; Cloud coverage: medium low	Inland permafrost regions of Antarctica at high elevation. Second highest winds of all regions.	0.0127	0.0128	3.18×10^{-3}	0.0112	0.0125	0.0142
6	Inland permafrost region: Elevation: high; Humidity: medium low; Temperature: low; Wind speed: very high; Showers: medium low; Cloud coverage: medium low	Inland permafrost regions of Antarctica at high elevation with the highest winds of all regions.	0.0156	0.0156	4.32×10^{-3}	0.0123	0.0148	0.0193
7	Inland permafrost region: Elevation: medium high; Humidity: medium high; Temperature: low; Wind speed: high; Showers: medium high; Cloud coverage: medium high	Inland permafrost regions of Antarctica nearest the coast at the lowest elevation for the permafrost regions. Cloudiest of the permafrost regions.	0.0212	0.0213	0.0127	0.0162	0.0190	0.0242
8	Inland permafrost region: Elevation: medium high; Humidity: medium high; Temperature: medium low; Wind speed: medium high; Showers: medium low; Cloud coverage: medium low	Inland permafrost regions of Greenland with the calmest winds and lowest cloud coverage of the permafrost regions.	0.0671	0.0673	0.0210	0.0159	0.0697	0.0882
9	Coastal regions: Elevation: low; Humidity: medium high; Temperature: medium low; Wind speed: medium high; Showers: high; Cloud coverage: very high	Coastal regions of Alaska and northern Canada, and northern and eastern Russia. Includes portions of coastal Indonesia and a strip of coastal Antarctica south of Argentina.	0.126	0.128	0.0483	0.0966	0.116	0.172
10	Coastal regions: Elevation: low; Humidity: medium high; Temperature: medium low; Wind speed: high; Showers: medium high; Cloud coverage: very high	Coastal Antarctica, coastal Iceland, northern coast of the British Isles and Norway, scattered coastal areas of the Philippines and Indonesia.	0.059	0.0643	0.0526	0.0231	0.0373	0.142
11	Inland regions: Elevation: medium low; Humidity: medium high; Temperature: medium low; Wind speed: medium low; Showers: medium high; Cloud coverage: medium high	Inland Alaska and the majority of Canada, inland Scandinavian peninsula, inland northern Russia, strip of land east of the Andes mountains, some inland parts of Indonesia and Papua New Guinea.	0.135	0.135	0.0370	0.0985	0.130	0.178
12	Inland regions: Elevation: very high; Humidity: low; Temperature: medium high; Wind speed: medium low; Showers: medium low; Cloud coverage: medium low	Tibet and other high elevation land in western China. Small portions of central Ethiopia and western Bolivia/northern Chile.	0.140	0.140	0.0454	0.0922	0.128	0.202
13	Inland regions near lakes: Elevation: medium low; Humidity: medium low; Temperature: medium high; Wind speed: medium low; Showers: medium high; Cloud coverage: medium high	Scattered inland areas near small lakes concentrated in central and northern Canada, Finland, and northeastern Russia. Includes scattered regions in northern Asia, eastern Africa, and northern and eastern South America.	0.140	0.141	0.0644	0.0978	0.129	0.204

UNCLASSIFIED

14	Inland regions: Elevation: medium low; Humidity: medium low; Temperature: medium high; Wind speed: medium low; Showers: medium low; Cloud coverage: medium low	Inland British Isles and France east through southern Russia (as far north as St. Petersburg) until the western edge of Mongolia. Includes Italy and land around the Black Sea. Also includes land around the Great Lakes, inland Japan, parts of the Andes Mountains, and parts of inland Indonesia.	0.193	0.193	0.0493	0.142	0.196	0.237
15	Inland regions near lakes: Elevation: medium high; Humidity: medium low; Temperature: medium high; Wind speed: medium low; Showers: medium low; Cloud coverage: medium low	Scattered inland areas near large lakes concentrated in central and northern Canada, eastern Africa, and southern Russia and Kazakhstan.	0.157	0.158	0.0782	0.0812	0.135	0.278
16	Inland regions: Elevation: medium high; Humidity: low; Temperature: medium high; Wind speed: medium low; Showers: medium low; Cloud coverage: medium low	Inland United States excluding the Great Lakes arid southwest regions, portions of central Canada, Argentina into southern Brazil, inland Spain, portion of southeastern Russia/northern China, small area of eastern Australia and south central Africa.	0.154	0.154	0.0686	0.0844	0.133	0.245
17	Coastal regions: Elevation: low; Humidity: medium low; Temperature: medium high; Wind speed: medium high; Showers: medium low; Cloud coverage: medium low	Coastal regions concentrated in the mid-latitudes. Includes nearly all of the west coast of the Americas, the east coast of the United States and Mexico, the coastal Caribbean, coastal Argentina to southern Brazil, coastal Europe including the coastal Mediterranean, Black, and Caspian Seas, coastal India, China, Japan, & Korea, coastal Philippines and southeast Asia, portions of the coasts of Australia and sub-Saharan Africa.	0.195	0.199	0.104	0.101	0.164	0.357
18	Inland regions: Elevation: medium high; Humidity: low; Temperature: medium high; Wind speed: medium low; Showers: low; Cloud coverage: low	Central Australia, parts of Yemen, Djibouti, and northern Somalia, Mediterranean north Africa west to the southern Caspian Sea including Israel, Syria, and northern Iran, central and northern India, areas east of the Caspian Sea into southern Mongolia and northwestern China.	0.236	0.237	0.0968	0.0799	0.241	0.359
19	Inland regions: Elevation: medium high; Humidity: low; Temperature: medium high; Wind speed: medium low; Showers: very low; Cloud coverage: very low	Arid portions of the southwestern United States and central Mexico, northern Chile and northwest Argentina, inland South Africa north through Namibia and Zimbabwe.	0.113	0.113	0.0679	0.0632	0.0991	0.172
20	Inland regions: Elevation: medium high; Humidity: low; Temperature: medium high; Wind speed: low; Showers: low; Cloud coverage: low	Eastern India and eastern China, Korea. Includes Myanmar, northern Thailand, Laos, and northern Vietnam.	0.568	0.570	0.139	0.410	0.546	0.782
21	Coastal regions: Elevation: low; Humidity: medium low; Temperature: high; Wind speed: medium high; Showers: medium high; Cloud coverage: medium high	Islands and equatorial coastal areas. Includes the Caribbean, coastal north, eastern, and southern South America, the east coast of Africa & Madagascar, the northern and southern coasts of Australia, coastal New Zealand, and coastal Philippines, Indonesia, and Papua New Guinea.	0.103	0.102	0.0453	0.0634	0.0929	0.139
22	Inland regions: Elevation: medium low; Humidity: low Temperature: high; Wind speed: medium low; Showers: medium low; Cloud coverage: medium low	Inland Central America and Cuba, near coastal but inland South America from Colombia east and south through Argentina, near coastal but inland east Africa, inland Madagascar, inland New Zealand, and inland Australia excluding the central region.	0.0945	0.0948	0.0484	0.0592	0.0801	0.143
23	Inland regions: Elevation: medium high; Humidity: medium low; Temperature: high; Wind speed: very low; Showers: medium high; Cloud coverage: high	Inland equatorial South America and Africa, including parts of southeast Asia and Malaysia/Indonesia and Central America. Regions associated with tropical rainforest and jungle.	0.139	0.139	0.0874	0.0699	0.0979	0.250
24	Inland regions: Elevation: medium high; Humidity: very low; Temperature: high; Wind speed: medium low; Showers: very low; Cloud coverage: very low	The Sahara desert and majority of the Arabian Peninsula, Iraq east through southern Iran and Pakistan into northwest India. Includes the Taklamakan desert in China.	0.394	0.395	0.0900	0.294	0.378	0.527
25	Inland regions: Elevation: medium high; Humidity: low; Temperature: very high; Wind speed: low; Showers: medium high; Cloud coverage: medium high	Region of sub-Saharan Africa just north of the equator from Guinea and Sierra Leone east to western Ethiopia.	0.425	0.425	0.106	0.286	0.419	0.575

Blue = Ocean; Orange = Permafrost; Red = Inland; Green = Lake; Purple = Coast

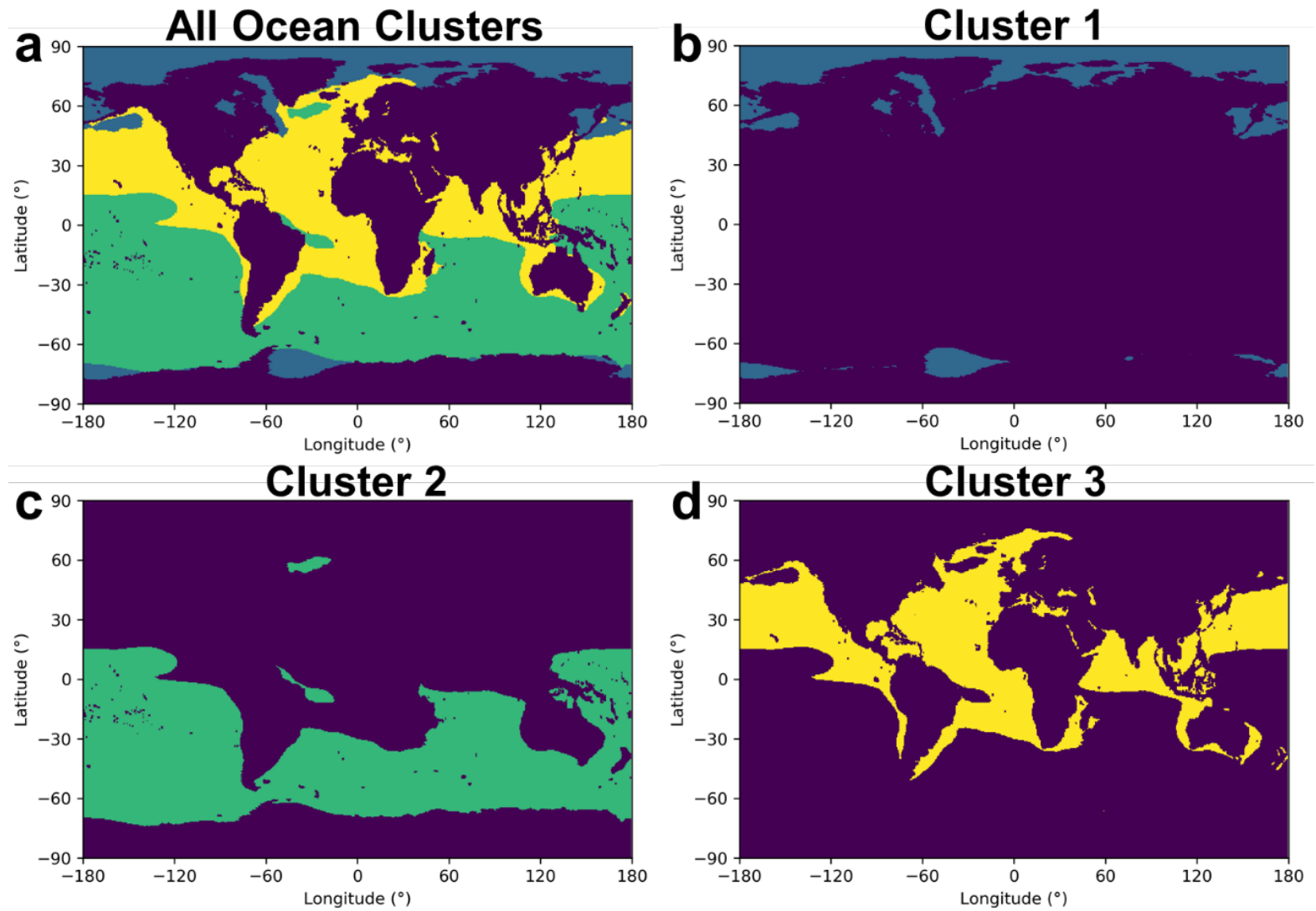


Figure D2. Ocean clusters 1 – 3 from the 2012 – 2016 Mar/Apr/May global MERRA-2 dataset.

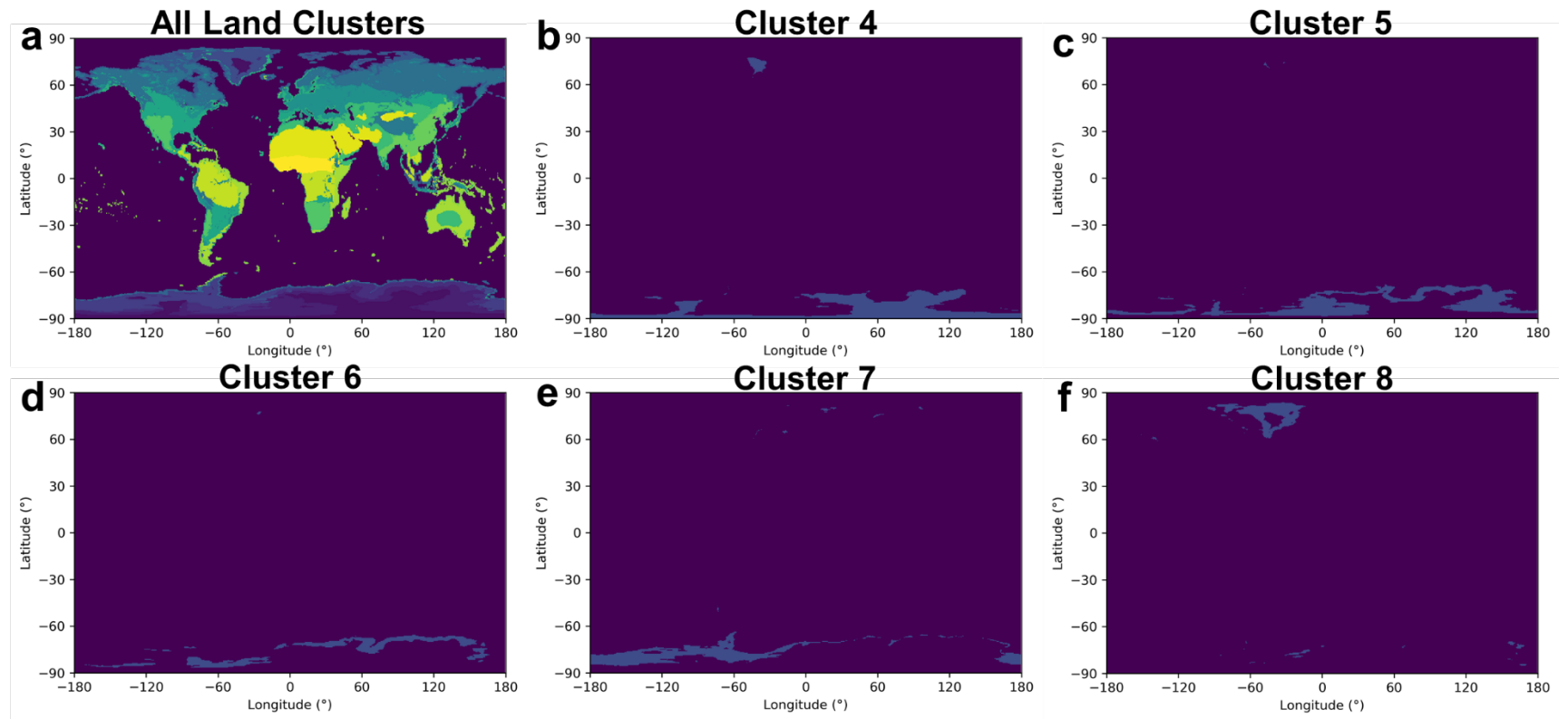


Figure D3. Land clusters 4 – 8 from the 2012 – 2016 Mar/Apr/May global MERRA-2 dataset.

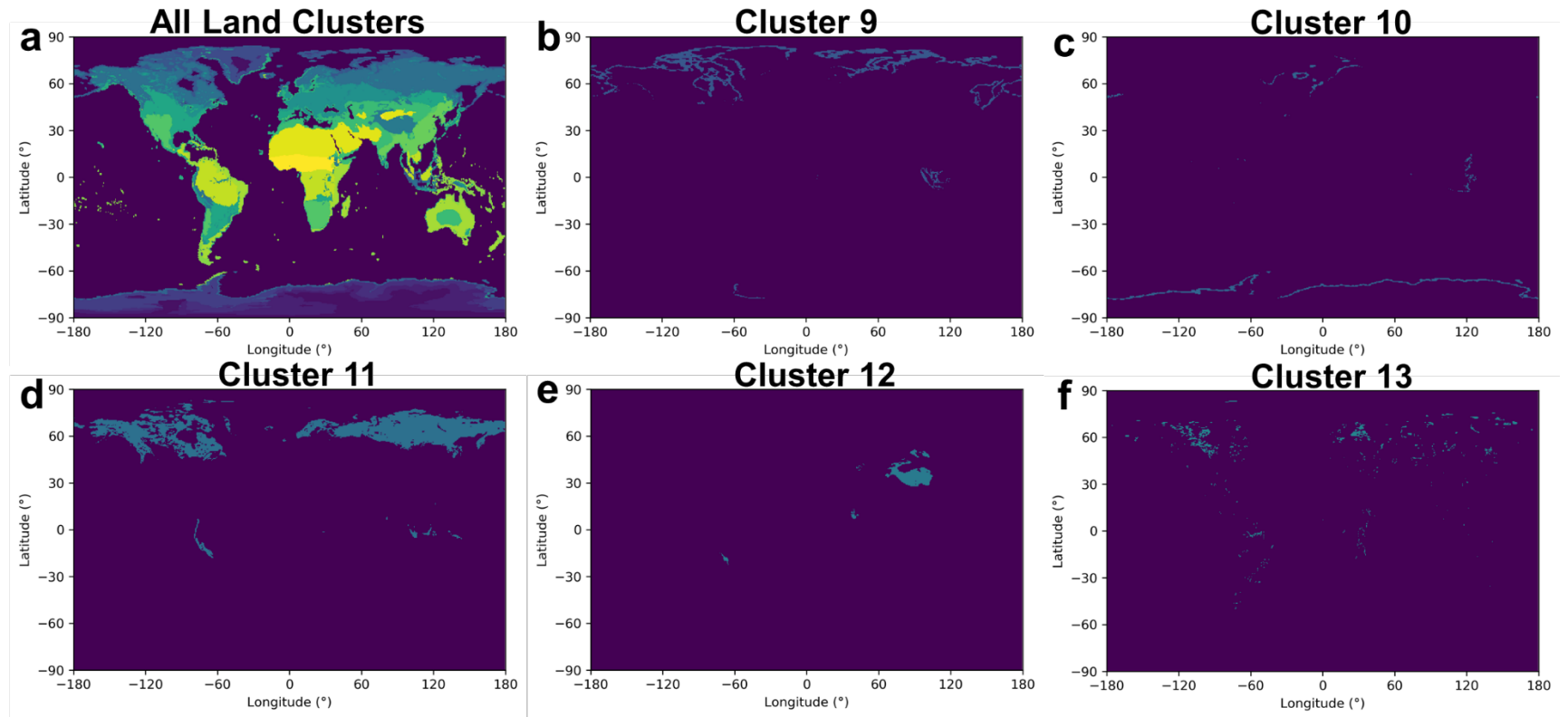


Figure D4. Land clusters 9 – 13 from the 2012 – 2016 Mar/Apr/May global MERRA-2 dataset.

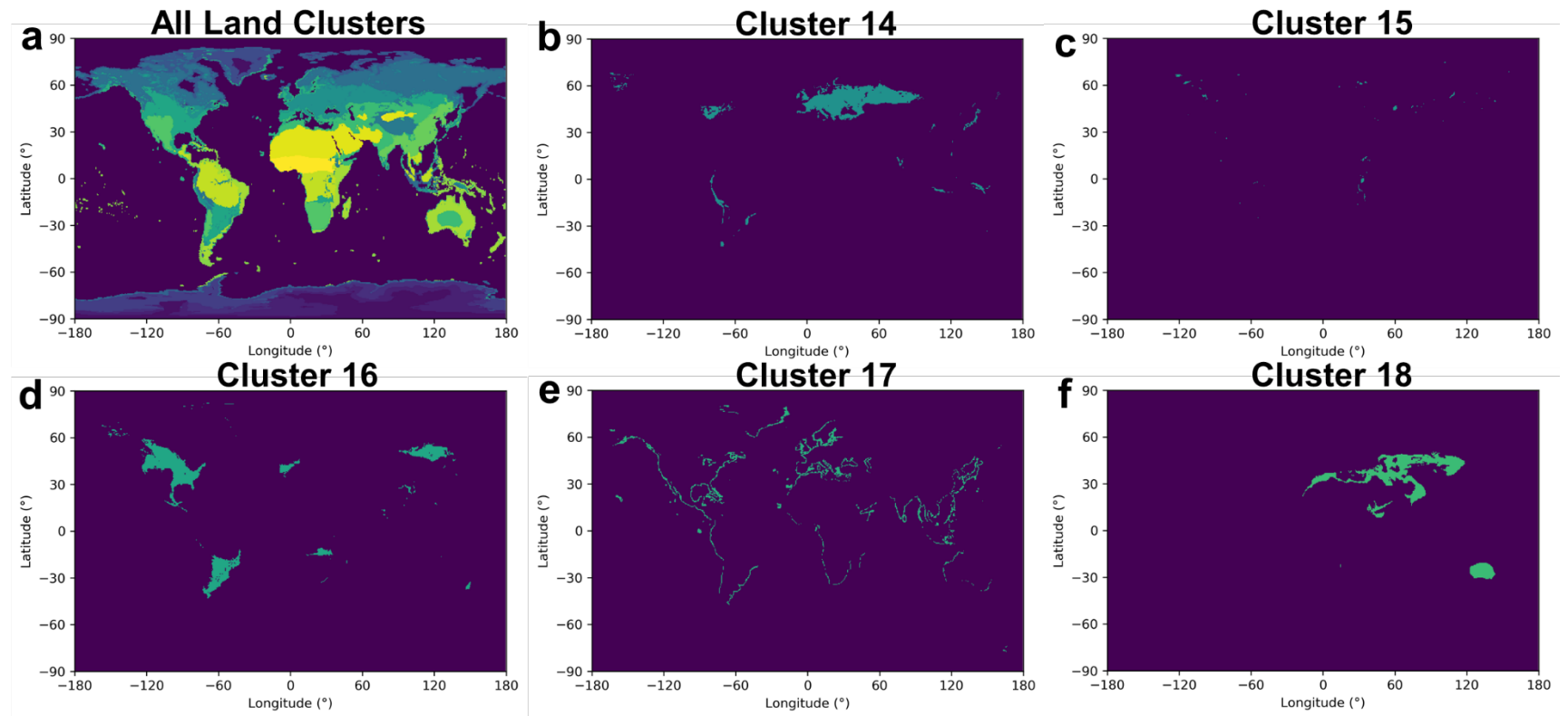


Figure D5. Land clusters 14 – 18 from the 2012 – 2016 Mar/Apr/May global MERRA-2 dataset.

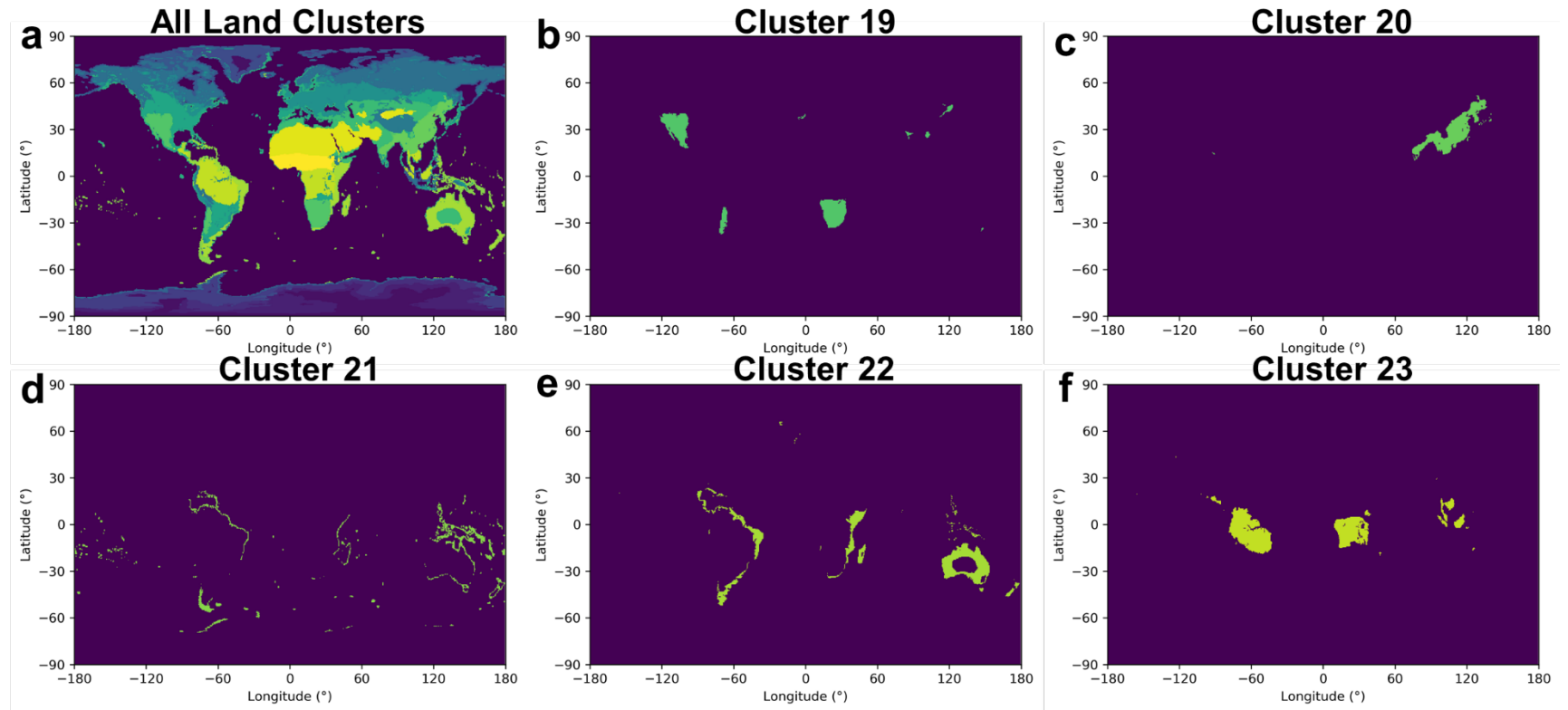


Figure D6. Land clusters 19 – 23 from the 2012 – 2016 Mar/Apr/May global MERRA-2 dataset.

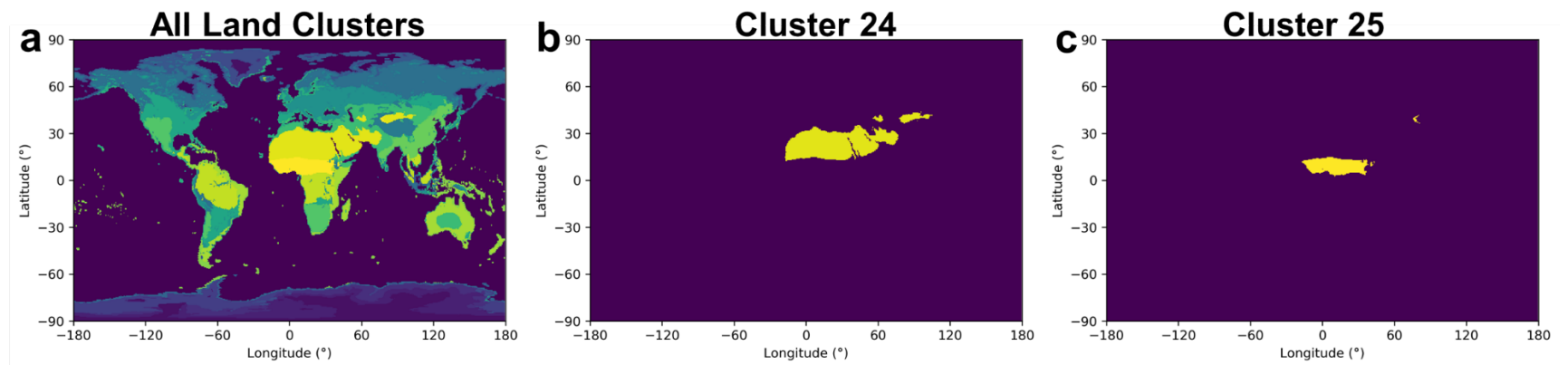


Figure D7. Land clusters 24 – 25 from the 2012 – 2016 Mar/Apr/May global MERRA-2 dataset.

E. 2012 to 2016 June through August Global Clusters

Table E1. Cluster centroids for the 2012 – 2016 Jun/Jul/Aug global MERRA-2 dataset.

Cluster	Fraction of Lake	Fraction of Land	Fraction of Land-Ice	Fraction of Ocean	Elevation (meters)	Relative Humidity (%)	2-m Air Temperature (K)	10-m Wind Speed (m s ⁻¹)	Fraction Showers	Fraction Cloud Coverage	Fraction Black Carbon	Fraction Dust	Fraction Organic Carbon	Fraction Sea Salt	Fraction Sulfate	Total AOD
1	0	2.53x10 ⁻⁸	3.28x10 ⁻⁸	1.00	5.51x10 ⁻⁵	87.5	274	8.86	0.955	0.804	0.0210	0.0408	0.0865	0.685	0.167	0.0913
2	2.99x10 ⁻⁹	2.67x10 ⁻⁷	1.34x10 ⁻⁷	1.00	0.0244	92.3	278	5.22	0.904	0.788	0.0618	0.122	0.381	0.115	0.319	0.123
3	6.65x10 ⁻¹⁰	2.01x10 ⁻⁷	0	1.00	-0.0310	79.3	297	6.23	0.862	0.562	0.0350	0.118	0.101	0.502	0.244	0.121
4	0	0	1.00	0	2939	98.6	215	7.91	0.969	0.835	0.0503	0.0738	0.240	0.197	0.440	0.0155
5	0	0	0.996	3.56x10 ⁻³	1776	85.1	231	12.4	0.848	0.707	0.0494	0.0839	0.214	0.257	0.397	0.0179
6	2.95x10 ⁻⁴	2.13x10 ⁻³	0.949	0.0479	486	89.5	240	6.38	0.899	0.785	0.0401	0.0729	0.185	0.347	0.355	0.0236
7	3.29x10 ⁻³	0.0508	0.892	0.0540	1608	86.5	268	4.87	0.725	0.519	0.0769	0.148	0.473	0.0176	0.285	0.0698
8	5.12x10 ⁻³	0.130	0.0601	0.805	38.4	91.0	277	4.36	0.862	0.718	0.0652	0.130	0.440	0.0737	0.292	0.114
9	9.64x10 ⁻³	0.975	0.0149	1.46x10 ⁻⁵	4128	53.5	281	2.66	0.634	0.450	0.0643	0.322	0.157	0.0220	0.438	0.0959
10	4.06x10 ⁻³	0.124	0.0379	0.834	30.6	84.5	284	6.25	0.892	0.704	0.0295	0.0919	0.106	0.507	0.265	0.105
11	0.0200	0.912	0.0160	0.0518	364	84.9	284	2.93	0.864	0.648	0.0628	0.138	0.436	0.0440	0.320	0.137
12	0.190	0.786	3.99x10 ⁻³	0.0202	305	75.2	289	2.95	0.768	0.552	0.0633	0.142	0.411	0.0501	0.335	0.170
13	0.0140	0.977	4.59x10 ⁻³	4.46x10 ⁻³	487	74.7	289	1.75	0.770	0.561	0.0663	0.129	0.492	0.0181	0.296	0.204
14	1.82x10 ⁻³	0.941	0	0.0565	268	54.6	290	3.85	0.495	0.291	0.0581	0.230	0.182	0.245	0.286	0.0753
15	4.28x10 ⁻³	0.985	0	0.0102	1113	37.2	291	3.47	0.284	0.162	0.148	0.0768	0.522	0.0503	0.203	0.145
16	0.0120	0.958	2.11x10 ⁻³	0.0278	426	76.2	291	2.34	0.768	0.566	0.0589	0.125	0.254	0.0826	0.481	0.159
17	3.90x10 ⁻³	0.159	8.32x10 ⁻⁴	0.836	28.7	81.7	293	4.28	0.704	0.532	0.0529	0.135	0.200	0.182	0.431	0.169
18	6.58x10 ⁻³	0.985	1.62x10 ⁻⁵	8.23x10 ⁻³	688	53.9	293	2.82	0.504	0.328	0.0828	0.133	0.381	0.0874	0.317	0.113
19	0.625	0.371	8.44x10 ⁻⁶	3.85x10 ⁻³	482	62.9	294	3.76	0.660	0.468	0.0705	0.200	0.390	0.0401	0.300	0.202
20	6.20x10 ⁻³	0.977	7.08x10 ⁻⁶	0.0161	858	46.8	295	3.04	0.529	0.348	0.0550	0.245	0.215	0.0444	0.442	0.174
21	0.0175	0.935	0	0.0475	564	66.3	297	1.52	0.755	0.563	0.136	0.0378	0.685	0.0142	0.127	0.516
22	8.78x10 ⁻³	0.929	0	0.0617	459	74.4	298	2.61	0.774	0.605	0.0617	0.115	0.120	0.0433	0.662	0.451
23	0.0103	0.958	3.79x10 ⁻⁵	0.0316	433	74.9	298	3.15	0.866	0.691	0.0353	0.377	0.160	0.168	0.261	0.233
24	2.58x10 ⁻³	0.975	2.53x10 ⁻⁵	0.0217	751	23.7	302	3.81	0.221	0.121	0.0318	0.651	0.0709	0.0371	0.212	0.298
25	2.09x10 ⁻³	0.910	0	0.0879	429	34.4	306	3.86	0.571	0.395	0.0165	0.753	0.0436	0.0446	0.145	0.557

Blue = Ocean; Orange = Permafrost; Red = Inland; Green = Lake; Purple = Coast

Table E2. Category percentiles for the ocean subset of the 2012 – 2016 Jun/Jul/Aug MERRA-2 dataset.

Percentile	Fraction of Lake	Fraction of Land	Fraction of Land-Ice	Fraction of Ocean	Elevation (meters)	Relative Humidity (%)	2-m Air Temperature (K)	10-m Wind Speed (m s^{-1})	Fraction Showers	Fraction Cloud Coverage	Fraction Black Carbon	Fraction Dust	Fraction Organic Carbon	Fraction Sea Salt	Fraction Sulfate	Total AOD
10	0	0	0	1.00	0	76.6	271	4.67	0.819	0.436	0.0160	0.0277	0.0532	0.0642	0.128	0.0622
25	0	0	0	1.00	0	79.5	274	5.28	0.876	0.576	0.0210	0.0383	0.0699	0.246	0.160	0.0801
50	0	0	0	1.00	0	83.5	288	6.43	0.920	0.721	0.0325	0.0671	0.102	0.521	0.220	0.0970
75	0	0	0	1.00	0	91.3	298	8.05	0.951	0.813	0.0522	0.132	0.185	0.687	0.295	0.125
90	0	0	0	1.00	0	94.7	301	9.51	0.971	0.876	0.0658	0.161	0.461	0.750	0.376	0.168

Table E3. Category percentiles for the land subset of the 2012 – 2016 Jun/Jul/Aug MERRA-2 dataset.

Percentile	Fraction of Lake	Fraction of Land	Fraction of Land-Ice	Fraction of Ocean	Elevation (meters)	Relative Humidity (%)	2-m Air Temperature (K)	10-m Wind Speed (m s^{-1})	Fraction Showers	Fraction Cloud Coverage	Fraction Black Carbon	Fraction Dust	Fraction Organic Carbon	Fraction Sea Salt	Fraction Sulfate	Total AOD
10	0	0	0	0	6.67	39.4	215	1.63	0.434	0.260	0.0307	0.0655	0.0870	0.0102	0.202	0.0158
25	0	0	0	0	105	65.5	248	2.67	0.669	0.463	0.0441	0.0770	0.171	0.0276	0.271	0.0261
50	0	0.831	0	0	412	81.8	285	3.92	0.825	0.619	0.0540	0.120	0.242	0.0862	0.356	0.115
75	3.77×10^{-3}	1.00	1.00	0.0160	1476	91.8	296	6.12	0.922	0.763	0.0650	0.158	0.398	0.213	0.432	0.191
90	0.0261	1.00	1.00	0.820	2850	98.7	301	9.34	0.973	0.853	0.0737	0.437	0.495	0.340	0.469	0.298

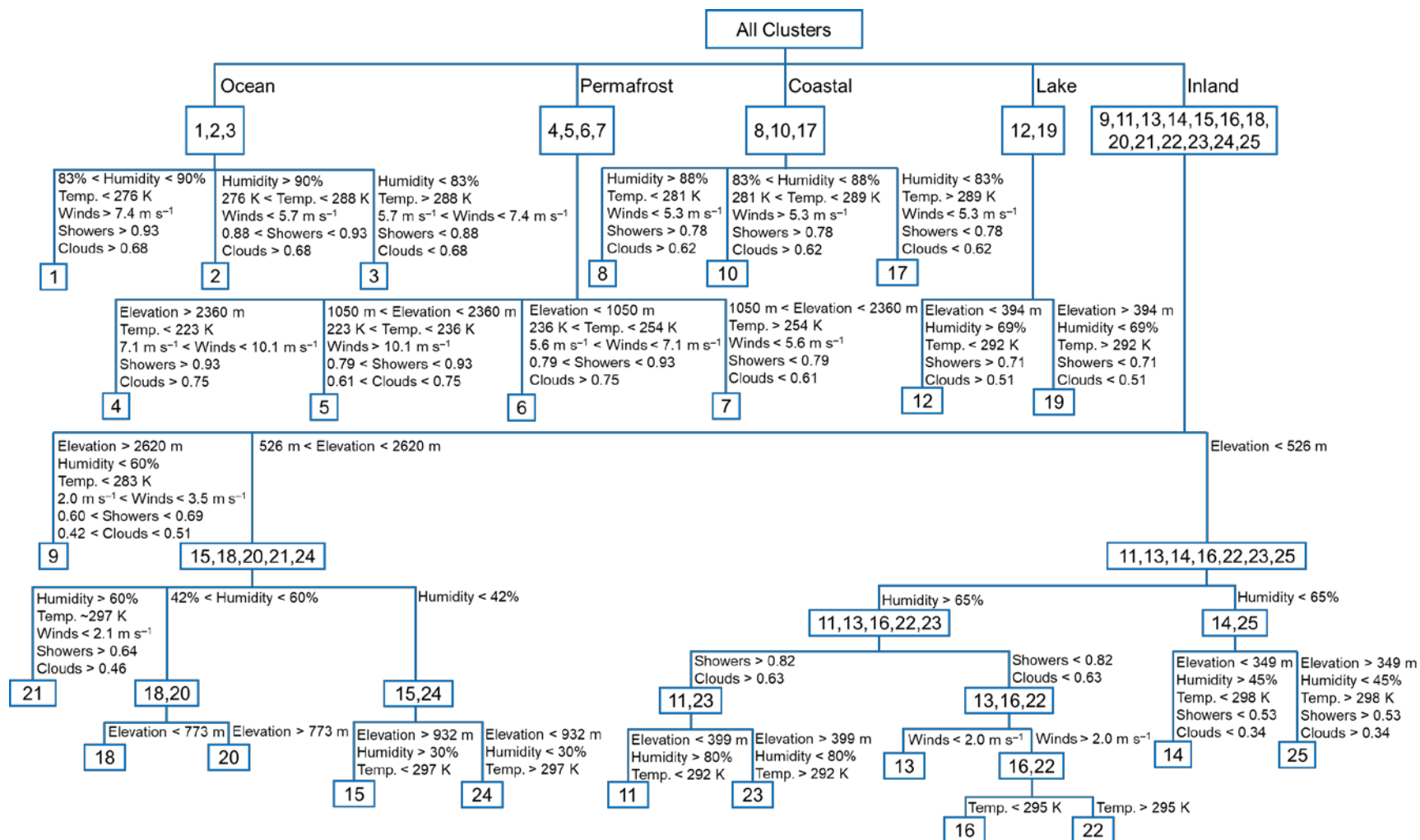


Figure E1. Decision tree for clusters describing the 2012 – 2016 Jun/Jul/Aug global MERRA-2 dataset.

Table E4. Cluster descriptions and AOD₅₅₀ statistics for the 2012 – 2016 Jun/Jul/Aug global MERRA-2 dataset.

Cluster	Qualitative Description	Characteristic Regions	AOD Centroid	AOD Average	AOD Standard Deviation	AOD 10th Percentile	AOD 50th Percentile	AOD 90th Percentile
1	Regions of the oceans: Humidity: medium high; Temperature: low; Wind speed: high; Showers: high; Clouds coverage: medium high	The Southern Ocean excluding a region closest to Antarctica southeast of Argentina. Includes some of the southern Pacific and Atlantic oceans.	0.0913	0.0908	0.0315	0.0386	0.0970	0.128
2	Regions of the oceans: Humidity: high; Temperature: medium low; Wind speed: low; Showers: medium low; Clouds coverage: medium high	Arctic Ocean, Hudson Bay north through the Davis Strait, northern Atlantic and Pacific Oceans, Bearing Sea, Atlantic Ocean just west of Africa, Sea of Japan and Sea of Okhotsk.	0.123	0.124	0.0746	0.0818	0.0977	0.186
3	Regions of the oceans: Humidity: low; Temperature: medium high; Wind speed: medium low; Showers: low; Clouds coverage: low	Central Pacific and Atlantic Oceans, Indian Ocean, Mediterranean, Black, Red, and Caspian Seas.	0.121	0.120	0.0802	0.0645	0.0966	0.196
4	Inland permafrost regions: Elevation: very high; Humidity: high; Temperature: very low; Wind speed: high; Showers: high; Cloud coverage: high	Inland permafrost regions of Antarctica at very high elevation and with high winds. Rainiest and Cloudiest of the permafrost regions.	0.0155	0.0155	1.93x10 ⁻³	0.0132	0.0153	0.0176
5	Inland permafrost regions: Elevation: high; Humidity: medium high; Temperature: low; Wind speed: very high; Showers: medium high; Cloud coverage: medium high	Inland permafrost regions of Antarctica at high elevation with the highest wind speed of all regions.	0.0179	0.0179	3.31x10 ⁻³	0.0137	0.0177	0.0221
6	Inland permafrost regions: Elevation: medium high; Humidity: medium high; Temperature: low; Wind speed: high; Showers: medium high; Cloud coverage: high	Inland permafrost regions of Antarctica at the lowest elevation of the permafrost regions.	0.0236	0.0237	4.36x10 ⁻³	0.0197	0.0226	0.0293
7	Inland permafrost regions: Elevation: high; Humidity: medium high; Temperature: medium low; Wind speed: medium high; Showers: medium low; Cloud coverage: medium low	Inland permafrost regions of Greenland. Calmest winds, least rainy, and least cloudy of the permafrost regions.	0.0698	0.0700	0.0103	0.0578	0.0690	0.0841
8	Coastal regions: Elevation: low; Humidity: medium high; Temperature: medium low; Wind speed: medium high; Showers: medium high; Cloud coverage: medium high	Coastal regions concentrated in the north. Coastal Alaska, Greenland, and northern Canada, coastal Northern and eastern Russia.	0.114	0.114	0.0385	0.0793	0.105	0.152
9	Inland regions: Elevation: very high; Humidity: low; Temperature: medium low; Wind speed: low; Showers: low; Cloud coverage: low	Tibet and the Andes Mountains from central Peru to central Chile and Argentina. Highest elevation of all regions.	0.0959	0.0961	0.0482	0.0385	0.0881	0.162
10	Coastal regions: Elevation: low; Humidity: medium high; Temperature: medium low; Wind speed: high; Showers: medium high; Cloud coverage: medium high	Islands and coastal regions concentrated near the equator. Includes the Caribbean, northern, eastern, and far southern coastal South America, coastal Malaysia, Indonesia, and Papua New Guinea, coastal northern and southern Australia, coastal New Zealand, coastal western India and Eastern Africa, and coastal Antarctica.	0.105	0.106	0.0760	0.0297	0.0880	0.206
11	Inland regions: Elevation: medium low; Humidity: medium high; Temperature: medium low; Wind speed: medium low; Showers: medium high; Cloud coverage: medium high	Inland regions of Alaska and northern and eastern Canada, inland north Scandinavian peninsula, inland far north Russia, inland north Brazil, and strip of Africa from Liberia to South Sudan.	0.137	0.137	0.0363	0.0926	0.136	0.180
12	Inland regions near lakes: Elevation: medium low; Humidity: medium low; Temperature: medium high; Wind speed: medium low; Showers: medium low; Cloud coverage: medium low	Scattered inland regions near small lakes concentrated in central and northern Canada, Finland and Russia, South America, and eastern Africa.	0.170	0.170	0.0644	0.108	0.159	0.245
13	Inland regions: Elevation: medium high; Humidity: medium low; Temperature: medium high; Wind speed: low; Showers: medium low; Cloud coverage: medium low	Inland regions of central Canada, central and eastern Russia, portion of Brazil south of the Amazon river, and strip of Africa including southern Cameroon and the northern Democratic Republic of the Congo.	0.204	0.204	0.0646	0.111	0.205	0.290

UNCLASSIFIED

14	Inland regions: Elevation: medium low; Humidity: low; Temperature: medium high; Wind speed: medium low; Showers: low; Cloud coverage: low	Inland Australia, eastern and near coastal Brazil, portions of near coastal Madagascar and east Africa, near coastal Mediterranean regions, land around the Gulf of California.	0.0753	0.0752	0.0337	0.0499	0.0650	0.106
15	Inland regions: Elevation: medium high; Humidity: very low; Temperature: medium high; Wind speed: medium low; Showers: very low; Cloud coverage: very low	Inland areas of southern Africa from northern South Africa to central Angola and Tanzania. Excludes Mozambique.	0.145	0.146	0.0583	0.0919	0.130	0.215
16	Inland regions: Elevation: medium high; Humidity: medium low; Temperature: medium high; Wind speed: low; Showers: medium low; Cloud coverage: medium low	Inland eastern United States and southeastern Canada, inland northern Mexico, British Isles and France east through Europe including southern Russia to western Mongolia, southeastern Russia, southwestern China south to southern Vietnam, inland Malaysia, Indonesia, and Papua New Guinea, southeastern Australia, northern portion of the Andes mountains, and southeastern Brazil through southern Argentina.	0.159	0.159	0.0542	0.0913	0.162	0.222
17	Coastal regions: Elevation: low; Humidity: medium low; Temperature: medium high; Wind speed: medium high; Showers: medium low; Cloud coverage: medium low	Coastal regions concentrated in the mid-latitudes. Includes the coastal United States and Great Lakes, coastal western Mexico, all of Coastal Europe, coastal China, Korea, and Japan, coastal eastern Russia, coastal Malaysia and Indonesia, portions of coastal Australia, coastal South Africa, and coastal western and southeastern South America.	0.169	0.170	0.0642	0.104	0.161	0.254
18	Inland regions: Elevation: medium high; Humidity: low; Temperature: medium high; Wind speed: medium low; Showers: low; Cloud coverage: low	Inland northwestern United States from Minnesota to Washington and south into central California, central Brazil to northern Argentina, and inland but near coastal eastern and southern Africa. Includes inland Madagascar.	0.113	0.113	0.0371	0.0660	0.110	0.171
19	Inland regions near lakes: Elevation: medium high; Humidity: low; Temperature: medium high; Wind speed: medium low; Showers: low; Cloud coverage: medium low	Scattered inland regions near large lakes concentrated in northwestern Canada, Russia, Kazakhstan, Uzbekistan, and eastern Africa.	0.202	0.202	0.0764	0.119	0.198	0.295
20	Inland regions: Elevation: medium high; Humidity: low; Temperature: medium high; Wind speed: medium low; Showers: low; Cloud coverage: low	Inland arid portions of the southwestern United States and northern Mexico, regions of central and eastern Argentina, inland Mediterranean Europe from Spain east including land around the Black sea, Turkey, land east of the northern part of the Caspian Sea through Mongolia into Inner Mongolia in China.	0.174	0.174	0.0525	0.108	0.181	0.227
21	Inland regions: Elevation: medium high; Humidity: medium low; Temperature: high; Wind speed: very low; Showers: medium low; Cloud coverage: medium low	Inland equatorial Africa including Gabon, the Republic of the Congo, and the Democratic Republic of the Congo. Regions of Africa associated with tropical rainforest and jungle with the calmest winds of any region.	0.516	0.518	0.156	0.312	0.519	0.726
22	Inland regions: Elevation: medium high; Humidity: medium low; Temperature: high; Wind speed: low; Showers: medium low; Cloud coverage: medium low	Inland eastern China, Japan, Korea, and northeastern India.	0.451	0.453	0.142	0.296	0.420	0.660
23	Inland regions: Elevation: medium high; Humidity: medium low; Temperature: high; Wind speed: medium low; Showers: medium high; Cloud coverage: medium high	Inland southern Mexico and Central America, northern near coastal portion of South America, sub-Saharan Africa just north of the equator from Guinea east to Ethiopia, and northern and western India.	0.233	0.238	0.101	0.105	0.224	0.369
24	Inland regions: Elevation: medium high; Humidity: very low; Temperature: very high; Wind speed: medium low; Showers: very low; Cloud coverage: very low	Northern half of the Sahara desert approaching the Mediterranean Sea, northern Arabian peninsula to the Azerbaijan, majority of Iran and Afghanistan north to southern Kazakhstan. Includes regions around the central Taklamakan and Gobi deserts in China.	0.298	0.299	0.0777	0.211	0.287	0.404
25	Inland regions: Elevation: medium high; Humidity: very low; Temperature: very high; Wind speed: medium low; Showers: low; Cloud coverage: low	Southern half of the Sahara desert, southern Arabian Peninsula to northern Somalia, and southern Iran to northeastern India. Includes center of the Taklamakan desert in China.	0.557	0.560	0.101	0.419	0.562	0.695

Blue = Ocean; Orange = Permafrost; Red = Inland; Green = Lake; Purple = Coast

UNCLASSIFIED

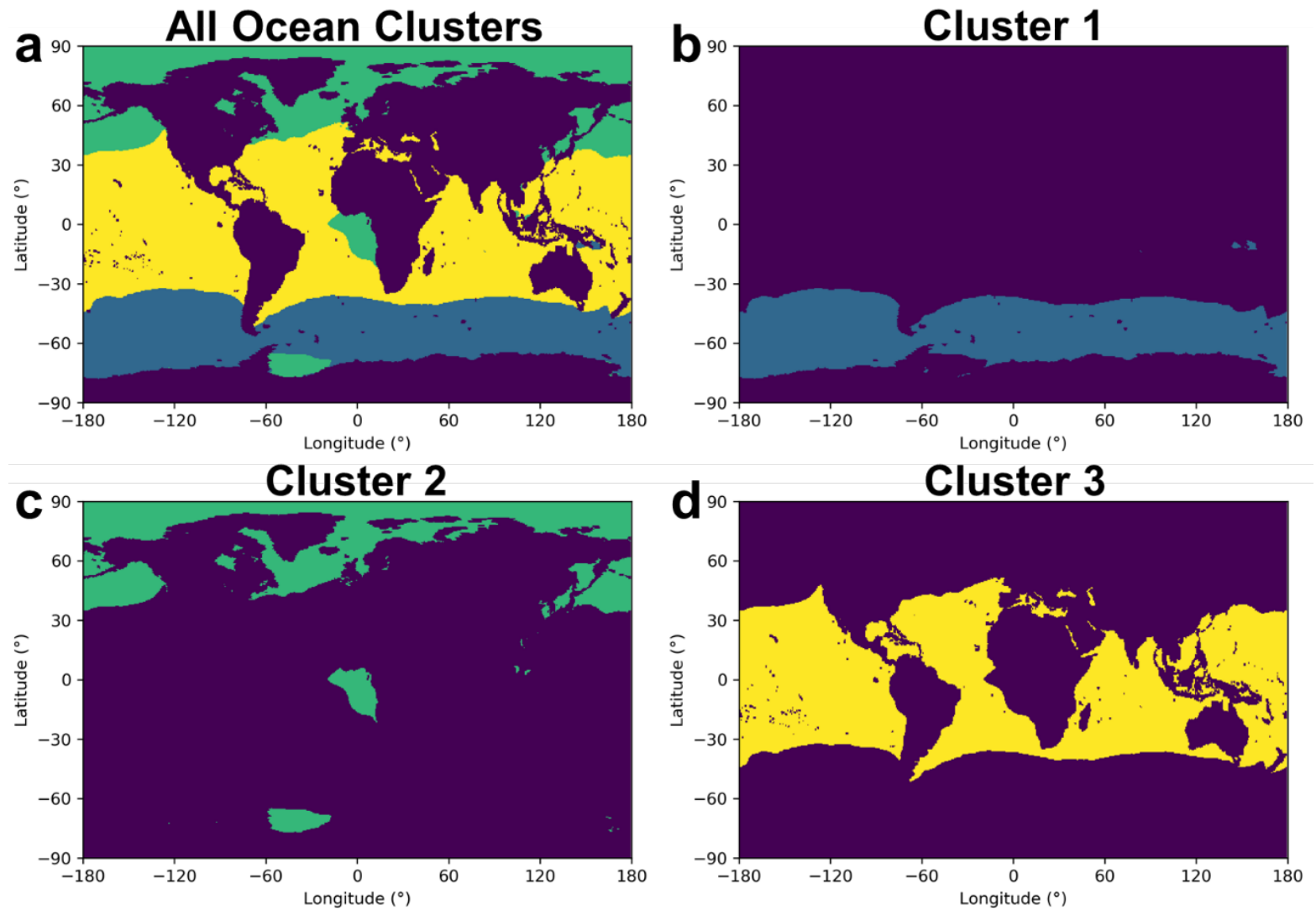


Figure E2. Ocean clusters 1 – 3 from the 2012 – 2016 Jun/Jul/Aug global MERRA-2 dataset.

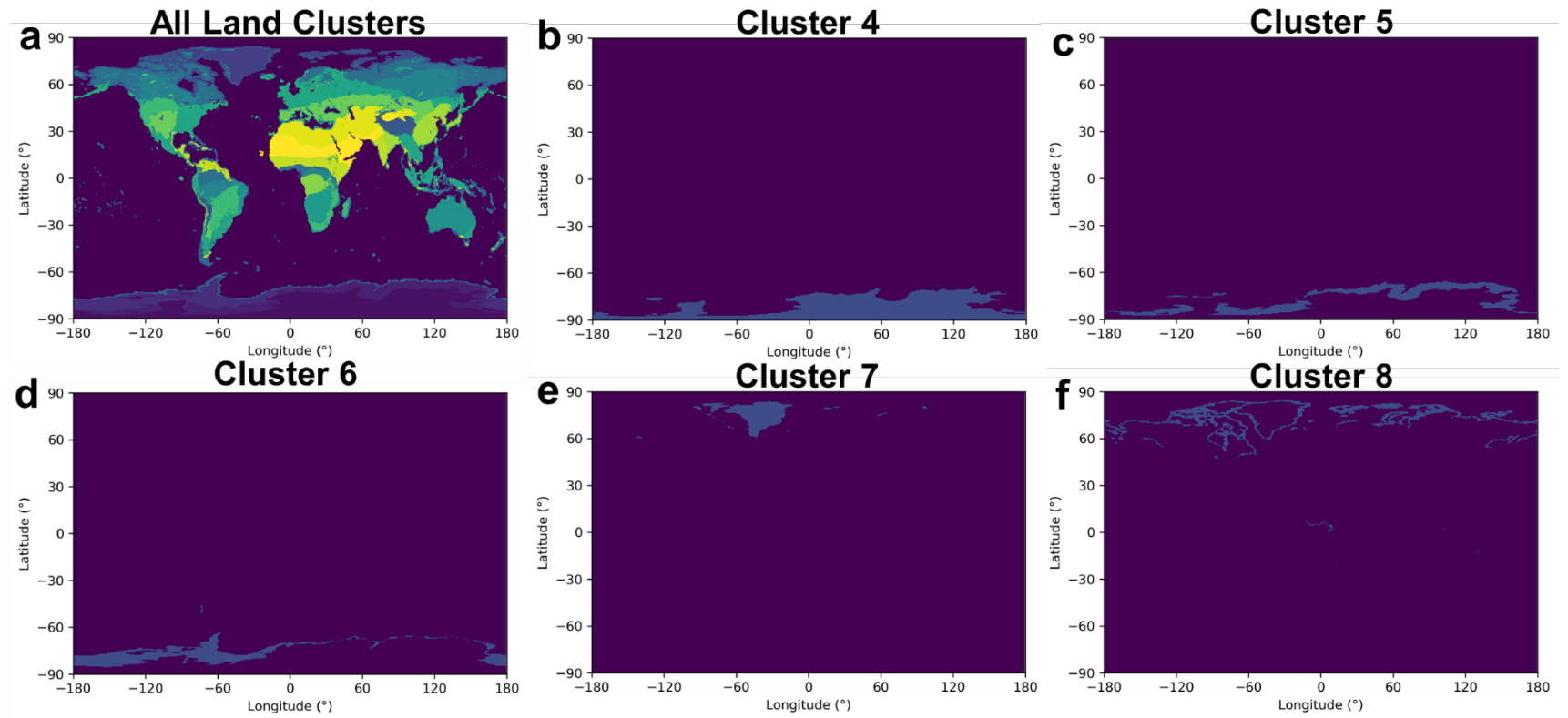


Figure E3. Land clusters 4 – 8 from the 2012 – 2016 Jun/Jul/Aug global MERRA-2 dataset.

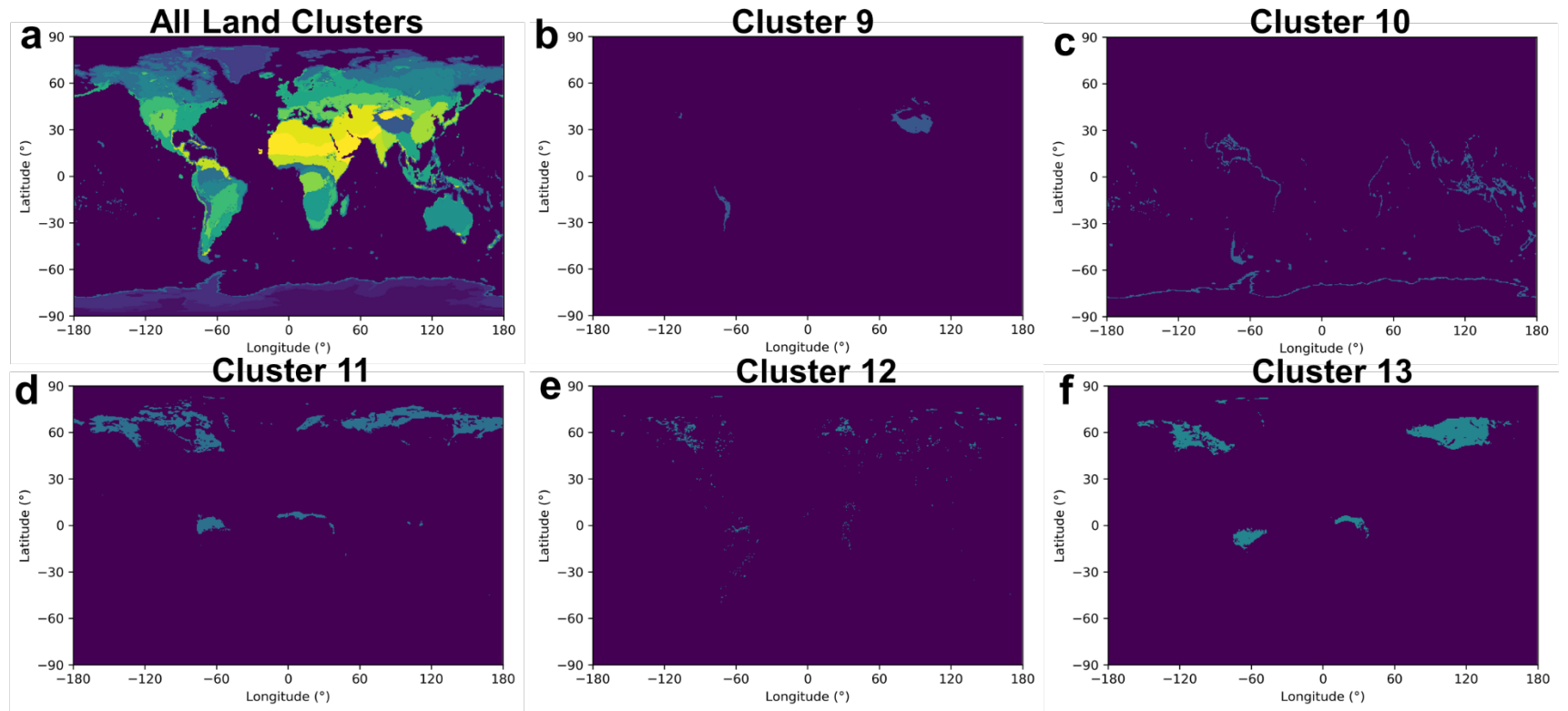


Figure E4. Land clusters 9 – 13 from the 2012 – 2016 Jun/Jul/Aug global MERRA-2 dataset.

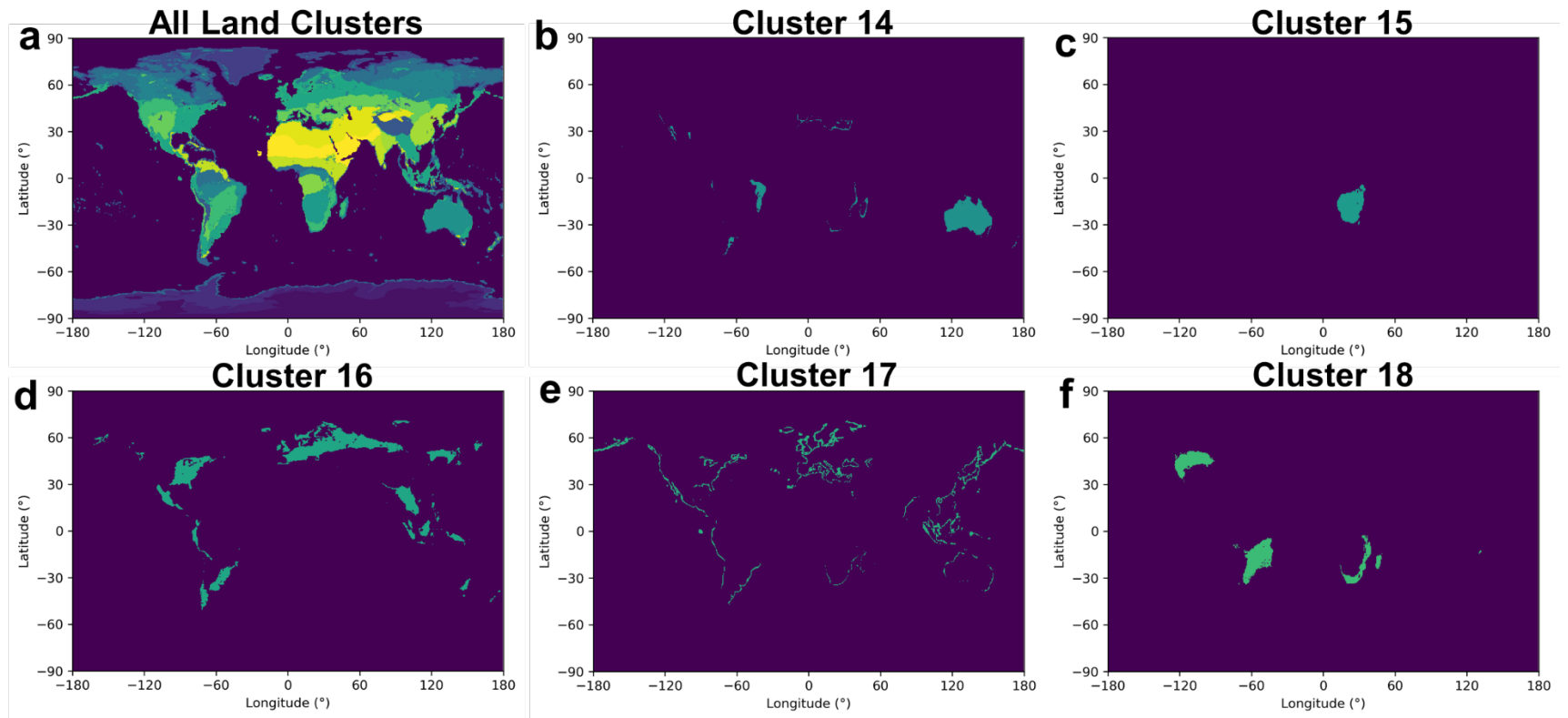


Figure E5. Land clusters 14 – 18 from the 2012 – 2016 Jun/Jul/Aug global MERRA-2 dataset.

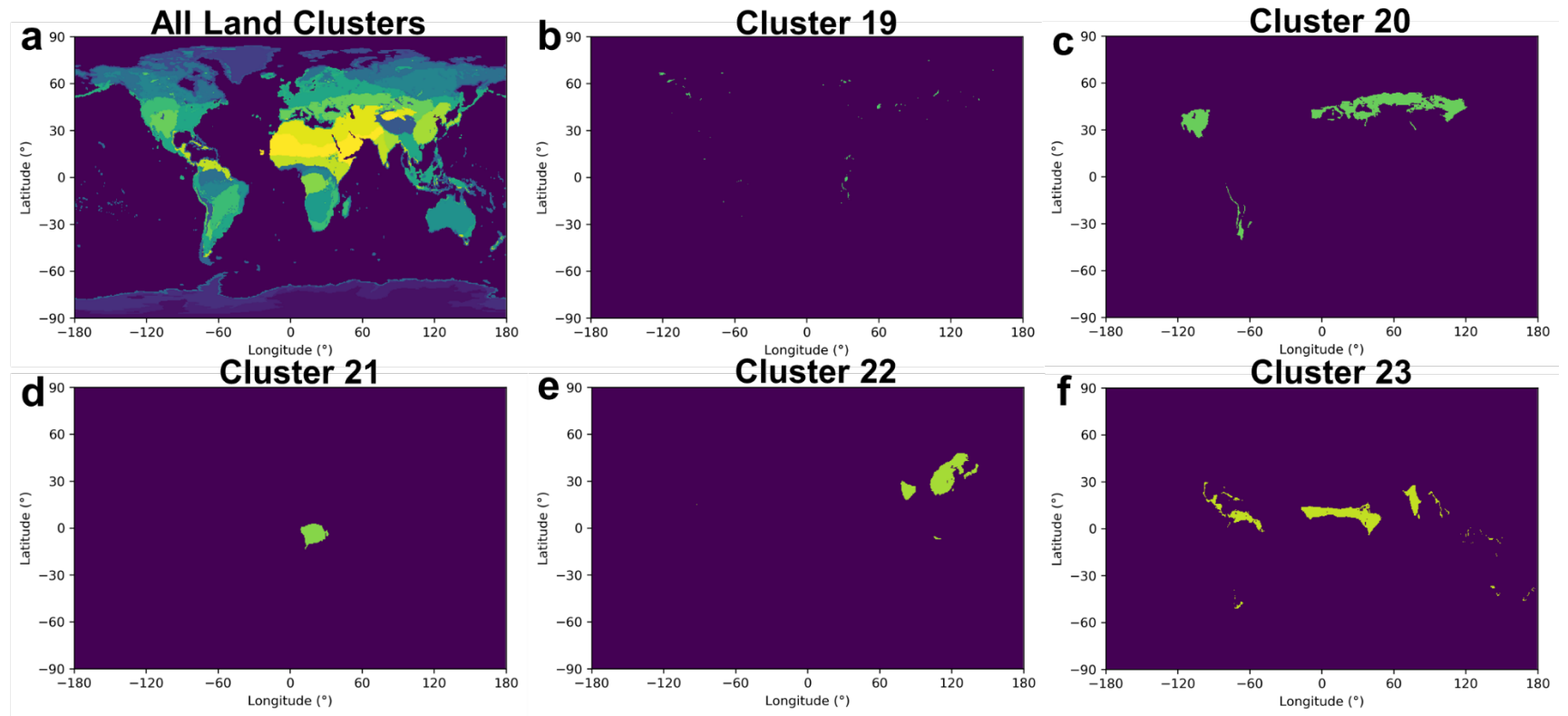


Figure E6. Land clusters 19 – 23 from the 2012 – 2016 Jun/Jul/Aug global MERRA-2 dataset.

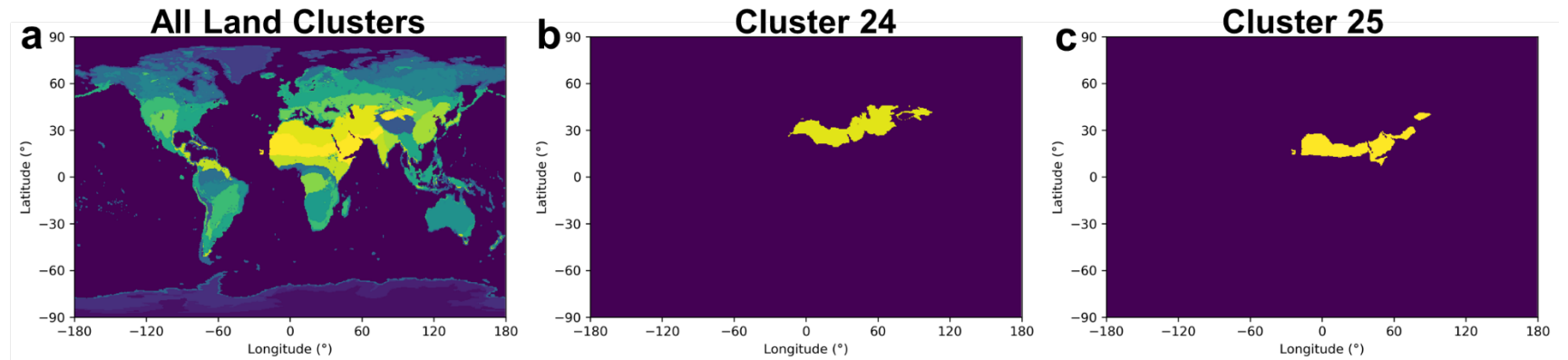


Figure E7. Land clusters 24 – 25 from the 2012 – 2016 Jun/Jul/Aug global MERRA-2 dataset.

F. 2012 to 2016 September through November Global Clusters

Table F1. Cluster centroids for the 2012 – 2016 Sep/Oct/Nov global MERRA-2 dataset.

Cluster	Fraction of Lake	Fraction of Land	Fraction of Land-Ice	Fraction of Ocean	Elevation (meters)	Relative Humidity (%)	2-m Air Temperature (K)	10-m Wind Speed (m s ⁻¹)	Fraction Showers	Fraction Cloud Coverage	Fraction Black Carbon	Fraction Dust	Fraction Organic Carbon	Fraction Sea Salt	Fraction Sulfate	Total AOD
1	3.23x10 ⁻⁹	1.90x10 ⁻⁷	2.12x10 ⁻⁷	1.00	8.02x10 ⁻³	93.6	262	5.88	0.960	0.864	0.0584	0.115	0.236	0.193	0.396	0.0562
2	8.06x10 ⁻¹¹	4.62x10 ⁻⁸	3.20x10 ⁻⁹	1.00	7.09x10 ⁻⁵	86.0	277	8.70	0.940	0.760	0.0332	0.0624	0.148	0.558	0.197	0.110
3	7.66x10 ⁻¹⁰	2.18x10 ⁻⁷	3.02x10 ⁻⁹	1.00	-0.0193	79.4	294	6.06	0.861	0.587	0.0486	0.0784	0.161	0.461	0.250	0.118
4	5.37x10 ⁻⁹	6.30x10 ⁻⁵	1.00	4.43x10 ⁻⁶	2983	96.8	226	6.53	0.865	0.624	0.0717	0.122	0.340	0.0857	0.380	0.0223
5	1.27x10 ⁻⁵	1.43x10 ⁻³	0.998	1.38x10 ⁻⁴	2514	91.9	231	9.48	0.722	0.516	0.0764	0.138	0.320	0.0932	0.372	0.0218
6	3.52x10 ⁻⁵	1.29x10 ⁻³	0.986	0.0122	1273	78.8	245	11.3	0.717	0.582	0.0666	0.132	0.316	0.140	0.345	0.0276
7	1.39x10 ⁻³	0.0193	0.918	0.0611	525	88.3	250	5.63	0.802	0.668	0.0618	0.121	0.292	0.172	0.353	0.0337
8	0.0141	0.940	0.0135	0.0322	408	93.2	267	3.01	0.914	0.729	0.0595	0.129	0.226	0.0724	0.513	0.0852
9	0.136	0.825	8.81x10 ⁻³	0.0293	275	89.3	271	3.18	0.883	0.689	0.0600	0.127	0.233	0.0895	0.491	0.0939
10	4.62x10 ⁻³	0.132	0.0588	0.805	33.6	88.2	272	5.49	0.894	0.748	0.0591	0.107	0.215	0.209	0.410	0.0876
11	0.010	0.973	0.0166	1.60x10 ⁻⁵	4200	49.1	274	2.99	0.536	0.380	0.0808	0.239	0.204	0.0188	0.457	0.0621
12	7.87x10 ⁻³	0.965	1.04x10 ⁻³	0.0253	481	78.2	281	3.01	0.789	0.602	0.0623	0.135	0.168	0.0582	0.576	0.145
13	0.530	0.467	5.45x10 ⁻⁶	3.68x10 ⁻³	451	74.0	282	3.36	0.767	0.584	0.0656	0.174	0.284	0.0678	0.408	0.138
14	6.28x10 ⁻³	0.979	1.37x10 ⁻⁴	0.0139	1139	54.7	282	3.18	0.580	0.416	0.0488	0.387	0.112	0.0438	0.408	0.149
15	7.12x10 ⁻³	0.982	2.05x10 ⁻³	8.11x10 ⁻³	766	64.5	286	2.98	0.602	0.440	0.0756	0.129	0.274	0.0702	0.451	0.116
16	3.66x10 ⁻³	0.0877	0.0129	0.896	18.5	81.8	288	6.22	0.881	0.676	0.0408	0.0608	0.154	0.457	0.287	0.114
17	9.94x10 ⁻³	0.942	2.10x10 ⁻⁶	0.0481	369	71.4	291	2.48	0.594	0.418	0.0810	0.132	0.139	0.0276	0.621	0.456
18	2.80x10 ⁻³	0.169	0	0.828	34.8	75.1	293	4.76	0.597	0.411	0.0580	0.194	0.159	0.197	0.392	0.181
19	2.26x10 ⁻³	0.971	7.68x10 ⁻⁶	0.0266	695	33.6	294	3.14	0.344	0.229	0.0347	0.598	0.0635	0.0419	0.261	0.240
20	0.0143	0.846	5.11x10 ⁻³	0.134	255	74.3	294	3.14	0.829	0.632	0.0565	0.0755	0.249	0.321	0.297	0.143
21	1.47x10 ⁻³	0.979	0	0.0188	358	35.5	296	3.68	0.395	0.241	0.0852	0.219	0.296	0.154	0.245	0.131
22	0.0144	0.970	1.67x10 ⁻⁶	0.0154	474	73.3	298	1.24	0.861	0.688	0.0834	0.0671	0.531	0.0911	0.227	0.233
23	5.93x10 ⁻³	0.987	0	7.02x10 ⁻³	849	40.4	298	3.66	0.578	0.387	0.111	0.0621	0.508	0.104	0.215	0.196
24	8.34x10 ⁻³	0.966	0	0.0254	443	71.7	298	1.80	0.822	0.638	0.0518	0.480	0.246	0.0451	0.177	0.241
25	1.98x10 ⁻³	0.990	0	7.82x10 ⁻³	479	29.0	299	3.40	0.523	0.366	0.0201	0.802	0.0534	0.0177	0.107	0.336

Blue = Ocean; Orange = Permafrost; Red = Inland; Green = Lake; Purple = Coast

Table F2. Category percentiles for the ocean subset of the 2012 – 2016 Sep/Oct/Nov MERRA-2 dataset.

Percentile	Fraction of Lake	Fraction of Land	Fraction of Land-Ice	Fraction of Ocean	Elevation (meters)	Relative Humidity (%)	2-m Air Temperature (K)	10-m Wind Speed (m s^{-1})	Fraction Showers	Fraction Cloud Coverage	Fraction Black Carbon	Fraction Dust	Fraction Organic Carbon	Fraction Sea Salt	Fraction Sulfate	Total AOD
10	0	0	0	1.00	0	77.0	263	4.99	0.817	0.481	0.0261	0.0196	0.0809	0.154	0.156	0.0552
25	0	0	0	1.00	0	79.2	273	5.61	0.875	0.582	0.0333	0.0496	0.114	0.344	0.186	0.0766
50	0	0	0	1.00	0	82.0	287	6.52	0.921	0.691	0.0448	0.0676	0.163	0.458	0.227	0.100
75	0	0	0	1.00	0	87.3	298	7.71	0.953	0.806	0.0569	0.107	0.223	0.570	0.324	0.122
90	0	0	0	1.00	0	92.8	301	9.08	0.972	0.863	0.0684	0.128	0.258	0.648	0.452	0.150

Table F3. Category percentiles for the land subset of the 2012 – 2016 Sep/Oct/Nov MERRA-2 dataset.

Percentile	Fraction of Lake	Fraction of Land	Fraction of Land-Ice	Fraction of Ocean	Elevation (meters)	Relative Humidity (%)	2-m Air Temperature (K)	10-m Wind Speed (m s^{-1})	Fraction Showers	Fraction Cloud Coverage	Fraction Black Carbon	Fraction Dust	Fraction Organic Carbon	Fraction Sea Salt	Fraction Sulfate	Total AOD
10	0	0	0	0	6.67	42.4	226	1.83	0.474	0.321	0.0402	0.0519	0.0969	0.0181	0.197	0.0213
25	0	0	0	0	105	68.5	251	2.71	0.662	0.475	0.0537	0.106	0.174	0.0410	0.319	0.0327
50	0	0.831	0	0	412	84.6	273	4.12	0.809	0.608	0.0644	0.127	0.252	0.0918	0.375	0.0886
75	3.77×10^{-3}	1.00	1.00	0.0160	1476	94.1	294	6.22	0.894	0.706	0.0735	0.152	0.330	0.158	0.478	0.151
90	0.0261	1.00	1.00	0.820	2850	96.7	300	8.64	0.942	0.767	0.0860	0.399	0.355	0.267	0.553	0.255

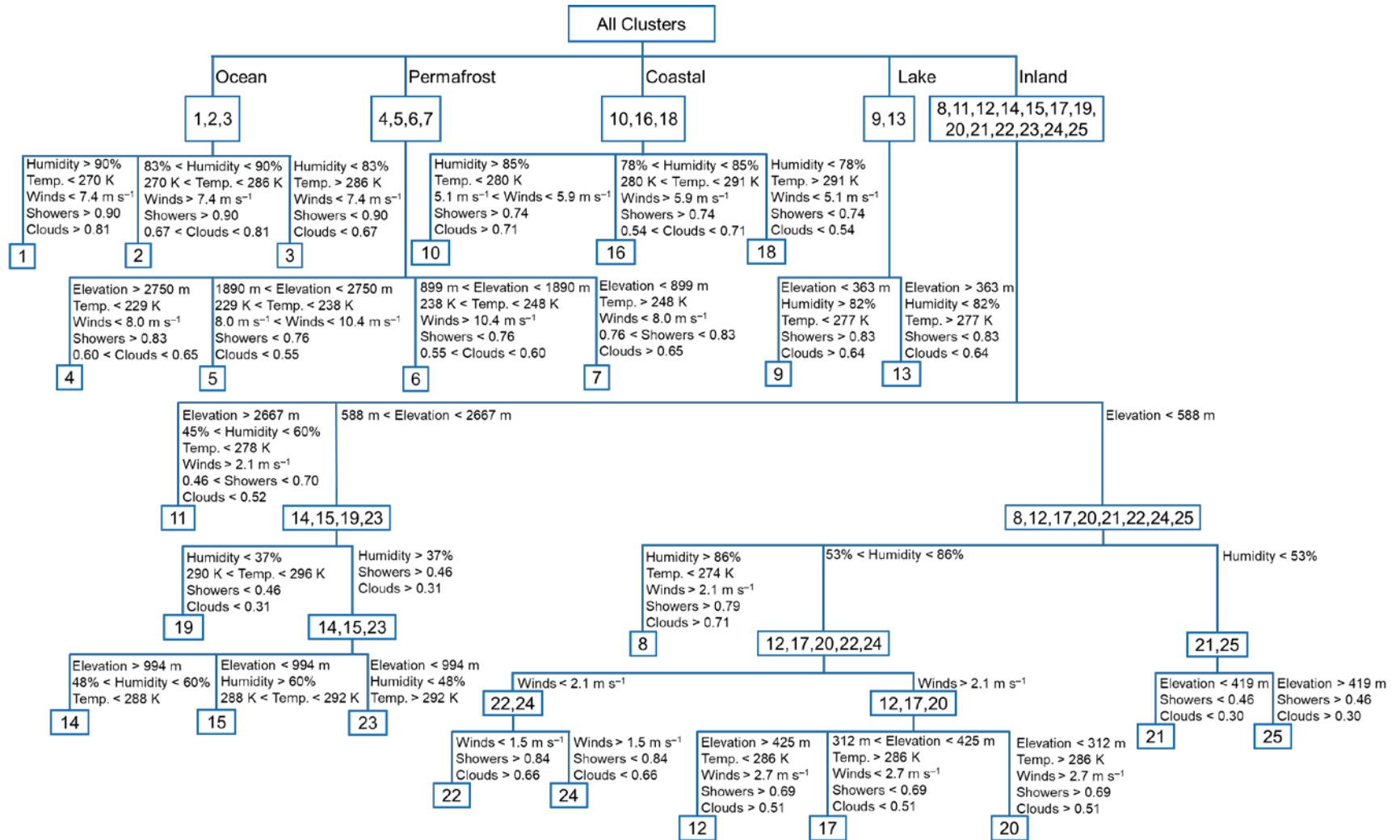


Figure F1. Decision tree for clusters describing the 2012 – 2016 Sep/Oct/Nov global MERRA-2 dataset.

Table F4. Cluster descriptions and AOD₅₅₀ statistics for the 2012 – 2016 Sep/Oct/Nov global MERRA-2 dataset.

Cluster	Qualitative Description	Characteristic Regions	AOD Centroid	AOD Average	AOD Standard Deviation	AOD 10th Percentile	AOD 50th Percentile	AOD 90th Percentile
1	Regions of the oceans: Humidity: very high; Temperature: very low; Wind speed: medium low; Showers: high; Clouds coverage: very high	Arctic Ocean, Hudson Bay north through the Davis Strait, portions of the Southern Ocean nearest Antarctica.	0.0562	0.0561	0.0121	0.0410	0.0543	0.0730
2	Regions of the oceans: Humidity: medium high; Temperature: medium low; Wind speed: high; Showers: medium high; Clouds coverage: medium high	Northern Atlantic Ocean west and north of the British Isles stretching north of Norway and western Russia, Bearing Sea, far northern Pacific Ocean, strip of the central Pacific Ocean, southern Pacific and Atlantic Oceans.	0.110	0.109	0.0215	0.0790	0.112	0.134
3	Regions of the oceans: Humidity: medium low; Temperature: medium high; Wind speed: medium low; Showers: low; Clouds coverage: medium low	Majority of the central Atlantic, Pacific, and Indian Oceans except small central regions of the Pacific and Indian Oceans.	0.118	0.118	0.0485	0.0758	0.104	0.176
4	Inland permafrost regions: Elevation: very high; Humidity: very high; Temperature: low; Wind speed: high; Showers: medium high; Cloud coverage: medium high	Inland permafrost regions of central Antarctica and Greenland at very high elevation. Rainiest of the permafrost regions.	0.0223	0.0222	3.77×10^{-3}	0.0190	0.0212	0.0283
5	Inland permafrost regions: Elevation: high; Humidity: medium high; Temperature: low; Wind speed: very high; Showers: medium low; Cloud coverage: medium low	Inland permafrost regions of Antarctica and Greenland at high elevation with very high winds. Least cloudy of the permafrost regions.	0.0218	0.0217	4.83×10^{-3}	0.0171	0.0207	0.0260
6	Inland permafrost regions: Elevation: medium high; Humidity: medium low; Temperature: low; Wind speed: very high; Showers: medium low; Cloud coverage: medium low	Inland permafrost regions of Antarctica and a small portion of Greenland with the highest winds of all regions. Least rainy of the permafrost regions.	0.0276	0.0275	4.07×10^{-3}	0.0233	0.0269	0.0331
7	Inland permafrost regions: Elevation: medium high; Humidity: medium high; Temperature: low; Wind speed: medium high; Showers: medium low; Cloud coverage: medium high	Inland permafrost regions of Antarctica and Greenland nearest the coasts. Lowest elevation and cloudiest of the permafrost regions.	0.0337	0.0336	0.0106	0.0262	0.0313	0.0432
8	Inland regions: Elevation: medium low; Humidity: medium high; Temperature: medium low; Wind speed: medium low; Showers: high; Cloud coverage: high	Inland regions of northern Canada, northern Russia, and the Scandinavian peninsula. Includes the northern part of the Andes in Colombia, portions of southern Central America, and the northern half of the island of New Guinea.	0.0852	0.0847	0.0264	0.0535	0.0807	0.121
9	Inland regions near lakes: Elevation: medium low; Humidity: medium high; Temperature: medium low; Wind speed: medium low; Showers: medium high; Cloud coverage: medium high	Scattered inland regions near small lakes concentrated in central Canada, the Scandinavian peninsula, Finland, and Russia.	0.0939	0.0949	0.0354	0.0611	0.0880	0.140
10	Coastal regions: Elevation: low; Humidity: medium high; Temperature: medium low; Wind speed: medium high; Showers: high; Cloud coverage: high	Coastal regions concentrated along the coast of Antarctica and in the northern latitudes. Includes the coast of Canada, Alaska, Greenland, northern Europe, northern and eastern Russia, and coastal southeast Asia, Malaysia, and Indonesia.	0.0876	0.0873	0.0544	0.0438	0.0752	0.154
11	Inland regions: Elevation: very high; Humidity: low; Temperature: medium high; Wind speed: medium low; Showers: low; Cloud coverage: low	Tibet, regions of the Andes mountains from southern Peru to central Chile. Highest elevation of all regions.	0.0621	0.0618	0.0266	0.0378	0.0535	0.0954
12	Inland regions: Elevation: medium high; Humidity: medium low; Temperature: medium high; Wind speed: medium low; Showers: medium low; Cloud coverage: medium low	Parts of the United States and Canada around the Great Lakes and along the St. Lawrence River, central Mexico, southern India, southeast Asia, Japan, and northern Spain through France to Denmark east through southern Russia.	0.145	0.144	0.0491	0.0930	0.137	0.198
13	Inland regions near lakes: Elevation: medium high; Humidity: medium low; Temperature: medium high; Wind speed: medium low; Showers: medium low; Cloud coverage: medium low	Scattered global regions near large lakes concentrated in northwestern Canada, eastern Africa, Finland, and Russia.	0.138	0.139	0.0725	0.0695	0.114	0.236

UNCLASSIFIED

14	Inland regions: Elevation: medium high; Humidity: low; Temperature: medium high; Wind speed: medium low; Showers: low; Cloud coverage: low	Northwestern near coastal Africa, inland areas near the Horn of Africa, Turkey east to the Caspian Sea, southern Kazakhstan east through Mongolia into Inner Mongolia and Ningxia in China.	0.149	0.149	0.0505	0.0867	0.142	0.217
15	Inland regions: Elevation: medium high; Humidity: low; Temperature: medium high; Wind speed: medium low; Showers: low; Cloud coverage: low	Inland United States except regions around the Great Lakes and in the far southwest, northern Mexico, southern Brazil and Argentina, southern Spain, Myanmar, strip of southeast South Africa and Australia.	0.116	0.115	0.0441	0.0680	0.105	0.172
16	Coastal regions: Elevation: low; Humidity: medium low; Temperature: medium high; Wind speed: high; Showers: medium high; Cloud coverage: medium high	Islands and coastal regions concentrated near the equator. Includes the Caribbean, coastal Central America and southern Mexico, coastal Colombia through Brazil, coastal Iceland, northern British Isles, and Norway, coastal Philippines and New Guinea, coastal New Zealand, southern tip of South America, parts of coastal Antarctica.	0.114	0.114	0.0394	0.0615	0.113	0.162
17	Inland regions: Elevation: medium low; Humidity: medium low; Temperature: medium high; Wind speed: low; Showers: low; Cloud coverage: low	Inland eastern China, Korea, northern India, and Bangladesh.	0.456	0.454	0.113	0.327	0.434	0.612
18	Coastal regions: Elevation: low; Humidity: medium low; Temperature: medium high; Wind speed: medium high; Showers: low; Cloud coverage: low	Coastal regions concentrated in the mid-latitudes. Coastal United States, coastal Great Lakes, coastal Mediterranean, Black, and Caspian Seas, coastal Arabian Peninsula, coastal India, China, Korea, and Japan, coastal Australia, coastal southern Africa and Madagascar, coastal southeastern and western South America.	0.181	0.180	0.0793	0.101	0.163	0.305
19	Inland regions: Elevation: medium high; Humidity: very low; Temperature: high; Wind speed: medium low; Showers: very low; Cloud coverage: very low	Eastern Sahara, Arabian Peninsula, Iraq, majority of Syria and Iran, Afghanistan and Pakistan north to Uzbekistan, and the Gobi Desert in China.	0.240	0.239	0.0670	0.163	0.227	0.337
20	Inland regions: Elevation: medium low; Humidity: medium low; Temperature: high; Wind speed: medium low; Showers: medium high; Cloud coverage: medium high	Inland Central America, inland but near coastal areas of South American from Colombia through Brazil, inland British Isles, inland parts of Ethiopia, Kenya, and Somalia, inland parts of Indonesia and Papua New Guinea, and inland New Zealand.	0.143	0.142	0.0534	0.0833	0.134	0.208
21	Inland regions: Elevation: medium low; Humidity: very low; Temperature: high; Wind speed: medium low; Showers: very low; Cloud coverage: very low	Inland Australia, southwestern United States and the land around the Gulf of California, central Argentina, northwestern India, some inland but near coastal parts of southern Africa.	0.131	0.131	0.0628	0.0827	0.115	0.184
22	Inland regions: Elevation: medium high; Humidity: medium low; Temperature: high; Wind speed: very low; Showers: medium high; Cloud coverage: medium high	Equatorial regions of South America, Africa, and Indonesia/Malaysia. Regions associated with tropical rainforest and jungle. Includes an area of northwest Canada. Lowest winds of all regions.	0.233	0.232	0.0822	0.135	0.220	0.340
23	Inland regions: Elevation: medium high; Humidity: very low; Temperature: high; Wind speed: medium low; Showers: low; Cloud coverage: low	Inland sub-Saharan Africa from central Angola and Tanzania south to include the majority of South Africa, inland Madagascar, eastern Brazil, region in Paraguay and Bolivia.	0.196	0.196	0.0579	0.116	0.194	0.276
24	Inland regions: Elevation: medium high; Humidity: medium low; Temperature: high; Wind speed: very low; Showers: medium high; Cloud coverage: medium high	Strip of sub-Saharan Africa north of the equator from Senegal and Liberia east to western Ethiopia and Kenya.	0.241	0.240	0.0496	0.181	0.238	0.305
25	Inland regions: Elevation: medium high; Humidity: very low; Temperature: high; Wind speed: medium low; Showers: low; Cloud coverage: low	Western and southern Sahara desert, including parts of Chad and Sudan, and the Taklamakan Desert in China.	0.336	0.334	0.0678	0.243	0.341	0.411

Blue = Ocean; Orange = Permafrost; Red = Inland; Green = Lake; Purple = Coast

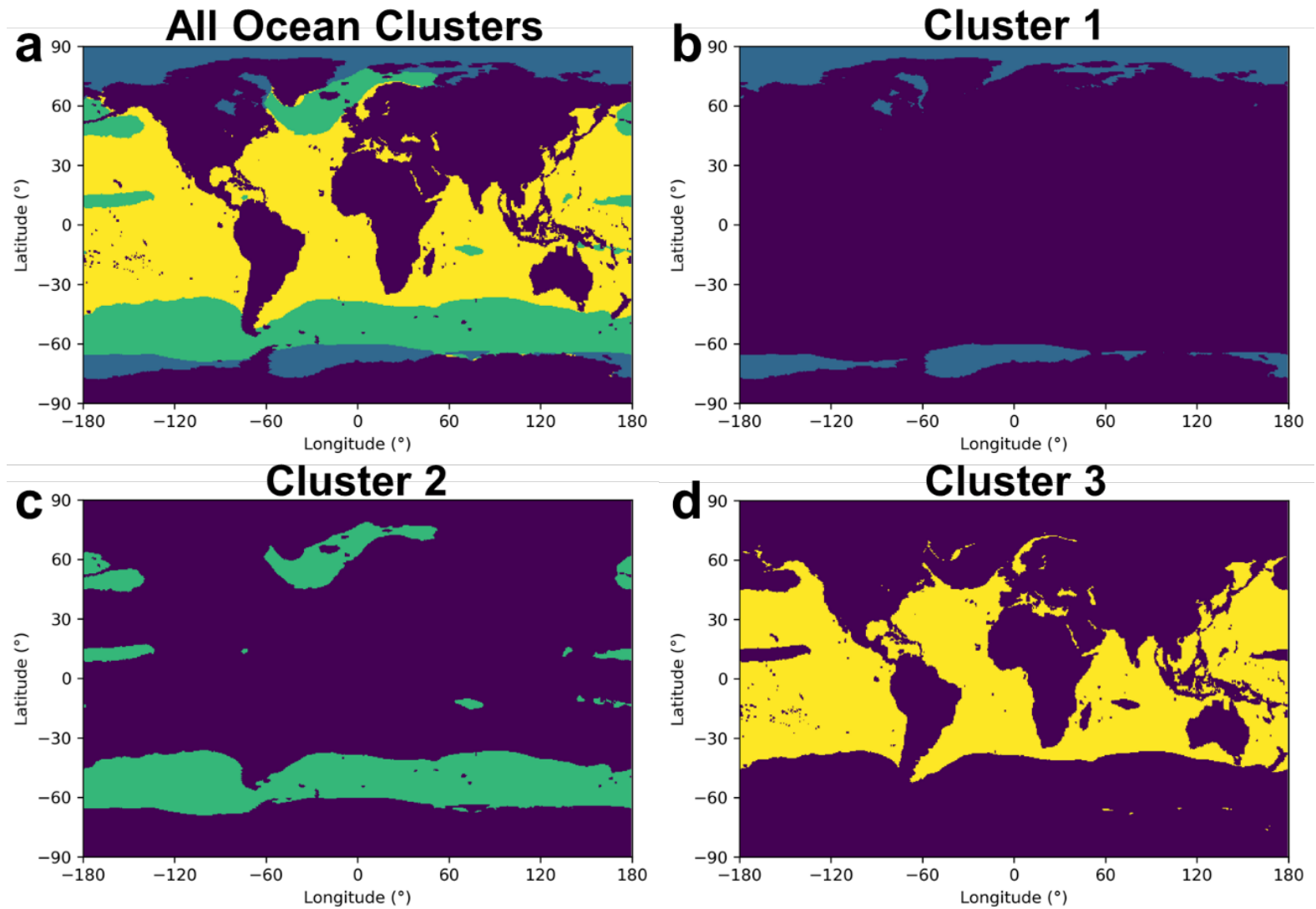


Figure F2. Ocean clusters 1 – 3 from the 2012 – 2016 Sep/Oct/Nov global MERRA-2 dataset.

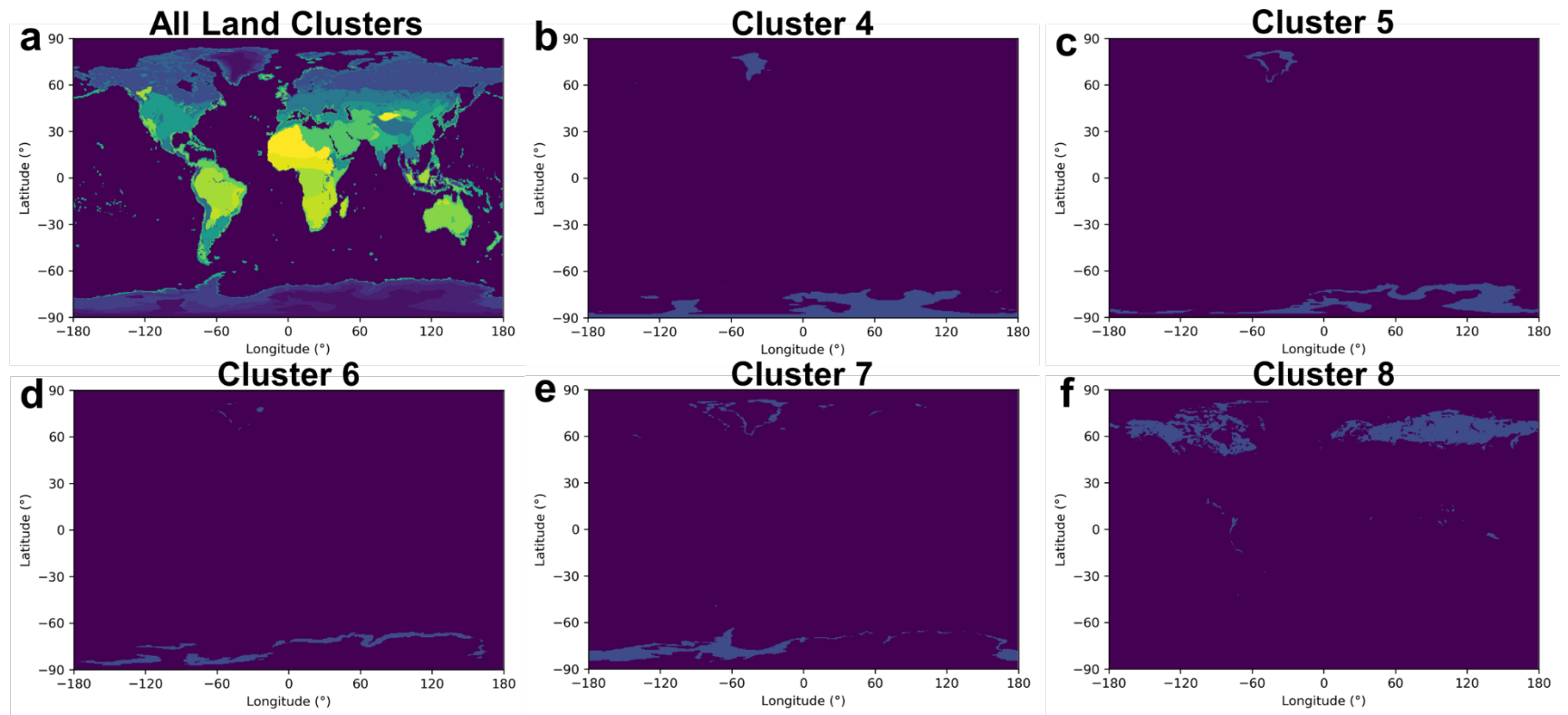


Figure F3. Land clusters 4 – 8 from the 2012 – 2016 Sep/Oct/Nov global MERRA-2 dataset.

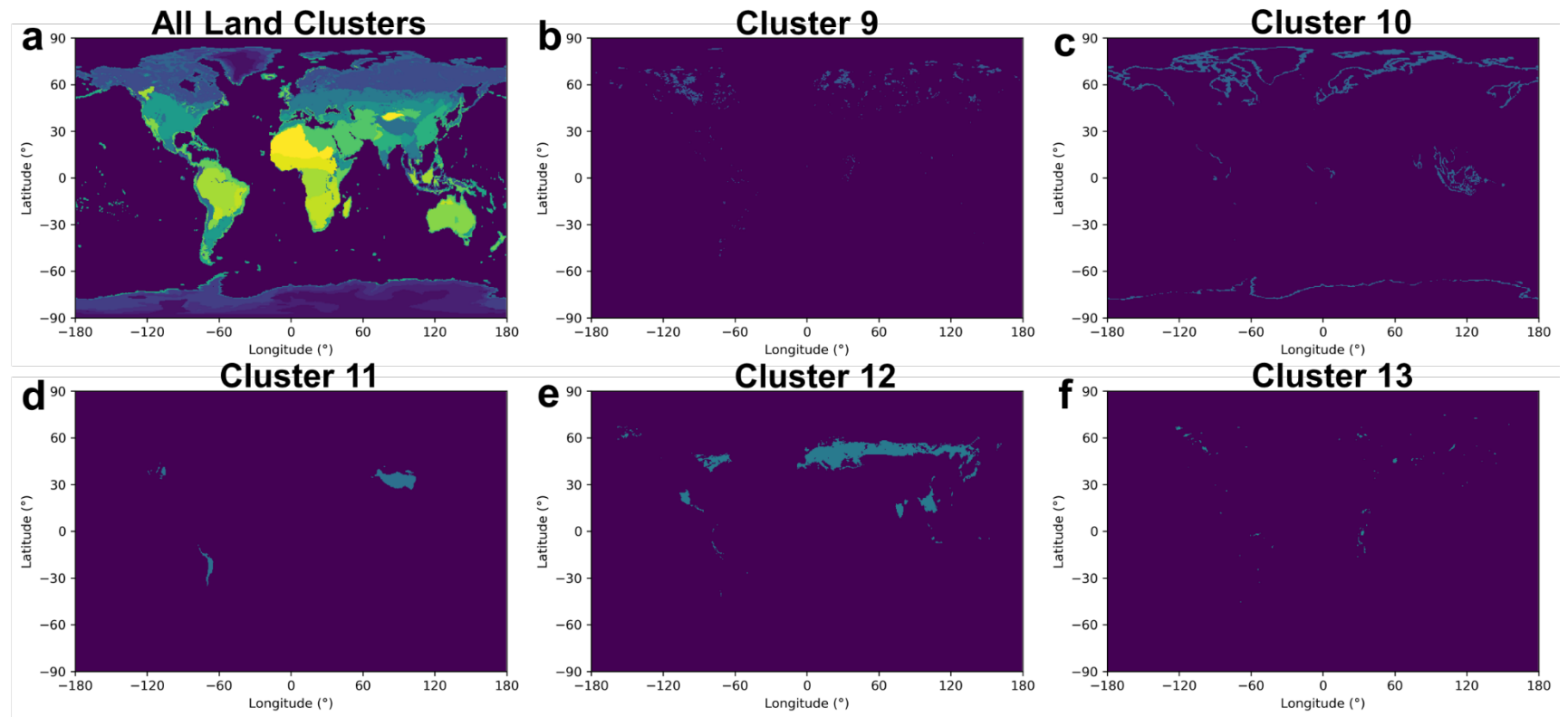


Figure F4. Land clusters 9 – 13 from the 2012 – 2016 Sep/Oct/Nov global MERRA-2 dataset.

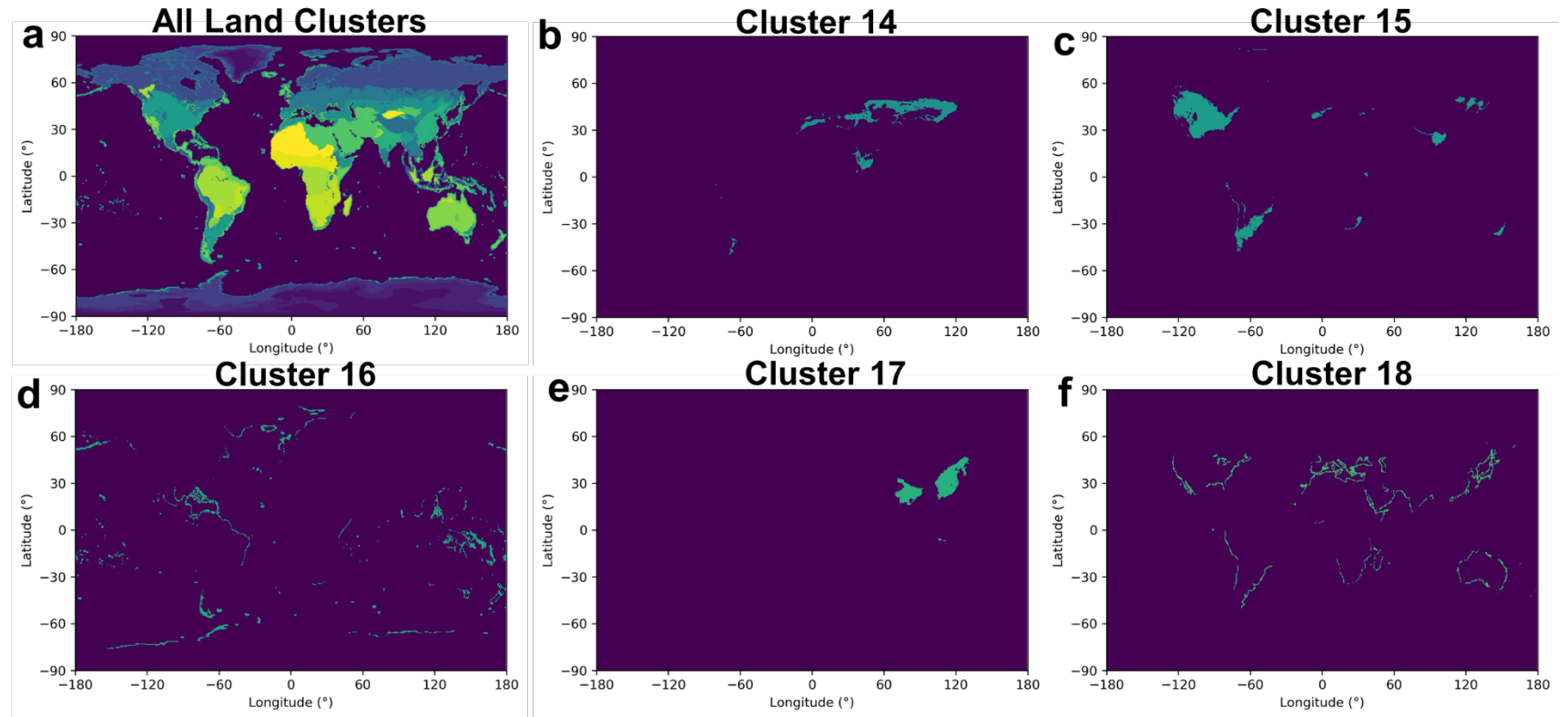


Figure F5. Land clusters 14 – 18 from the 2012 – 2016 Sep/Oct/Nov global MERRA-2 dataset.

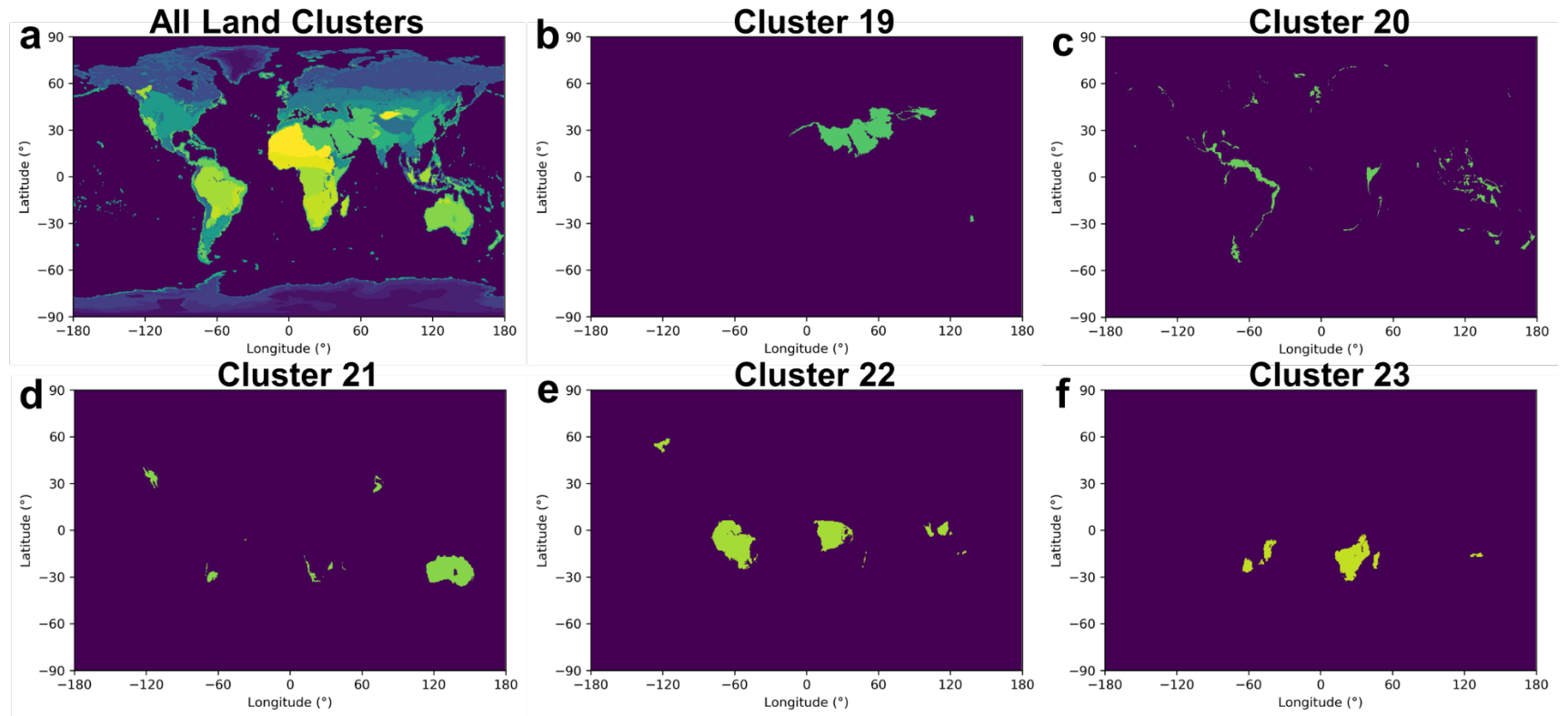


Figure F6. Land clusters 19 – 23 from the 2012 – 2016 Sep/Oct/Nov global MERRA-2 dataset.

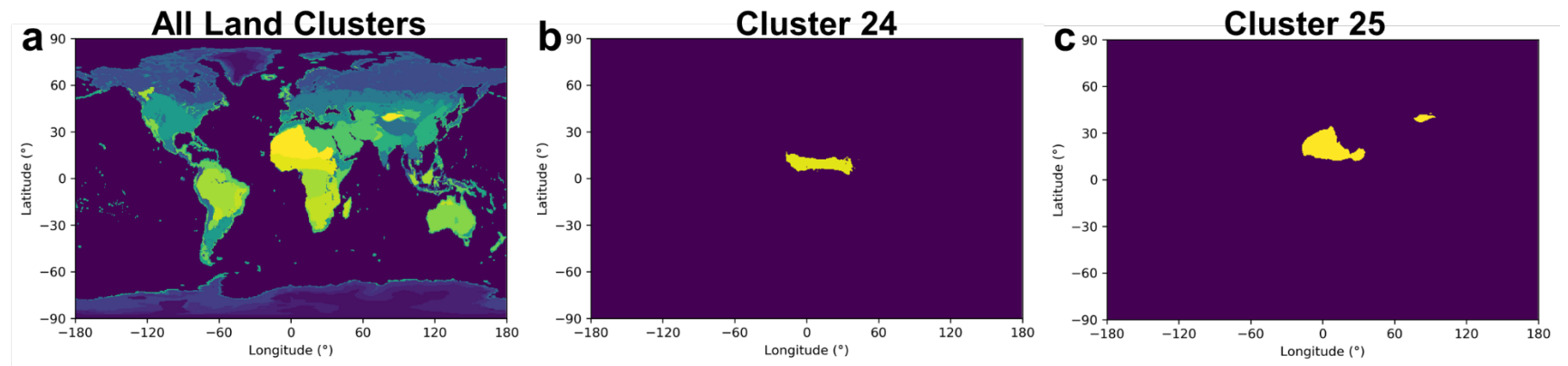


Figure F7. Land clusters 24 – 25 from the 2012 – 2016 Sep/Oct/Nov global MERRA-2 dataset.

G. Comparison of Cluster AOD Predictions with AERONET AOD

This section compares AERONET AOD statistics at 555, 551, and 532 nm over the respective time of year with the corresponding cluster AOD statistics. The locations of the AERONET sites compared is presented in Figure G1. We compared the average, 10th percentile, 50th percentile, and 90th percentile AOD in plots G2 through G6. The dotted line in each plot represents perfect correlation between experimental AOD and the cluster AOD_{550} derived from MERRA-2. The fractional error given in Table 4 in the main text is represented graphically as the horizontal distance from each point to the dotted line divided by the y-axis value of that point (AERONET AOD). We averaged this fractional error for all points in each set of statistics, then averaged the average fractional difference to arrive at the fractional error values in Table 4.

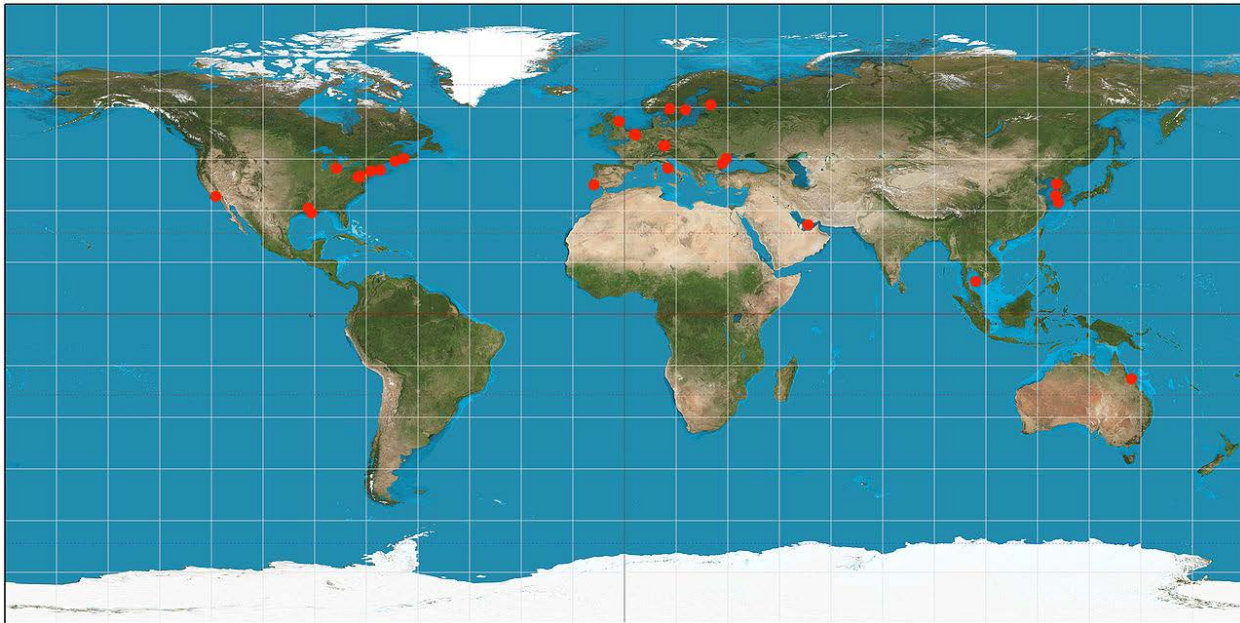


Figure G1. Locations of AERONET sites included in comparison with clusters.

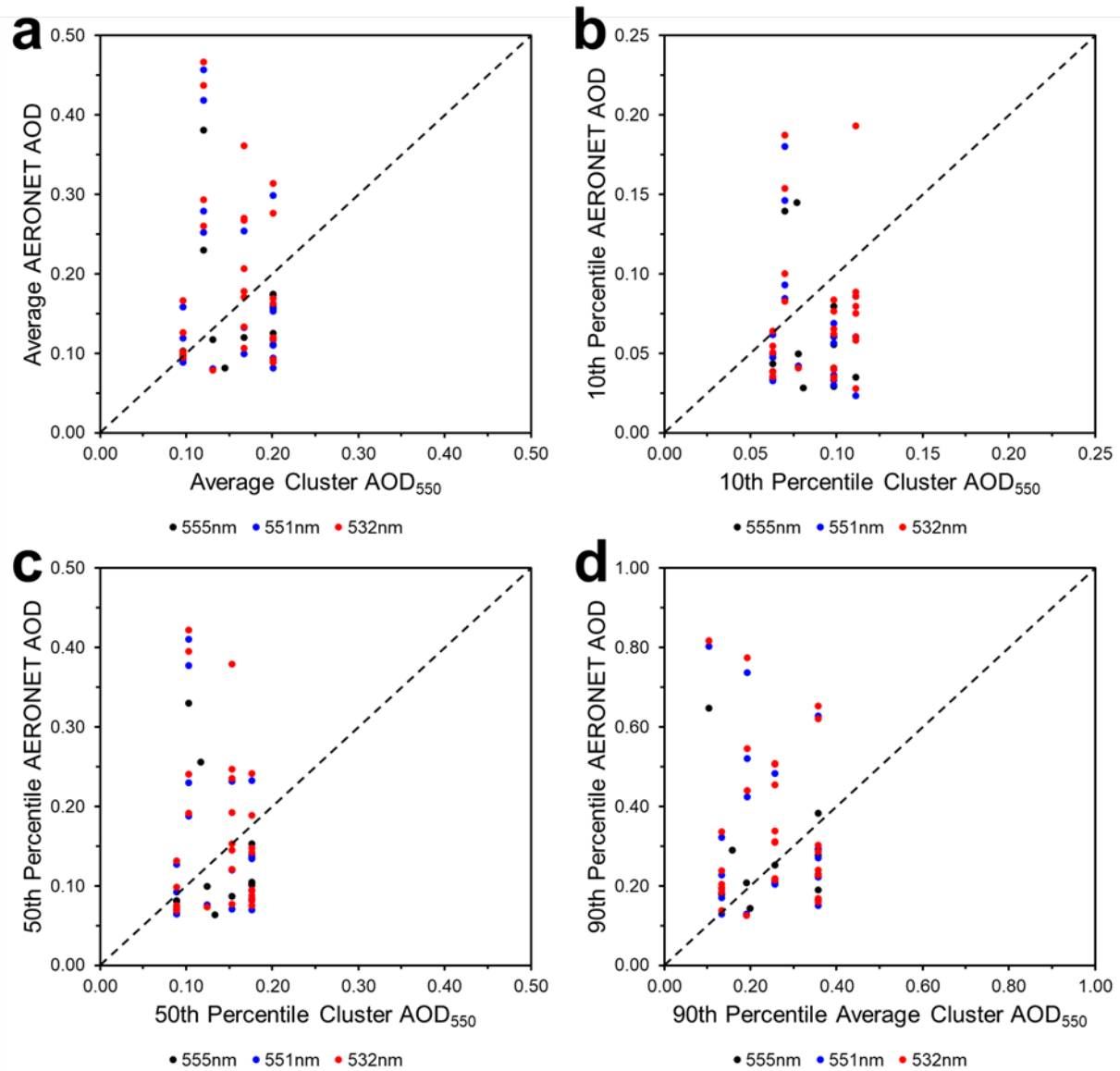


Figure G2. Annual AERONET AOD statistics vs. corresponding 2016 annual cluster AOD₅₅₀ statistics.

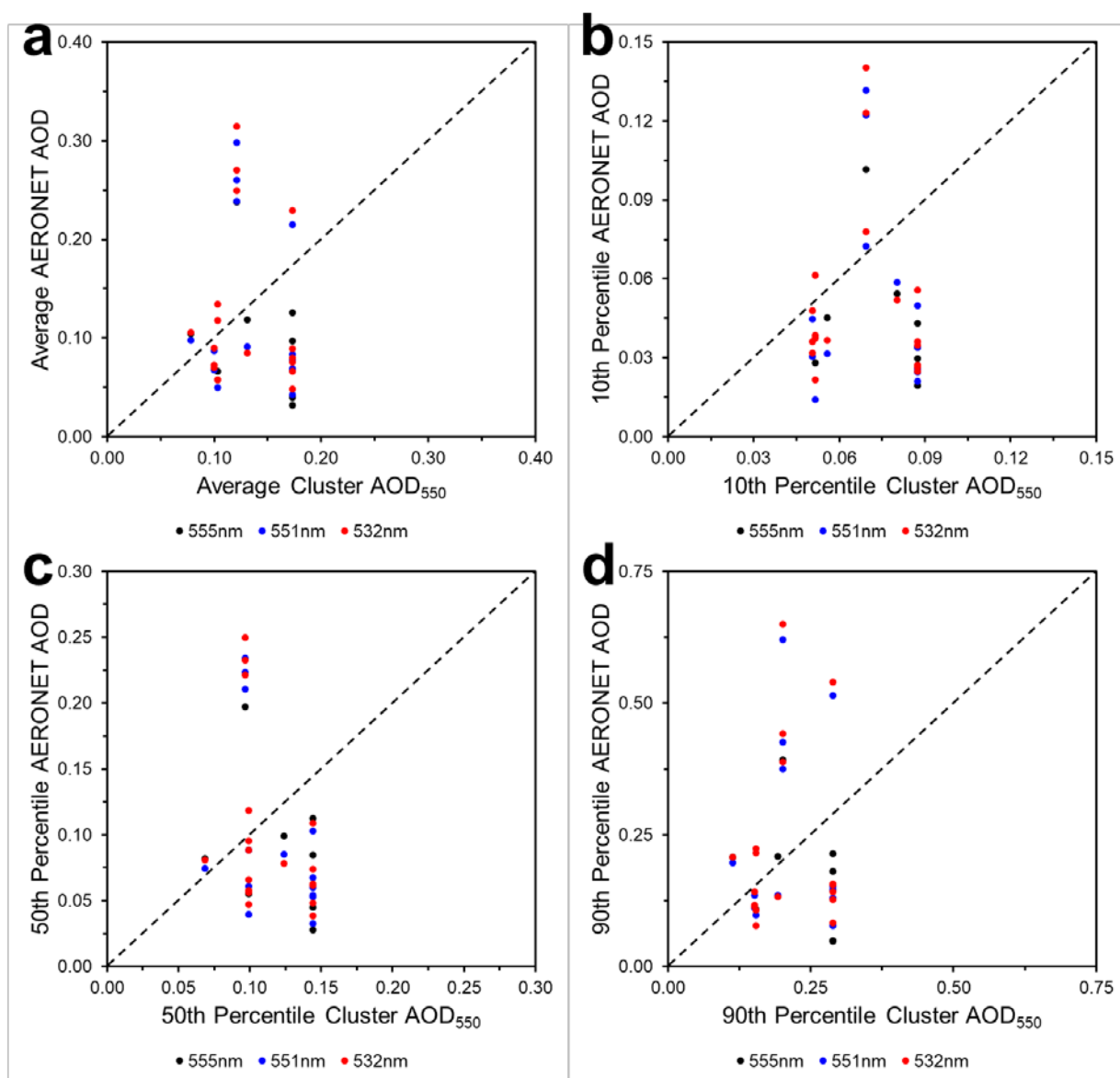


Figure G3. Dec/Jan/Feb AERONET AOD statistics vs. corresponding 2012 – 2016 Dec/Jan/Feb cluster AOD₅₅₀ statistics.

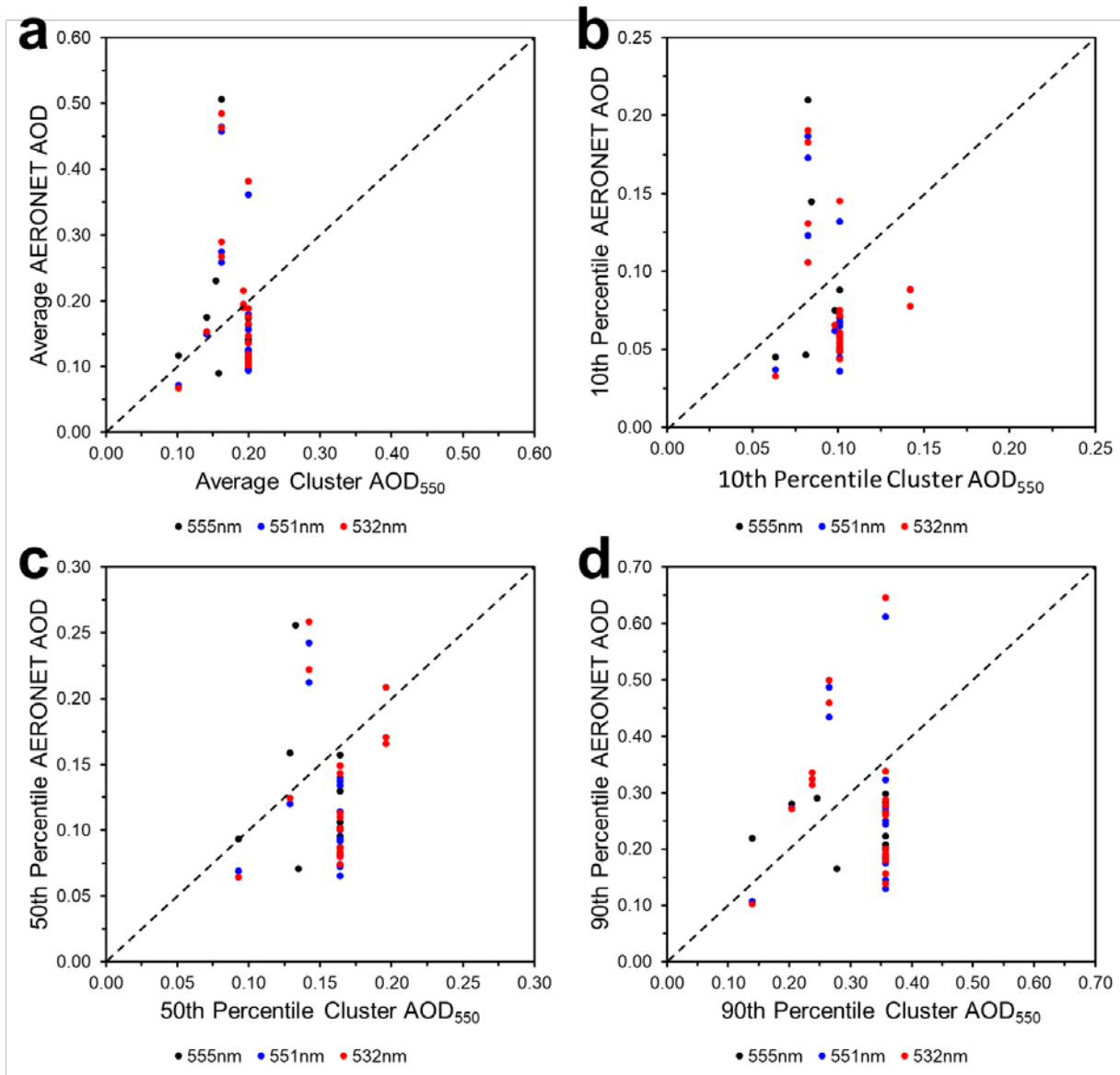


Figure G4. Mar/Apr/May AERONET AOD statistics vs. corresponding 2012 – 2016 Mar/Apr/May cluster AOD₅₅₀ statistics.

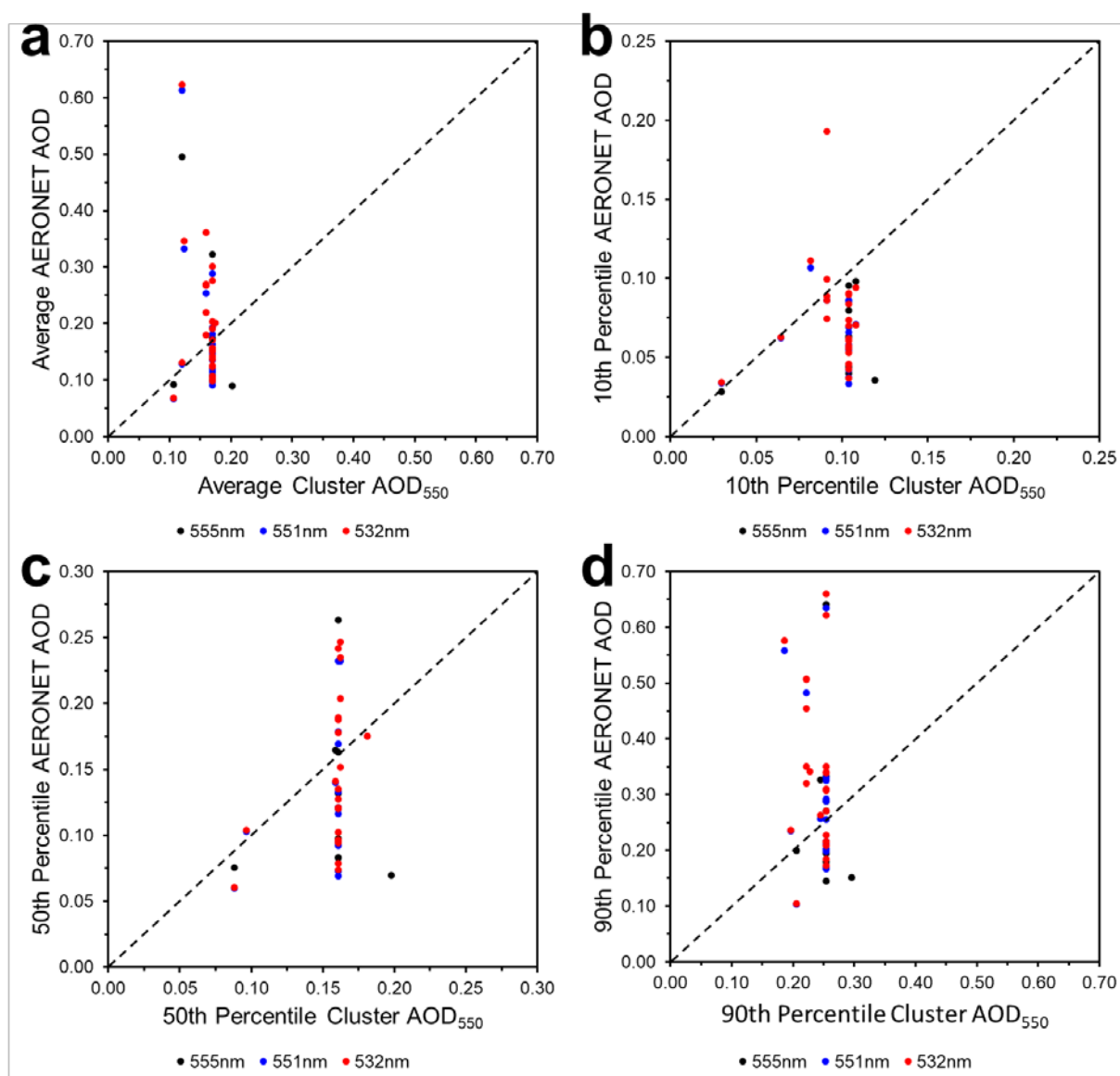


Figure G5. Jun/Jul/Aug AERONET AOD statistics vs. corresponding 2012 – 2016 Jun/Jul/Aug cluster AOD₅₅₀ statistics.

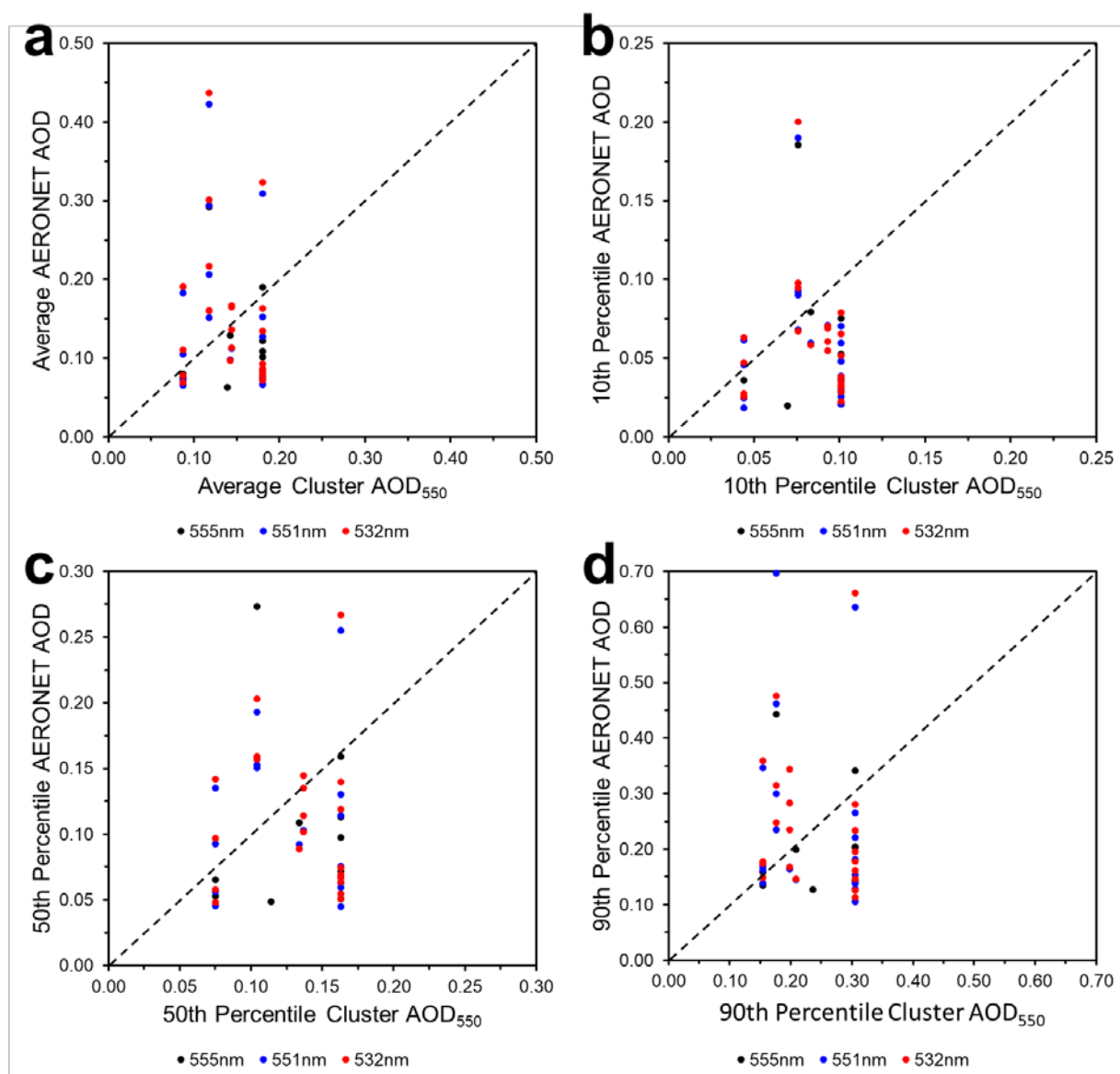


Figure G6. Sep/Oct/Nov AERONET AOD statistics vs. corresponding 2012 – 2016 Sep/Oct/Nov cluster AOD₅₅₀ statistic.

H. References

- Satopää, V., J. Albrecht, D. Irwin, and B. Raghavan, 2011: Finding a ‘kneedle’ in a haystack: Detecting knee points in system behavior. *2011 31st International Conference on Distributed Computing Systems Workshops*, 166–171, <https://doi.org/10.1109/ICDCSW.2011.20>
- Tolsa X., 2000: Principal values for the Cauchy integral and rectifiability. *Proceedings of the American Mathematical Society*, **128**, 7, 2111–2119, <https://doi.org/10.1090/S0002-9939-00-05264-3>
- Zhao, Q., V. Hautamaki, and P. Fränti, 2008: Knee point detection in BIC for detecting the number of clusters. *Advanced Concepts for Intelligent Vision Systems*, J. Blanc-Talon, S. Bourennane, W. Philips, D. Popescu, and P. Scheunders, Eds., Springer-Verlag, 664–673.

REPORT DOCUMENTATION PAGE				Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.					
1. REPORT DATE (DD-MM-YYYY) 04-23-2020		2. REPORT TYPE Technical Report		3. DATES COVERED (From - To)	
4. TITLE AND SUBTITLE Statistical Cluster Analysis of Global Aerosol Optical Depth for Simplified Atmospheric Modeling				5a. CONTRACT NUMBER HQ0034-14-D-0001	
				5b. GRANT NUMBER _____	
				5c. PROGRAM ELEMENT NUMBER _____	
6. AUTHOR(S) Noah T. Plymale (SED); Joshua E. Szekely (SED); Anna H. Rubinstein (N/A);				5d. PROJECT NUMBER DC-1-3994 (IDA)	
				5e. TASK NUMBER _____	
				5f. WORK UNIT NUMBER _____	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Institute for Defense Analyses 4850 Mark Center Drive Alexandria, Virginia 22311-1882				8. PERFORMING ORGANIZATION REPORT NUMBER 13184 H 2020-000163	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) Defense Threat Reduction Agency 8725 John J Kingman Rd #6201, Fort Belvoir, VA 22060				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER _____	
12. DISTRIBUTION / AVAILABILITY STATEMENT Distribution Statement A - Approved for public release; distribution is unlimited.					
13. SUPPLEMENTARY NOTES _____					
14. ABSTRACT Atmospheric aerosols originating from natural and anthropogenic sources have implications for modeling atmospheric phenomena, as aerosols conditions can change significantly and rapidly because depending on local geography and atmospheric conditions. A computational k-means clustering algorithm is applied to a global set of data obtained from the Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2) to yield a set of 25 clusters that discriminate based on land type, elevation, and atmospheric conditions to predict statistical aerosol optical depth (AOD) information. Analysis of annual 2016 MERRA-2 data in addition to data from four meteorological seasons for the five year period from 2012 through 2016 resulted in five separate sets of 25 clusters. The clustered AOD information is available with decision trees, qualitative cluster descriptions, and color-coded cluster maps to assist in identifying which cluster to use in retrieving AOD information. The results of this analysis have applications in atmospheric modeling where knowledge of approximate or typical aerosol conditions are needed in lookup table form without requiring access to large atmospheric databases or computationally intensive aerosol models.					
15. SUBJECT TERMS Aerosol Optical Depth (AOD); atmospheric aerosols; atmospheric modeling; MERRA-2; k-means clustering					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT Unclassified	b. ABSTRACT Unclassified	c. THIS PAGE Unclassified			Joanna Ingraham
			Unlimited	110	19b. TELEPHONE NUMBER (include area code) (571) 616-4208