remote sensing

MDPI

Article

Studying the Aerosol Effect on Deep Convective Clouds over the Global Oceans by Applying Machine Learning Techniques on Long-Term Satellite Observation

Xuepeng Zhao [1,*], James Frech [2], Michael J. Foster [3] and Andrew K. Heidinger [4]

[1] National Centers for Environmental Information (NCEI), NOAA/NESDIS, Silver Spring, MD 21910, USA
[2] Cooperative Institute for Satellite Earth System Studies (CISESS)/Earth System Science Interdisciplinary Center (ESSIC), University of Maryland, College Park, MD 20742, USA; james.frech@noaa.gov
[3] Space Science and Engineering Center (SSEC), University of Wisconsin—Madison, Madison, WI 53706, USA; mike.foster@ssec.wisc.edu
[4] Office of Geostationary Earth Orbit Observations, NOAA/NESDIS, Madison, WI 53706, USA; andrew.heidinger@noaa.gov
* Correspondence: xuepeng.zhao@noaa.gov; Tel.: +1-301-713-4842; Fax: +1-301-713-3300

Abstract: Long-term (1982–2019) satellite climate data records (CDRs) of aerosols and clouds, reanalysis data of meteorological fields, and machine learning techniques are used to study the aerosol effect on deep convective clouds (DCCs) over the global oceans from a climatological perspective. Our analyses are focused on three latitude belts where DCCs appear more frequently in the climatology: the northern middle latitude (NML), tropical latitude (TRL), and southern middle latitude (SML). It was found that the aerosol effect on marine DCCs may be detected only in NML from long-term averaged satellite aerosol and cloud observations. Specifically, cloud particle size is more susceptible to the aerosol effect compared to other cloud micro-physical variables (e.g., cloud optical depth). The signature of the aerosol effect on DCCs can be easily obscured by meteorological covariances for cloud macro-physical variables, such as cloud cover and cloud top temperature (CTT). From a machine learning analysis, we found that the primary aerosol effect (i.e., the aerosol effect without meteorological feedbacks and covariances) can partially explain the aerosol convective invigoration in CTT and that meteorological feedbacks and covariances need to be included to accurately capture the aerosol convective invigoration. From our singular value decomposition (SVD) analysis, we found the aerosol effects in the three leading principal components (PCs) may explain about one third of the variance of satellite-observed cloud variables and significant positive or negative trends are only observed in the lead PC1 of cloud and aerosol variables. The lead PC1 component is an effective mode for detecting the aerosol effect on DCCs. Our results are valuable for the evaluation and improvement of aerosol-cloud interactions in the long-term climate simulations of global climate models.

Keywords: aerosol indirect effect (AIE); aerosol optical thickness (AOT); deep convective cloud (DCC); aerosol-cloud interaction (ACI); satellite observation; machine learning (ML); artificial intelligence (AI)

Citation: Zhao, X.; Frech, J.; Foster, M.J.; Heidinger, A.K. Studying the Aerosol Effect on Deep Convective Clouds over the Global Oceans by Applying Machine Learning Techniques on Long-Term Satellite Observation. *Remote Sens.* **2024**, *16*, 2487. https://doi.org/10.3390/rs16132487

Academic Editors: Michal Segal-Rosenheimer and Haochi Che

Received: 13 May 2024
Revised: 1 July 2024
Accepted: 4 July 2024
Published: 7 July 2024

1. Introduction

Deep convective clouds (DCCs) play an important role in the hydrological and energy cycles associated with atmospheric circulations, as well as in regional and local weather and climate systems [1,2]. The formation and development of DCCs involve complicated and entangled dynamical, thermodynamic, and micro-physical processes. As such, the potential impacts of globally increased anthropogenic aerosols on DCCs since preindustrial times are extremely complex and widely debated and are still the most actively studied subject of aerosol cloud interactions (ACIs) [3–9].

One of the long-standing challenges in ACI studies is to separate the aerosol effect on DCCs from the covariance of meteorological conditions, especially from an observational

perspective. Model simulations are frequently used in the literature for disentangling the aerosol effect from the dynamical and thermodynamic effects in aerosol and DCC interactions, e.g., [10–12]. Recently, due to the fast development of artificial intelligence (AI) technology, machine learning (ML) techniques have begun to be applied in the study of aerosol cloud interactions. For example, Zang et al. [13] used the artificial neural network technique to separate the primary aerosol convective invigoration (PAI) from the covarying meteorology-aerosol invigoration (MAI) effect on tropical DCCs over land areas by quantifying the sensitivity of DCC properties to aerosol changes under fixed meteorological conditions. They found that PAI and MAI contribute comparably in magnitudes to the aerosol invigoration effect on tropical DCCs over land.

In this paper, we use AI-based techniques along with nearly 40 years of global satellite climate data records (CDRs) of DCC and aerosol index (AIX) to study the aerosol effect on DCCs from a global and long-term observational perspective. The objective is to identify potential signatures or imprints of the aerosol effect on DCCs over the global oceans by disentangling the aerosol micro-physical effect from the covariance of meteorological conditions, with the help of an artificial neural network technique, in the analysis of long-term averaged (or climatology of) aerosol and cloud variables from operational satellite observations. The results will be beneficial for validating global long-term climate model simulations of the interactions between aerosols and DCCs, which are still plagued with large uncertainties in the ACI treatment for DCCs in global climate models.

2. Data

This study uses three long-term data products with climate quality. The first is version 4.0 (v4.0) of the Advanced Very High-Resolution Radiometer (AVHRR) aerosol optical thickness (AOT) CDR from the National Oceanic and Atmospheric Administration (NOAA) Polar Operational Environmental Satellites (POES) and the European Meteorological Operational (MetOp) satellites. The second is version 6.0 (v6.0) of the NOAA AVHRR + High-Resolution Infrared Sounder (HIRS) Pathfinder Atmospheres-Extended (PATMOS-x) cloud CDR product. The third is the climate forecast system reanalysis (CFSR) product of the National Centers for Environmental Prediction (NCEP). They are described briefly in the following.

2.1. Satellite Data

AVHRR AOT CDR v4.0 is the first satellite data used. It is available from the NOAA CDR website (https://www.ncei.noaa.gov/products/climate-data-records/avhrr-aerosol-opticalthickness, accessed on 1 January 2022), which is maintained and archived by the NOAA National Centers for Environmental Information (NCEI). AVHRR AOT is derived over the global ocean surface for λ_1 = 0.63 μm and λ_2 = 0.86 μm channels using a two-channel AVHRR aerosol retrieval algorithm [14] on AVHRR clear-sky daytime reflectance. The clear-sky reflectance is determined by the PATMOS-x AVHRR + HIRS v6.0 all-sky reflectance and cloud probability CDR products [15]. The data time period spans from 1982 to the present, and the spatial resolution is 0.1° × 0.1° on the equal-angle latitude/longitude grid. AOTs derived for λ_1 = 0.63 μm (τ_1) and λ_2 = 0.86 μm (τ_2) channels are used to calculate the aerosol Angström exponent α ($\alpha = -[\ln(\tau_1/\tau_2)/\ln(\lambda_1/\lambda_2)]$), which is used to further determine the aerosol index (AIX) defined as AIX = $\alpha \times \tau_1$. AIX from 1982 to 2019 will be used in our analysis since it is a better proxy than AOT for column aerosol number concentration [16–18].

The second satellite product used is AVHRR + HIRS PATMOS-x cloud CDR v6.0. It is a level-2b product on 0.1° × 0.1° equal-angle latitude/longitude grid [19,20], which is retrieved using the inter-calibrated AVHRR global area coverage (GAC) radiances collocated with the HIRS radiances on the same satellite platforms of NOAA POES and the European Organization for the Exploitation of Meteorological Satellites (EUMETSAT)/MetOp [21–23]. The intercalibrated and collocated AVHRR + HIRS GAC radiances are also cross-calibrated with more advanced National Aeronautics and Space Administration (NASA) moderate

resolution imaging spectroradiometer (MODIS) radiances [24–26]. The data can be downloaded from the NOAA CDR website (https://www.ncei.noaa.gov/products/climate-data-records/avhrr-hirs-cloud-properties-patmos, accessed on 5 January 2022). Six cloud variables are available and used in our analysis: cloud particle effective radius (CPER), cloud optical depth (COD), ice water path (IWP), cloud cover fraction (CCF), cloud top height (CTH), and cloud top temperature (CTT). These variables are output together with the inter- and cross-calibrated and collocated AVHRR + HIRS radiances along with selected ancillary data (e.g., surface types) as the level-2b daily CDR products, which allow other CDRs to be derived and generated from the PATMOS-x CDR data for more applications (such as the AOT CDR introduced above). Ice phase clouds are identified by cloud type flags in PATMOS-x cloud CDR [23,27]. In the current study, our focus is on DCCs which are determined from PATMOS-x ice phase clouds through a two-step filtering [28]. Only the data after 1982 is used in this study, as there are many missing observations in the first three years (1979–1981) of collocated AVHRR + HIRS data records. Thirty-eight years (1982–2019) of daily products from the AVHRR aerosol index and PATMOS-x AVHRR + HIRS cloud CDRs are averaged to obtain both monthly and long-term mean values, which are used in our study.

2.2. Reanalysis Data

The third dataset used in our analysis is the NCEP CFSR monthly mean product (ftp://nomads.ncdc.noaa.gov/CFSR/HP_monthly_means/, accessed on 5 January 2022) with a latitude and longitude resolution of 0.5° × 0.5°. Specifically, we selected 19 meteorological fields from the CFSR reanalysis, which are relevant and important to the formation and development of DCCs. The selected meteorological fields include surface convective available potential energy (CAPE) in J/kg, precipitable water of atmospheric column (PW) in kg/m^2, relative humidity (RH) in percentage in atmospheric column (RH_{clm}), at 850 mb and 400 mb pressure levels (RH_{850} and RH_{400}), and at 2 m altitude (RH_{2m}), vertical velocity (ω) in Pa/s at 850 mb and 400 mb pressure levels (ω_{850} and ω_{400}) and at 0.995 sigma vertical coordinate level (ω_{sig995}), U and V components of horizontal wind vectors in m/s at 10 m altitude (U_{10m} and V_{10m}), at 850 mb (U_{850} and V_{850}), and 400 mb (U_{400} and V_{400}) pressure levels, and atmospheric temperature in K at 850 mb and 400 mb pressure levels (T_{850} and T_{400}) and at 2 m altitude (T_{2m}). U and V components at 700 mb and 400 mb pressure levels are used to compute the vertical shear of horizontal wind (VSHW) between the two pressure levels for the middle troposphere using $VSHW = [(U700 - U400)^2 + (V700 - V400)^2]^{1/2}/(700 - 400)$. NCEP CFSR was designed and executed as a global, high-resolution, coupled atmosphere–ocean–land surface–sea ice system to provide the best estimate of the state of these coupled domains over the period from 1979 to present [29]. The selected meteorological variables from CFSR monthly mean products are averaged from January 1982 to December 2019 in order to obtain long-term mean values (or climatology). Both monthly and long-term averaged values of these meteorological variables are interpolated into the same spatial resolution (0.1° × 0.1°) as the above satellite cloud and aerosol CDR products. Twenty-six variables (1 aerosol, 6 cloud, and 19 meteorology) introduced above and used in the following analysis are summarized in Table 1.

Table 1. One aerosol, six cloud, and nineteen meteorological variables used in our analyzing study.

	Variables	Note
Aerosol	AIX	Satellite Observation
Cloud	CPER, COD, IWP, CCF, CTH, CTT	Satellite Observation
Meteorology	CAPE, PW, RH_{clm}, RH_{850}, RH_{400}, RH_{2m}, T_{850}, T_{400}, T_{2m}, U_{850}, U_{400}, U_{10m}, V_{850}, V_{400}, V_{10m}, ω_{850}, ω_{400}, ω_{sig995}, VSHW	CFSR Reanalysis

3. Approaches of Analysis

First, we used the SHapley Additive exPlanation (SHAP) approach [30–32], an explainable machine learning technique, to explore the nonlinear correlation among the six cloud variables (CPER, COD, IWP, CCF, CTH, and CTT) as well as their nonlinear correlation with AIX and the selected 19 CFSR meteorological variables. Specifically, we used Python's Extreme Gradient Boosting (XGBoost) nonlinear regression model for prediction and used the SHAP Python analysis package for interpretation [33–35]. This combination is powerful for understanding machine learning-based nonlinear regression model prediction and can tell us how each input variable has contributed to an individual prediction. By aggregating SHAP values, we can also understand trends across multiple predictions. Here, SHAP values refer to Shapley values applied to a conditional expectation function of a machine learning regression mode. The essence of a Shapley value is to measure the contributions to the final outcome from each player separately among the coalition while preserving the sum of contributions as equal to the final collective outcome. For a machine learning model, this means that SHAP values of all the input variables will always sum up to the difference between the baseline (expected) model output and the current model output for the prediction being explained.

The objective of our SHAP analysis is to identify the potential aerosol effect on DCCs (simply called the aerosol indirect effect [AIE] hereafter). DCCs appear more frequently over the three latitude belts over oceans as shown in Figure 1: the tropical latitude belt (TRL) of 15°S–15°N, the northern middle latitude belt (NML) of 30°N–60°N, and the southern middle latitude belt (SML) of 30°S–60°S. CCF shows the highest value over the tropical convergence zones in TRL due to strong convections. There are also relatively high CCF values in the NML and SML storm corridors over the Northern Pacific Ocean, the Northern Atlantic Ocean, the Southern Indian Ocean, and the Southern Atlantic Ocean, where mesoscale convective cloud systems (MCS) are active. CCF values are relatively low in the subtropical latitudes of both hemispheres, prevailing with subsidence motions. Thus, the SHAP analysis that is able to interpret the nonlinear dependence of correlated multiple variables was applied first to the above three latitude belts over the oceans to find the AIE signature of DCCs.

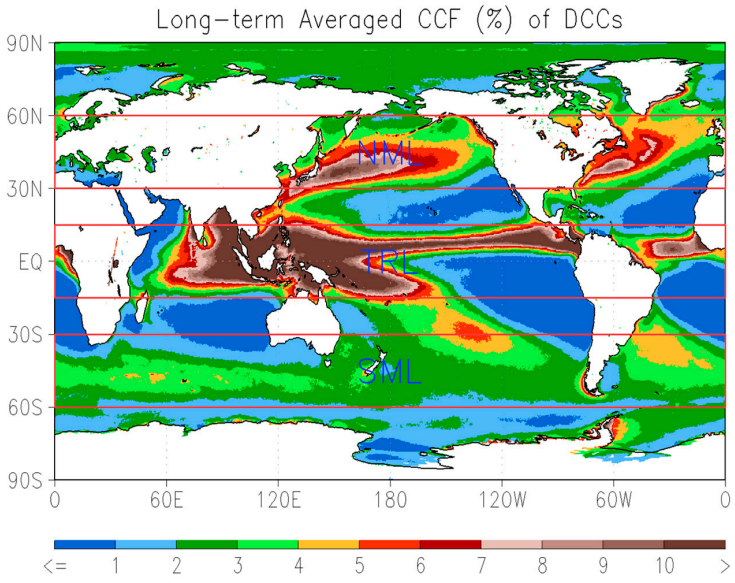

Figure 1. Distributions of cloud variable cloud cover fraction (CCF) of deep convective clouds (DCCs) for its long-term (1982–2019) averaged monthly mean values over the global oceans. The three latitude belts, the tropical latitude belt (TRL) of 15°S–15°N, northern middle latitude belt (NML) of 30°N–60°N, and southern middle latitude belt (SML) of 30°S–60°S, where DCCs appear more frequently, are marked by the three long rectangle boxes in red color.

A statistical analysis was then performed for the three latitude belts to derive the relationships of cloud variables with AIX and meteorological variables and to find the sensitive regime of AIX for the AIE of DCCs. To better explain the statistical relationships of cloud variables with AIX and meteorological variables, a back-propagation neural network (BPNN) model was employed to examine the AIE and the covariance of meteorological conditions, which still plagues the ACI studies based on satellite observations. Specifically, the BPNN model is used to perform nonlinear fitting in this study. At present, the BPNN model is commonly used to resolve complicated and non-linear problems in many fields due to its adaptive and self-learning abilities [13,36–38]. Thus, we developed this simple BPNN model to fit the changing relationship of a cloud variable with the increase of AIX. The relationship is determined by binning the cloud and meteorological variables according to AIX with 0.01 incremental interval. Since the BPNN model is specifically used to fit the changing relationship of a cloud variable with the increase of AIX, all the available data are used for training. We invested considerable time into optimizing the BPNN model, including selecting activation functions, tuning hyperparameters, and examining overfitting. This BPNN model includes one input layer, one hidden layer, and one output layer as shown in Figure 2. Tanh and Sigmoid activation functions are used for input and output layers, respectively. The learning rate, which is a hyperparameter of the BPNN model, is set at 0.5 based on optimal tests for minimizing the cost function of the nonlinear fit. There are 20 input variables (AIX and the 19 CFSR meteorological fields introduced in the above Section 2) and the predicted output is one of the six cloud variables of DCC. The hidden layer includes eight nodes. The training was performed with 15,000 iterations to achieve a mean squared error (MSE) of fitting less than 10^{-5}.

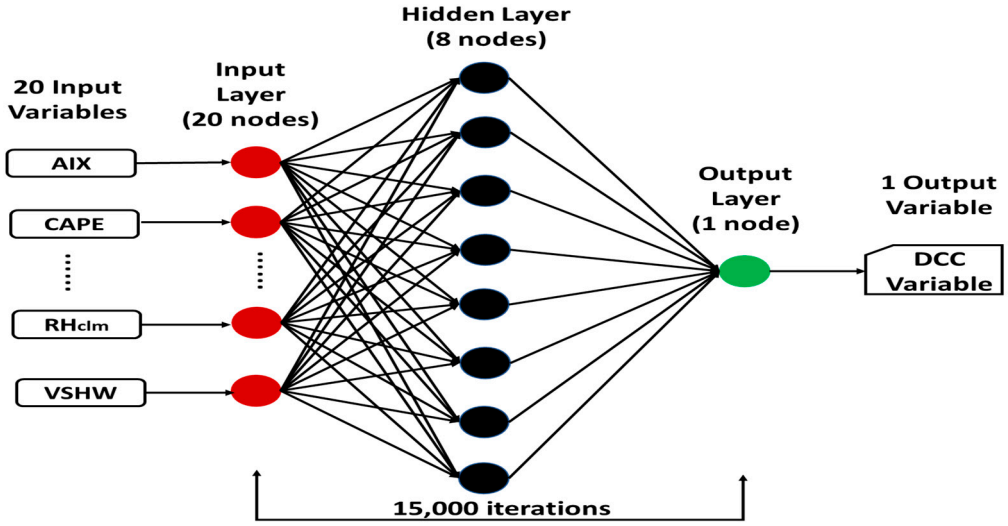

Figure 2. Schematic diagram of the three-layer (input, hidden, output) back-propagation neural network (BPNN) model.

Last, we performed singular value decomposition (SVD) analysis [39–41] on six cloud variables (CPER, COD, IWP, CCF, CTH, and CTT) versus AIX, respectively. SVD decomposes a matrix into three matrices: U (left singular vectors), Σ (singular values), and V (right singular vectors), revealing the principal components (PCs) which are the directions of maximum variance in the data. These components are used for dimensionality reduction, noise reduction, and data interpretation. Since SVD examines the coupled variability of two fields (e.g., CPER and AIX), it will identify those modes of behavior in which the variations of the two fields are strongly coupled. Thus, the aerosol effect of cloud variables can be examined for the major coupled modes, which are also named as the major principal components of SVD analysis. Although SVD analysis has been widely used in the study of large-scale climate dynamics e.g., [39–41], to our knowledge, this is the first time it has been used to analyze the climate effects of aerosol and cloud interactions primarily at a

small scale. The annual mean values of AIX and the six cloud variables from 1982 to 2019 over the global oceans were used in the SVD analysis.

4. Results

4.1. Correlation Relationship

Figure 3 shows a nonlinear correlation matrix from the SHAP analysis for the annual mean values of six DCC variables (CPER, COD, IWP, CCF, CTH, and CTT) from 1982 to 2019 over the three latitude belts defined in Figure 1. We use the annual mean values of the cloud variables rather than the monthly mean values because the former is much less noisy than the latter. The three micro-physical variables (CPER, COD, IWP) show the highest positive/negative correlation (or P-/N-Corr. hereafter) in TLR, which is slightly reduced in NML and is further reduced in SML. The N-Corr. of COD and IWP in TRL and NML turns to small P-Corr. in SML. For the three macro-physical variables (CCF, CTH, CTT), CTH and CTT show extremely high (−1.0), high (−0.9), and moderate (−0.7) N-Corr. in TRL, NML, and SML, respectively, which suggests that convection is the strongest in TRL and gradually reduced in NML and SML since a stronger convection results in a higher DCC convective tower which has a colder cloud top. CCF shows relatively low P-/N-Corr. with CTH/CTT, and the absolute values of the correlation coefficients are generally less than 0.4 and gradually decrease from TRL to SML and NML. Between the micro- and macro-physical cloud variables, CPER, COD, and IWP show moderate P-/N-Corr. with CCF in TLR but low P-/N-Corr. in SML and NML, and the absolute values of the correlation coefficients are generally less than 0.5. CPER and IWP show relatively high, moderate, and low P-/N-Corr. with CTH/CTT in SML, NML, and TLR, respectively. The correlation coefficients of COD with CTH/CTT are −0.3/0.3 in TLR, 0.1/−0.2 in NML, and 0.2/0.07 in SML. These correlation information between the cloud variables will facilitate our interpretation of the subsequent results from the SHAP, BPNN, and SVD analyses.

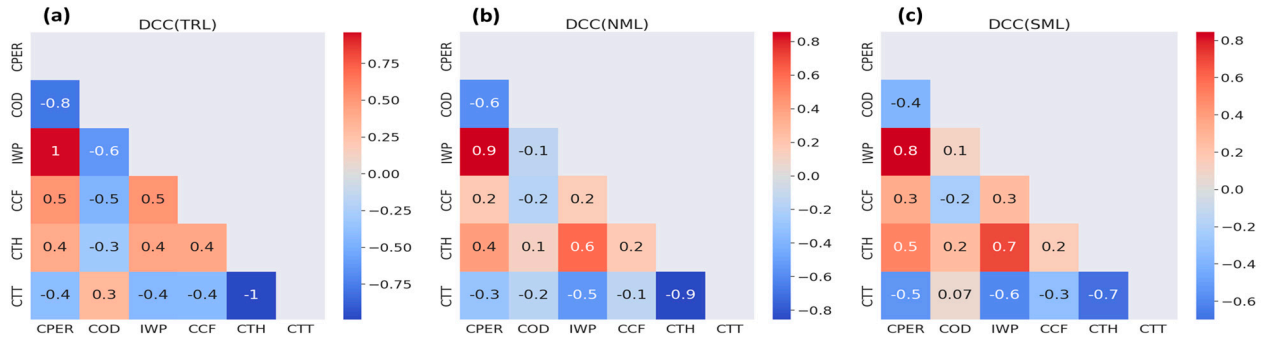

Figure 3. Nonlinear correlation matrix of six deep convective cloud (DCC) variables (cloud particle effective radius [CPER], cloud optical depth [COD], ice water path [IWP], cloud cover fraction [CCF], cloud top height [CTH], and cloud top temperature [CTT]) in the three latitude belts defined in Figure 1. The numbers are correlation coefficients.

For the prediction of a cloud variable (e.g., CPER) from the XGBoost nonlinear regression predictive model with 20 input variables (AIX and the 19 meteorological variables listed in Table 1), we calculated SHAP value for each input variable (which can be negative or positive for each of the variables) across all the grid points for their annual mean data from 1982 to 2019. Then, we took the mean of the absolute SHAP values, as we did not want positive and negative values to offset each other, and plotted them in Figure 4, which shows three bar plots of absolute SHAP values of 20 input variables for CPER prediction in the three latitude belts defined in Figure 1. There is one bar for each input variable. A longer bar means that its corresponding variable is more important for the prediction of cloud variable CPER. For NML (Figure 4a), the three most important variables for the CPER prediction are AIX, VSHW, and RH_{850}. The most significant contribution is from AIX, which suggests that AIE should be able to manifest easily in the changes of CPER for

marine DCCs of MCS and should not be obscured by meteorological covariances. For TRL (Figure 4b), the three most important variables are CAPE, ω_{400}, and RH_{400}, which suggests that AIE has difficulty manifesting in the changes of CPER for tropical marine DCCs, as the meteorological covariances associated with strong tropical convections can easily conceal the aerosol effect on the DCCs. For SML (Figure 4c), the three most important variables are PW_{clm}, VSHW, and CAPE, which also suggests that AIE has difficulty manifesting in the changes of CPER for marine DCCs of MCS over the southern middle latitude oceans, where moisture is abundant and aerosol loading is lower than in the northern hemisphere (NH), its counterpart. Thus, meteorological covariances may easily obscure AIE on the CPER of DCCs over SML.

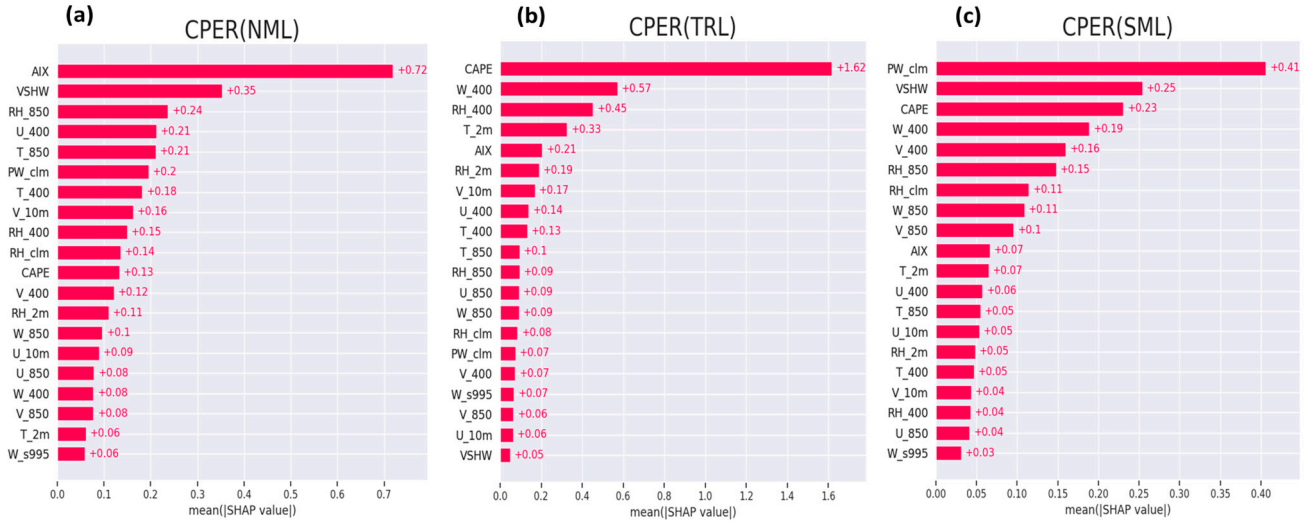

Figure 4. Bar plots of absolute SHapley Additive exPlanation (SHAP) values of 20 variables for cloud particle effective radius (CPER) prediction in the three latitude belts defined in Figure 1: (**a**) northern middle latitude (NML), (**b**) tropical latitude (TRL), and (**c**) southern middle latitude (SML).

Since AIE may only manifest in NML, we will focus our following analyses on NML. We also performed SHAP analysis for the other five cloud variables (COD, IWP, CCF, CTH, and CTT) in addition to CPER, and the three most important input variables for the prediction of the five cloud variables on NML are summarized in Table 2 together with those for CPER. We can see that the AIE signature may still be able to manifest for the other two observed cloud micro-physical variables (COD and IWP), albeit not as evidently as it does for CPER, since AIX is the second and third important variable for the prediction of COD and IWP, respectively. AIE can be easily obscured by meteorological covariances for the three macro-physical cloud variables (CCF, CTH, and CTT) since AIX is not even one of the three most important variables. Wind and moisture fields in the upper troposphere (UT) are most important for the prediction of CCF. Temperature, wind, and moisture fields in UT are critical for the prediction of CTH. Temperature field in UT, convective energy and moisture supply near the surface are critical for the CTT prediction.

Figure 5 shows the waterfall plot of SHAP values for the CPER in NML, which adds more information compared to the bar plot of Figure 4. The 20 arrow bars represent E[f(x)] − f(x) for the 20 individual variables. For example, AIX increased the predicted CPER SHAP value by +0.66 but T_{400} reduced the predicted CPER by −0.38. The difference between E[f(x)] and f(x) equals the summation of the SHAP values of the 20 variables. The differences between the SHAP values tell us how the variables have contributed collectively to the prediction of CPER when compared to their averaged prediction. A large positive/negative SHAP value difference indicates that the corresponding variable had a significant positive/negative impact on the CPER prediction. We can see AIX and CAPE have the most important positive contribution to the predicted CPER while T_{400} provides

the most important negative contribution, which is newly added information compared to Figure 4.

Table 2. The three most important input aerosol and meteorological variables in the SHapley Additive exPlanation (SHAP) analysis and their corresponding absolute mean SHAP values for the six cloud variables in the northern middle latitude (NML). The aggregated SHAP value of the 20 variables for each cloud variable is also listed.

Cloud Variable (Aggregated SHAP Value)	1st Important (SHAP Value)	2nd Important (SHAP Value)	3rd Important (SHAP Value)	Notes
CPER (3.26)	AIX (0.72)	VSHW (0.35)	RH_{850} (0.24)	AIE may manifest easily
COD (5.07)	T_{850} (1.04)	AIX (0.44)	RH_{400} (0.43)	AIE may not manifest easily
IWP (60.49)	RH_{400} (7.87)	VSHW (7.51)	AIX (6.20)	AIE may not manifest easily
CCF (0.02307)	V_{400} (0.00406)	RH_{400} (0.00403)	U_{400} (0.00395)	AIE may be concealed
CTH (0.63)	T_{400} (0.52)	U_{400} (0.11)	RH_{400} (0.10)	AIE may be concealed
CTT (2.93)	T_{400} (0.53)	CAPE (0.44)	RH_{2m} (0.31)	AIE may be concealed

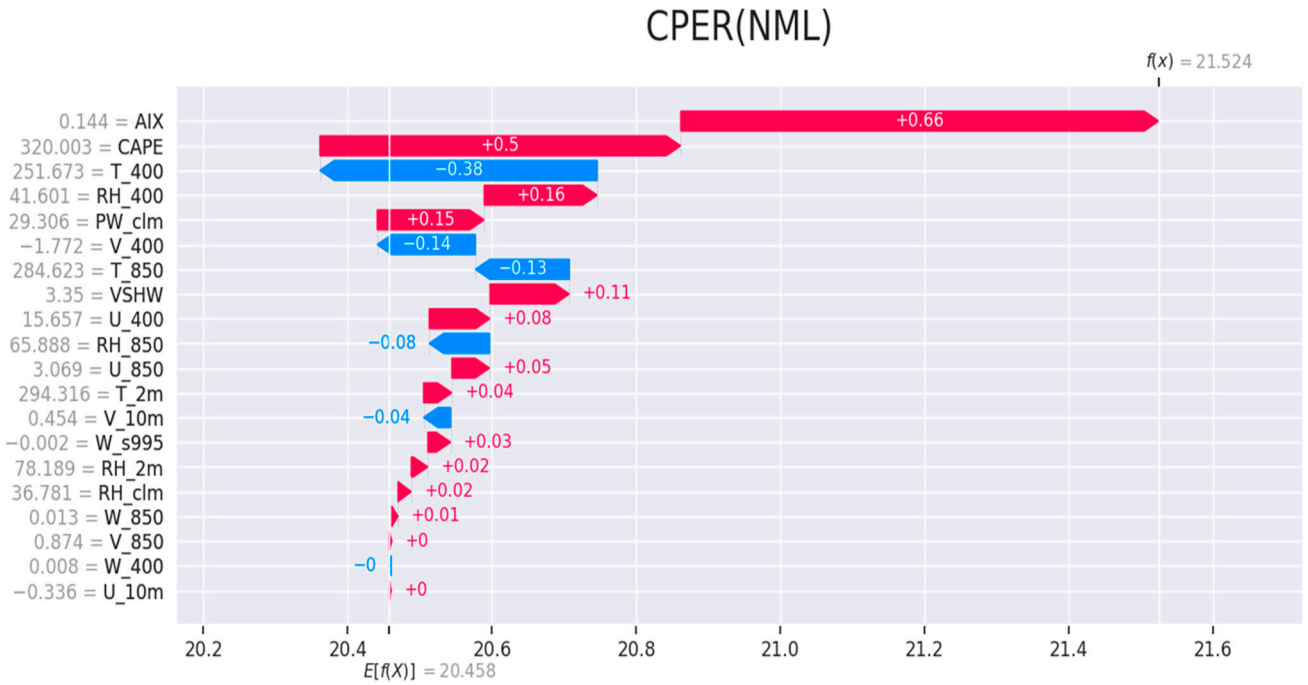

Figure 5. Waterfall plot of SHapley Additive exPlanation (SHAP) values for cloud particle effective radius (CPER) prediction in the northern middle latitude (NML). E[f(x)] = 20.458 is the average (or baseline value) of predicted values of CPER from 20 variables for their individual mean values (given in the caption of vertical coordinate) over all grid points. f(x) = 21.524 is the collectively predicted value of CPER from 20 variables over all grid points. The arrow bars show the differences between E[f(x)] and f(x) for individual variables. The difference between E[f(x)] and f(x) equals the summation of 20 SHAP value differences of the variables.

Figure 6 presents a heatmap of SHAP values of the 20 variables over all the grid points in NML for CPER prediction, which provides information for individual grid points (instances) as well as the mean value of all the grids (dark bars on the right side). The f(x) curve is the predicted CPER over all the grid points. We can see that AIX makes the most important positive contribution for CPER prediction over all the grid points in NML. There seem to exist two regimes, with the first one on the left side at about instance 800 and the second one on the right. The SHAP values for the meteorological variables are very different in these two regimes (most of their positive/negative signs are even reversed), but AIX is nearly always the most important positive contributor in these two regimes, which suggests that the AIE of DCCs should be able to manifest easily in the CPER change over the oceans of NML. Even though we still do not know why there are two regimes for most of the meteorological variables, this feature suggests that the effect of most of the meteorological variables on CPER is not monotonic and can be positive or negative, while the effects of AIX on CPER are mainly positive. Further in-depth analysis is needed in our future work to find out the exact causes of the two regimes.

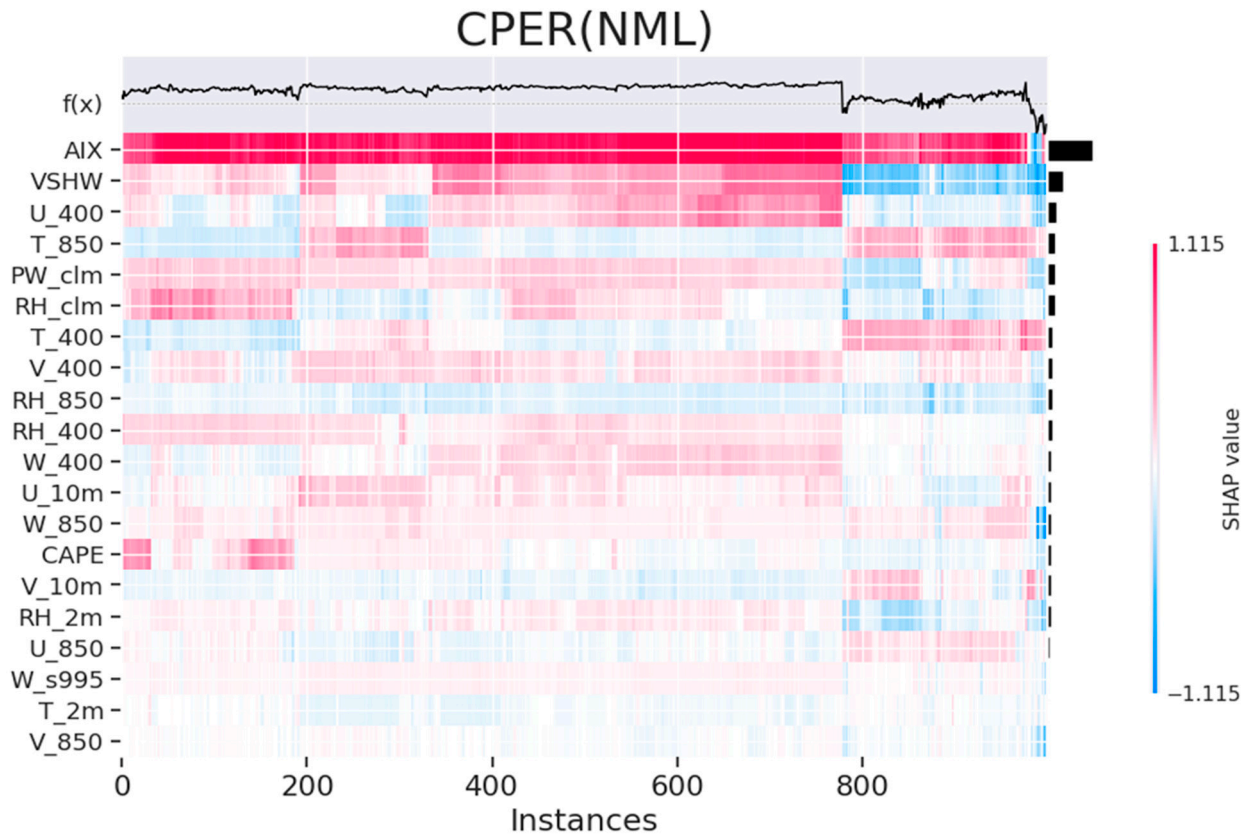

Figure 6. Heatmap of SHapley Additive exPlanation (SHAP) values of 20 variables over all the grid points (instances) in the northern middle latitude (NML) for cloud particle effective radius (CPER) prediction, which provides information for individual instances as well as their mean value for all the instances (dark bars on the right side). f(x) line is the predicted CPER over all the grids. Color bar on the right-hand side indicates the sign and magnitude of SHAP values.

4.2. Separating Entangled Effects in the Sensitive Regime of AIE

To find out how cloud variables change with the increase of aerosol loading, we binned the six cloud variables (CPER, IWP, COD, CCF, CTT, and CTT) according to AIX with 0.01 incremental intervals. The first AIX bin, which is different from the other AIX bins, is specifically set from 0 to 0.05 in order to contain sufficient sample numbers. The mean values for individual bins are computed and displayed as the black line with dots for CPER, CCF, and CTH in Figure 7 for NML as an example. The black vertical bars on

the black dots are the corresponding standard errors for each bin. The sampling numbers for individual bins are also displayed with a dashed black curve in Figure 7a. We can see the sample numbers are large enough that the standard errors are very small, except in the bins at the two end sides, where the sample numbers are smaller than about 200 and caution needs to be taken when interpreting the observed features. However, the standard deviation (not shown here to avoid a busy plot) of individual bins can be much larger than its standard error. This is because the standard deviation in a bin represents the spread of the large sample numbers in the bin while the corresponding standard error shows the range in which the average value falls with high probability. We can see the large variations of the cloud variables are mainly observed between 0.05 and 0.32 of AIX and most of the sample numbers also fall in this AIX range. Thus, the range of AIX from 0.05 to 0.32 is the potential regime (or sensitive regime) where the aerosol effect on DCCs might manifest. We can also see there is an AIX threshold (~0.18), below which CPER increases with AIX increase and above which CPER decreases with AIX increase until reaching AIX~0.32. Then, CPER levels off with fluctuations due to relatively small sample numbers. The variation tendency of IWP with AIX increase is very similar to that of CPER due to their strong positive correlation in the NML as shown in Figure 3b. For a similar reason, the variation tendency of COD with AIX increase is inversely similar to that of CPER, as COD and CPER are clearly anticorrelated (see Figure 3b). For CCF, it decreases from a peak at AIX~0.06 to a minimum at AIX ~ 0.12. It then increases to the second peak at AIX~0.21 and decreases to the second minimum at AIX~0.32. It gradually levels off with fluctuations beyond AIX~0.32. For CTH, there is also an AIX threshold (~0.20) that CTH increases with AIX increase below this threshold, then decreases with AIX increase after passing this threshold until it reaches AIX~0.32. Afterwards, CTH levels off with fluctuations and relatively large standard errors. CTT also has similar, albeit reversed variations to CTH due to their strong negative correlation as observed in Figure 3b.

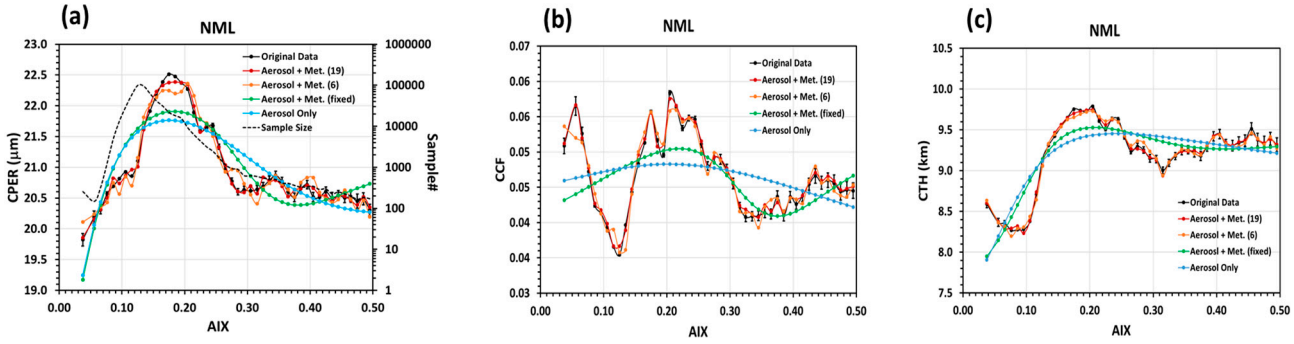

Figure 7. Statistic relationships (black curves) of (**a**) cloud particle effective radius (CPER), (**b**) cloud cover fraction (CCF), and (**c**) cloud top height (CTH) with aerosol index (AIX) derived from their long-term (1982–2019) averaged values in the northern middle latitude (NML) along with the fitting curves from the machine learning back-propagation neural network (BPNN) model with different input variables: AIX + 19 meteorological variables (red curves), AIX + 6 meteorological variables (brown curves), AIX + constant 19 meteorological variables (green curves), and solely the AIX variable (blue curves). Black dots and the vertical bars are the mean values and the standard errors of cloud variables, respectively, for individual AIX bins.

The above changing tendency of the cloud variables with AIX increase in the sensitive regime of aerosol effect is entangled with meteorological feedbacks and covariances (see Zhao and Foster [28]). Here, we used the machine learning BPNN model introduced in Section 3 for a regression analysis in order to separate the entangled effects. First, for each of the cloud variables, we use AIX and 19 meteorological variables listed in Table 1 as the input of the BPNN model to fit the black curves displayed in Figure 7 for AIX bins (red curves). Then, we reduced 19 meteorological variables to six variables (CAPE, PW, RH_{clm}, T_{850}, ω_{400}, and VSHW), which are the most important for DCC formation and development,

and repeated the BPNN fitting (brown curves). We also fixed 19 meteorological variables to their values in the first AIX bin (0.0–0.05, the cleanest scenario) of Figure 7 to perform BPNN fitting as a case with AIX increasing in a constant meteorological condition (green curves). Last, only AIX was used as the input variable to perform BPNN fitting as the most extreme case without the feedbacks and covariances of the meteorological condition (blue curves). The selection of 20 input variables for the BPNN model is based on our experience and previous studies, e.g., [3,10,13]. The selection of the six most important aerosol and meteorological variables used in one of our study cases follows Zang et al. [13], which suggests these six variables are critical for studying the aerosol indirect effect on DCCs.

Only the results for CPER, CCF, and CTH from the BPNN fitting are shown in Figure 7 as examples, since the results of IWP (or COD) are similar (or reversely similar) to CPER and the results of CTT are reversely similar to CTH due to their strong positive or negative correlations, as observed in Figure 3. We can see that for the cloud micro-physical variable CPER (Figure 7a), the cases of aerosol-only (blue curve) and aerosol with constant meteorological condition (green curve) can generally capture the basic variation (such as the hump-type shape centered at AIX~0.18) observed in the original data but miss the exact magnitude and the detailed variations. When 6 meteorological variables are added, the resulting fitting curve (brown) becomes much closer to the original curve. When 19 meteorological variables are included, the fitting curve (red) almost completely overlaps with the original curve. This suggests the primary aerosol effect (the aerosol effect without meteorological feedback) may capture basic CPER variation with AIX increase. The meteorological feedbacks to the primary aerosol effect and the covariances of meteorological condition need to be included to capture the detailed CPER variations accurately. For the macro-physical cloud variable CCF (Figure 7b), the primary aerosol effect cannot reproduce the first peak of CPER at AIX~0.06 but partially captures the second CPER peak at AIX~0.21 and the follow-on variations. This suggests that the first peak is probably solely due to the meteorological covariance, or it may be a spurious feature due to limited sample numbers at the low end of the AIX bins. Further study is warranted in a future work to find out the real cause. The second peak is due to both the primary aerosol effect and the corresponding meteorological feedback and covariance. For CTH (Figure 7c), the convective invigoration peaked at AIX~0.20 can only be partially captured by the primary aerosol effect (blue and green curves). The corresponding meteorological feedback and covariance need to be included for a more complete capture of the CTH variations (brown and red curves).

4.3. Coupled Analysis of Aerosol Effect on Cloud Variables

The above SHAP correlation analysis suggests that the aerosol effect on DCCs manifests more easily in NML than in the other two latitude belts (TRL and SML). Thus, we specifically selected two rectangular regions in NML, where aerosol loadings are influenced by the continental aerosols transported by offshore winds (see Figure 8), to perform SVD analysis for the six cloud variables versus AIX. The first region (120°E–180°E; 30°N–60°N) over the Western Pacific Ocean (WPO) is influenced by industrial pollutants and dust particles originating from China in addition to marine aerosols, and the second region (60°W–0°W; 30°N–60°N) over the North Atlantic Ocean (NAO) is influenced by industrial pollutants originating from the USA and Europe along with marine aerosols.

Figure 9 shows the time series of principal components (PCs) for the first, second, and third mode of SVD analysis in WPO for AIX versus the three cloud micro-physical variables (CPER, COD, and IWP). For CPER versus AIX, the three leading PCs (PC1, PC2, and PC3) explain ~34%, ~5%, and ~4% (total: 44%) of the variance, respectively, and they are dominated by a mixture of interannual and interdecadal fluctuations. PC1 of CPER and AIX shows a significant positive trend, PC2 shows a minor positive trend, and PC3 shows a somewhat negative trend. Pearson's correlation coefficients between PC1 of CPER and AIX is r = 0.70, r = 0.68 for PC2, and r = 0.74 for PC3. For COD versus AIX, PC1, PC2, and PC3 explain ~20%, ~7%, and ~6% (total: 33%) of the variance, respectively. They also

show a mixture of interannual and interdecadal fluctuations. PC1 of COD and AIX shows a significant positive trend while PC2 and PC3 show a minor negative and positive trend, respectively. The correlation coefficients of COD and AIX in PC1, PC2, and PC3 are r = 0.78, 0.73, and 0.84, respectively, and they are slightly higher than the values of the three leading PCs of CPER versus AIX. For IWP versus AIX, PC1, PC2, and PC3 explain ~28%, ~5%, and ~4% (total: 37%) of the variance, respectively. They are also dominated by a mixture of interannual and interdecadal oscillations. PC1 of IWP and AIX also shows a significant and congruous negative trend while PC2 and PC3 barely show a trend. The correlation coefficients of IWP and AIX in PC1, PC2, and PC3 are r = 0.90, 0.91, and 0.97, respectively, and they are higher than the values of the three leading PCs of CPER (or COD) versus AIX. Thus, the aerosol effect on DCC in WPO accounts for more variance of CPER (~44%) than IWP (~37%) and COD (~33%). However, the signature of the aerosol effect likely manifests more easily in the three PCs modes of IWP than in the corresponding modes of COD and CPER, especially in the PC3 mode, due to the highest correlation coefficients of IWP and AIX in PC1, PC2, and PC3 compared to COD and CPER. Therefore, IWP should be a more effective cloud micro-physical variable for detecting aerosol effects on DCCs in the PC modes of satellite observation.

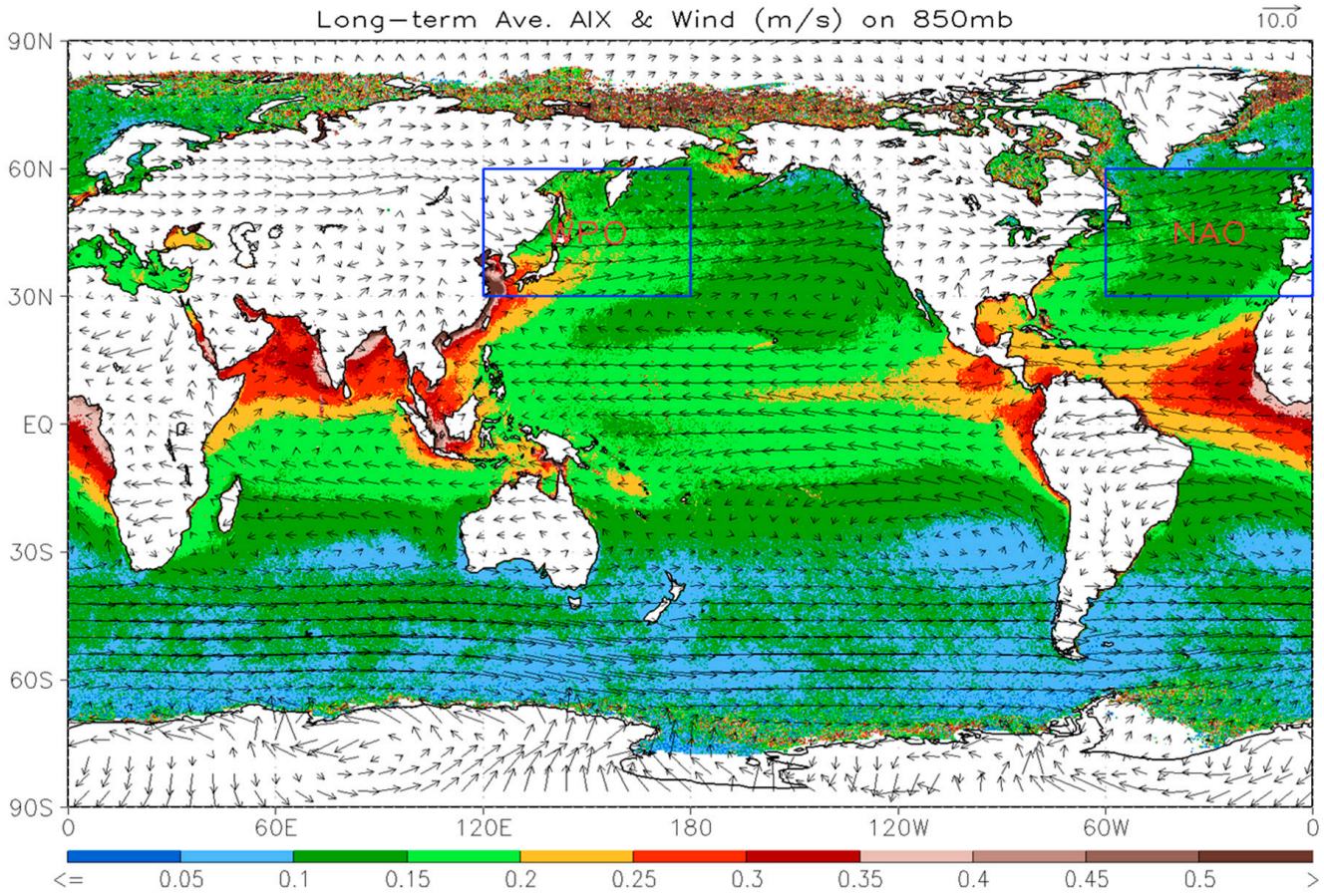

Figure 8. Long-term (1982–2019) averaged monthly mean aerosol index (AIX) over the global oceans, which is overlaid with long-term averaged horizontal wind vectors (m/s) on the 850 mb pressure level. Two blue rectangle boxes marked for WPO (Western Pacific Ocean) and NAO (North Atlantic Ocean) are the two regions selected for singular value decomposition (SVD) analysis.

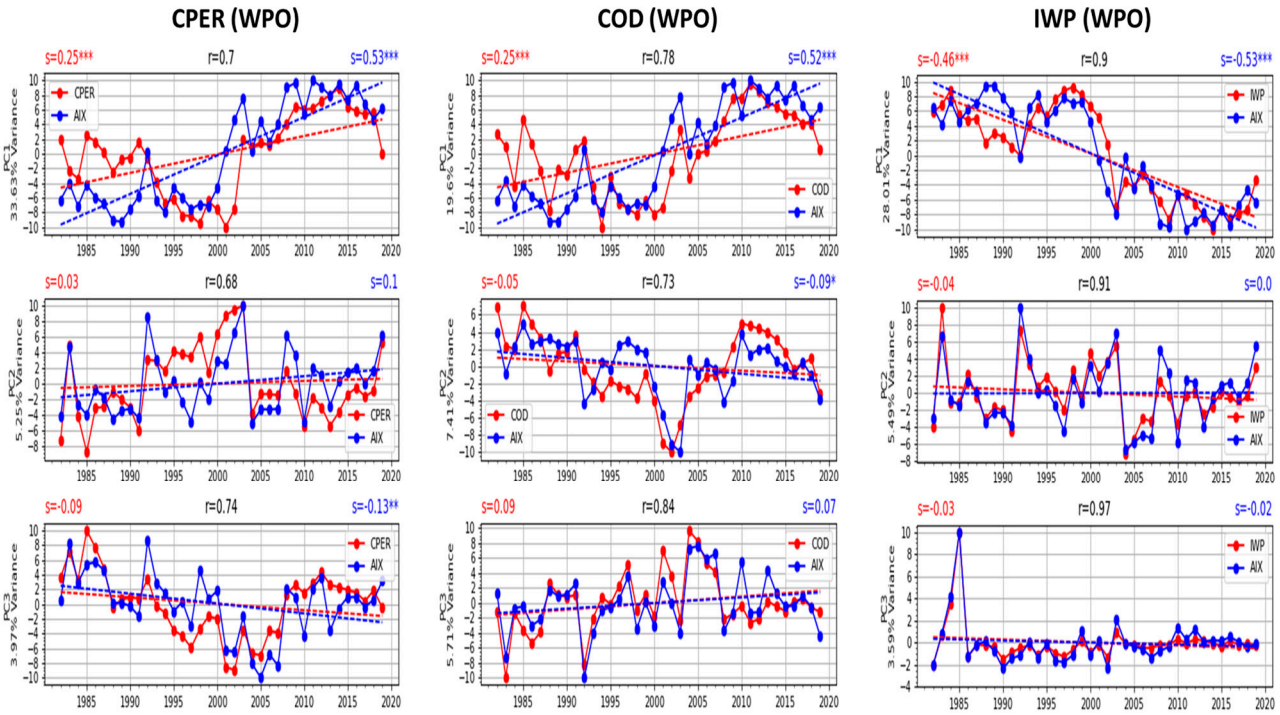

Figure 9. Time series of principal components (PCs) of the first (**top row**), second (**middle row**), and third (**bottom row**) singular value decomposition (SVD) mode in the Western Pacific Ocean (WPO) region for aerosol index (AIX) and three cloud micro-physical variables: cloud particle effective radius (CPER, **left column**), cloud optical depth (COD, **middle column**), and ice water path (IWP, **right column**). The two linear trends are given by the two dash lines, and the corresponding trend values are given by *s* with three, two, one, and no asterisks to indicate 99%, 95%, 90%, and less than 90% confidence levels, respectively. *r* values are the Pearson's correlation coefficients between the PCs of cloud variables and AIX. The corresponding trend values are given by *s* with three, two, one, and no asterisks to indicate 99%, 95%, 90%.

Figure 10 shows the time series of SVD PCs in the WPO region for AIX versus the three macro-physical variables (CCF, CTH, and CTT). For CCF versus AIX, PC1, PC2, and PC3 explain ~24%, ~7%, ~5% (total: 36%) of the variance, respectively. PC1 and PC2 show a mixture of interannual and interdecadal fluctuations, while PC3 is dominated by interannual oscillations. PC1 of CCF and AIX shows a significant and congruous negative trend, while PC2 and PC3 barely show a trend. The correlation coefficients of CCF and AIX in PC1, PC2, and PC3 are r = 0.83, 0.73, and 0.83, respectively. For CTH versus AIX, PC1, PC2, and PC3 explain ~24%, ~6%, ~4% (total: 34%) of the variance, respectively. PC1 and PC2 also show a mixture of interannual and interdecadal fluctuations while PC3 is dominated by interannual oscillations. PC1 of CTH and AIX shows a significant negative trend while PC2 and PC3 barely show a trend. The correlation coefficients of CTH and AIX in PC1, PC2, and PC3 are r = 0.84, 0.86, and 0.91, respectively. For CTT versus AIX, PC1, PC2, and PC3 explain ~22%, ~6%, ~4% (total: 32%) of the variance, respectively. PC1 and PC2 show a mixture of interannual and interdecadal fluctuations while PC3 is dominated by interannual oscillations. PC1 of CTT and AIX shows a significant positive trend, while PC2 and PC3 barely show a trend. The correlation coefficients of CTT and AIX in PC1, PC2, and PC3 are r = 0.88, 0.92, and 0.89, respectively. The relatively higher correlation coefficients of CTT (or CTH) and AIX in PC1, PC2, and PC3 than that of CCF and AIX suggest that the signature of the aerosol effect on DCC in the three PC modes may manifest somewhat clearer in the three PC modes of CTT (or CTH) than in the corresponding PC modes of CCF even though the aerosol effect on DCC in WPO accounts for more variance of CCF (~36%) than CTH (~34%) and CTT (~32%).

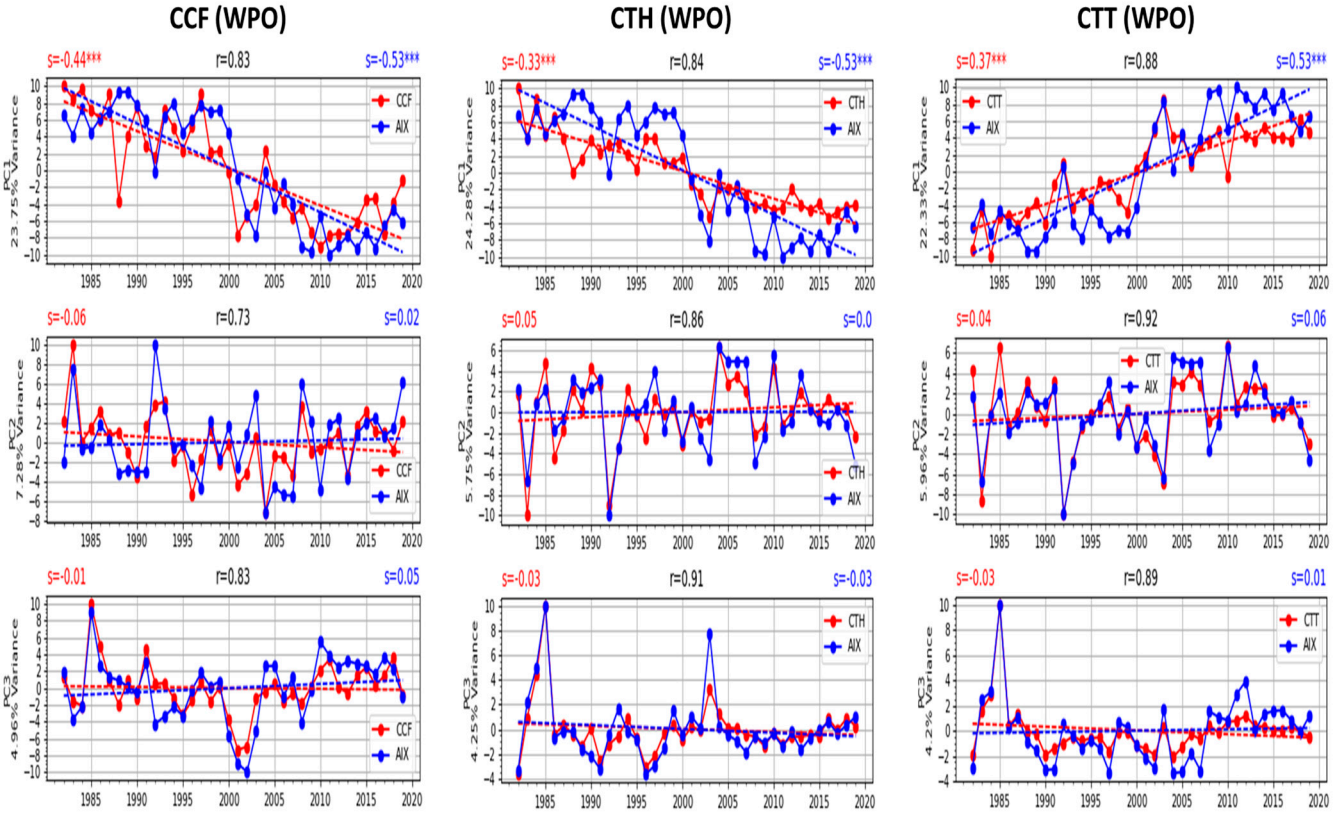

Figure 10. Similar to Figure 10 but for AIX and three cloud macro-physical variables: cloud cover fraction (CCF, **left column**), cloud top height (CTH, **middle column**), and cloud top temperature (CTT, **right column**). The corresponding trend value is given by *s* with three and no asterisks to indicate 99%.

Since the PC1 of AIX and the six cloud variables over the WPO displays evident positive or negative trends in Figures 9 and 10, it is worth having a further analysis on this changing tendency of mode 1 (or PC1 component) from SVD analysis. Table 3 summarizes the changing tendencies of mode 1 from SVD analysis for the six DCC variables versus AIX over the WPO. The sign of positive or negative trends of PC1 for AIX and the six DCC variables displayed in Figures 9 and 10 are extracted and listed in rows 1 and 4 of Table 3, respectively. The positive or negative sign of the PC1 eigenvector for AIX and the corresponding six DCC variables are given in rows 2 and 5, respectively. As examples to show the real meaning of a positive or negative sign of PC1 eigenvector for AIX versus the six DCC variables, Figure 11 shows three scenarios of PC1 eigenvector distribution over the WPO for CPER versus AIX (Figure 11a,d), CCF versus AIX (Figure 11b,e), and COD versus AIX (Figure 11c,f). We can see the eigenvectors of CPER and AIX are positive (the 1st scenario) over the WPO, the eigenvectors of CCF and AIX are mainly negative (the 2nd scenario), and the eigenvector of COD can be negative or positive (the 3rd scenario) while the corresponding eigenvector of AIX is positive.

Multiplying row 4 with row 5, we can determine the changing tendency of mode 1 variance for the DCC variables (row 6), while the corresponding mode 1 changing tendency of AIX (row 3) can be determined by multiplying row 1 with row 2. Actually, the sign of row 1 (or 4) and row 2 (or 5) does not matter; what matters is the sign of their product in row 3 (or row 6). We can see mode 1 of AIX (corresponding to the six DCC variables) show consistent and congruous positive changing trend over the WPO (row 3). Correspondingly, in row 6, CPER, IWP, CCF, and CTH show positive changing trend but the corresponding changing trend of COD and CTT can be negative or positive. Thus, the aerosol effect on CPER, IWP, CCF, and CTH of DCC are positive in PC1 component but can be negative or

positive for COD and CTT (see row 7). Aerosol DCC invigoration (CTH and CCF increase in row 6) in PC1 component is enhanced gradually due to the increase of AIX PC1 component (in row 3) from 1982 to 2019 over the WPO.

Table 3. Summary of the linear changing tendencies of PC1 component (or mode 1) from SVD analysis for the six DCC variables versus AIX over the Western Pacific Ocean (WPO), which indicate if the aerosol indirect effect (AIE) on the six DCC variables is positive (enhancement/increase) or negative (weakening/decrease) or both.

No	AIX/DCC PC1 Properties	Six DCC Variables						Comments
		CPER	COD	IWP	CCF	CTH	CTT	
1	AIX PC1 Trend	(−) P (+)	(−) P (+)	(+) N (−)	(+) N (−)	(+) N (−)	(−) P (+)	(−) P (+): (from negative to positive) Positive trend (+) N (−): (from positive to negative) Negative trend
2	AIX PC1 Eigenvector	+	+	−	−	−	+	+: positive value −: negative value
3	AIX PC1 Variance	(−) P (+)	(−) P (+)	(−) P (+)	(−) P (+)	(−) P (+)	(−) P (+)	row1 × row2
4	DCC PC1 Trend	(−) P (+)	(−) P (+)	(+) N (−)	(+) N (−)	(+) N (−)	(−) P (+)	(−) P (+): (from negative To positive) Positive trend (+) N (−): (from positive to negative) Negative trend
5	DCC PC1 Eigenvector	+	−/+	−	−	−	−/+	+: positive value −: negative value −/+: negative or positive
6	DCC PC1 Variance	(−) P (+)	(+) N (−)/(−) P (+)	(−) P (+)	(−) P (+)	(−) P (+)	(+) N (−)/(−) P (+)	row4 × row5
7	Summary	P-AIE	N-/P-AIE	P-AIE	P-AIE	P-AIE	N-/P-AIE	N/P-AIE: negative/positive AIE

Notes: (1) (−) P (+) means changing from negative value to positive value with a positive trend and (+) N (−) is reversed. (2) −/+, (+) N (−)/(−) P (+), and N-/P-AIE mean can be negative or positive but negative is somewhat dominant in the studying region. (3) Rows 1 and 4 are the PC1 trend of AIX and cloud variables extracted from the linear regression (blue and red dish lines) in Figures 9 and 10. (4) Rows 2 and 5 are determined by examining the distribution of PC1 Eigenvector for AIX and cloud variables as in Figure 11. (5) Sign and trend tendency of mode 1 variance for AIX (row 3) is determined from multiplying AIX PC1 eigenvalues (row 1) with AIX PC1 eigenvectors (row 2). Similarly, sign and trend tendency of mode 1 variance for the DCC variables (row 6) is determined from multiplying their PC1 eigenvalues (row 4) with PC1 eigenvectors (row 5).

Results in NAO are shown in Figure 12 for the three cloud micro-physical cloud variables and AIX. The three leading PC modes for CPER versus AIX explain ~34%, ~6%, and ~4% (total: 44%) of the variance, respectively. The corresponding values for the three leading PC modes of IWP versus AIX are ~24%, ~6%, and ~5% (total: 35%), respectively. Their temporal variations are similar, with interdecadal oscillations dominating in PC1 and PC2, but interannual fluctuations are evident in PC3. Their PC1 shows a significant negative trend, while their PC2 and PC3 show minor trends. Since the correlation coefficient of PC3 of CPER (or IWP) and AIX is above 0.95 and the corresponding values of PC1 and PC2 are below 0.90, the aerosol effect on DCC should be easier to detect in the third SVD mode than in the first and second modes for CPER and IWP. For COD versus AIX, the three leading PC modes explain ~22%, ~6%, and ~5% (total: 33%) of the variance, respectively. PC1 of AIX shows a much more significant positive trend than that of COD. Their PC2 shows a negative trend, and PC3 shows a positive trend. The corresponding correlation coefficients are r = 0.64, 0.89, and 0.74 for PC1, PC2, and PC3, respectively, which are generally lower than the values of IWP (or CPER) versus AIX. Thus, the aerosol effect on DCC should be

easier to manifest in the three leading PC modes of IWP and CPER than in those of COD over the NAO.

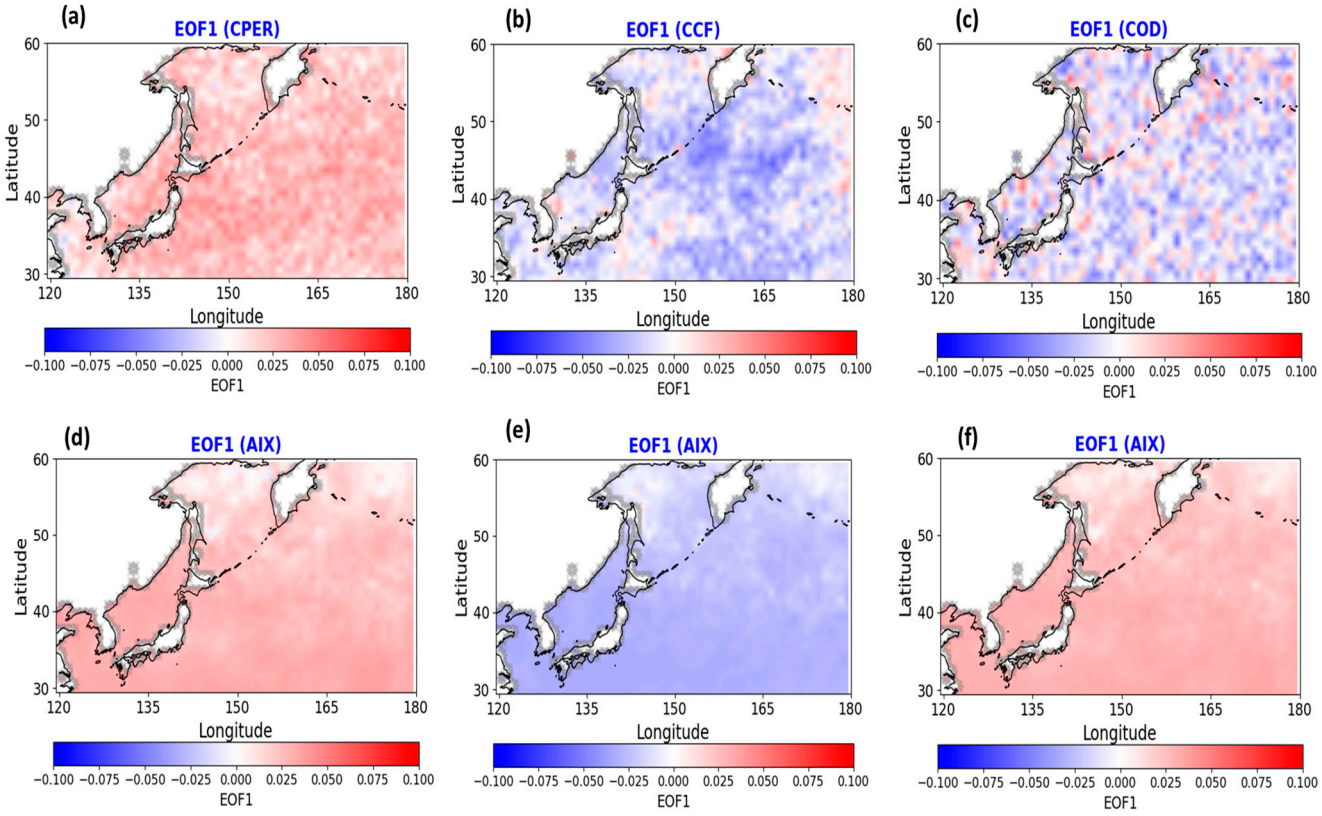

Figure 11. Examples of eigenvector distribution of empirical orthogonal function (EOF) mode 1 (or PC1 component) from the SVD analysis over the Western Pacific Ocean (WPO) for (**a**) CPER versus (**d**) AIX, (**b**) CCF versus (**e**) AIX, and (**c**) COD versus (**f**) AIX.

Figure 13 shows the results of SVD analysis in the NAO for the three cloud macro-physical variables versus AIX. The three leading PC modes for CCF versus AIX explain ~19%, ~8%, and ~6% (total: 33%) of the variance, respectively. Interdecadal oscillations are dominated in PC1 and PC2 while a mixture of interannual and interdecadal fluctuations is the major feature in PC3. PC1 of CCF and AIX shows a much more significant positive trend than PC2 and PC3. The correlation coefficients are r = 0.83, 0.77, and 0.88 for PC1, PC2, and PC3, respectively. For CTH versus AIX, PC1, PC2, and PC3 explain ~24%, ~7%, and ~6% (total: 37%) of the variance, respectively, and the corresponding values for CTT are ~18%, ~7%, and ~6% (total: 31%). Temporal oscillations of PC1 for CTH and CTT are similar but reversed in sign for PC2 and PC3. PC1 of CTH and AIX show a significant negative trend but their PC2 and PC3 show a minor positive trend. PC1 of CTT and AIX shows a much more significant negative trend than PC2 and PC3. The signature of the aerosol effect on DCC should manifest somewhat clearer in the three leading PCs of CTT compared to that of CTH (or CCF) since the corresponding correlation coefficients of PC1, PC2, and PC3 for CTT versus AIX are generally above 0.9 but below 0.9 for CTH and CCF. Thus, CTT should be a more effective cloud macro-physical variable for detecting aerosol invigoration on DCCs in the PC modes of satellite observation over the NAO.

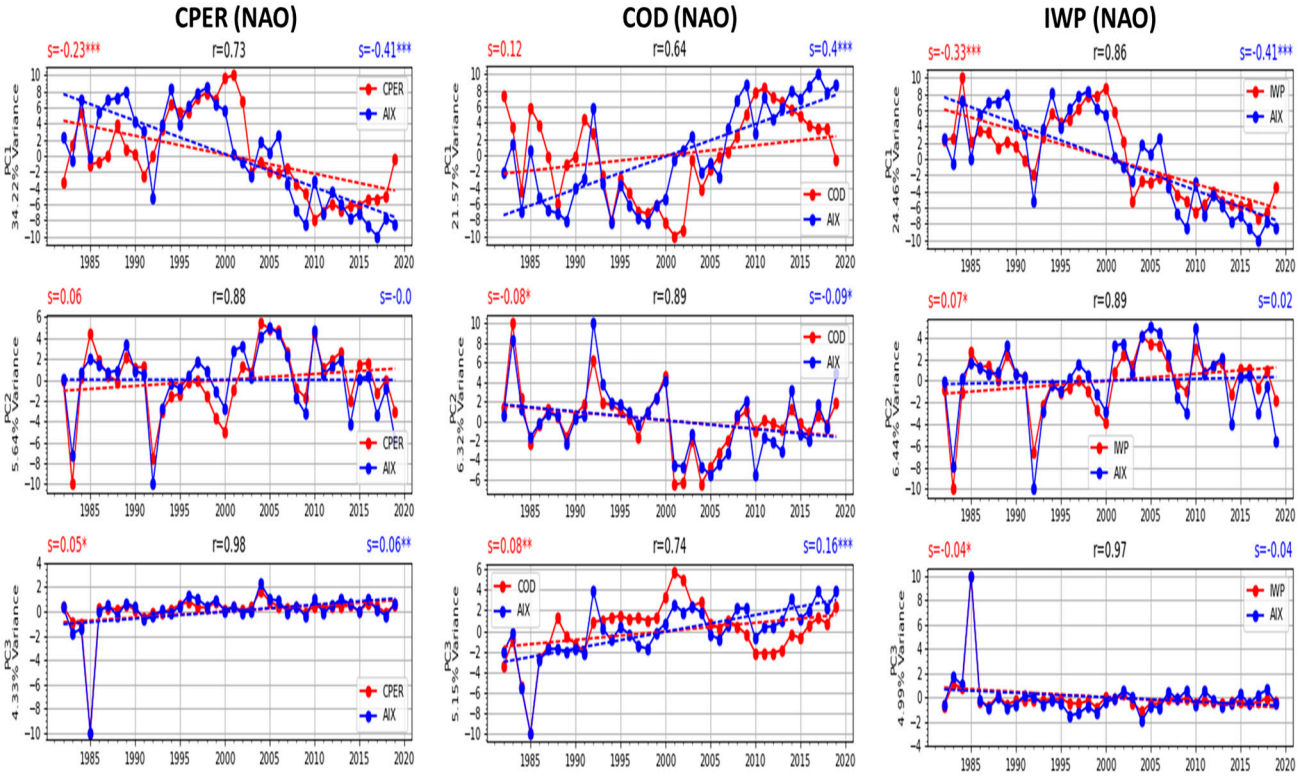

Figure 12. Time series of principal components (PCs) of the first (**top row**), second (**middle row**), and third (**bottom row**) singular value decomposition (SVD) mode in the North Atlantic Ocean (NAO) region for aerosol index (AIX) and three cloud micro-physical variables: cloud particle effective radius (CPER, **left column**), cloud optical depth (COD, **middle column**), and ice water path (IWP, **right column**). The two linear trends are given by the two dash lines and the corresponding trend values are given by *s* with three, two, one, and none asterisks to indicate 99%, 95%, 90% and less than 90% confidence level, respectively. *r* values are the Pearson's correlation coefficients between PCs of cloud variables and AIX.

We furtherly studied the changing tendency of mode 1 for AIX and the six DCC variables over the NAO, since they also display evident positive or negative trends in Figures 12 and 13 as in Figures 9 and 10 for WPO. The results are summarized in Table 4 for the NAO as in Table 3 for the WPO. We can see mode 1 of AIX (versus the six DCC variables) shows a consistent positive changing tendency (row 3) from 1982 to 2019 over the NAO. Correspondingly in row 6, CPER, IWP, and CTH show positive changing trends, COD and CTT show negative trends, and the changing trend of CCF can be positive or negative. As a result (row 7), the AIE of CPER, IWP and CTH of DCC are positive in PC1 component and negative for COD and CTT and can be positive or negative for CCF. DCC invigoration (CTH increase while CTT decrease) is enhanced in PC1 component due to the increase of AIX PC1 component from 1982 to 2019 over the NAO. Thus, the PC1 component is an effective mode for detecting the aerosol effect on DCCs over both WPO and NAO.

By comparing the summary row 7 of Table 3 for the WPO and 4 for the NAO, we can see that the major difference in the aerosol effect on DCCs in the PC1 component of these two regions are in COD, CCF, and CTT. The effect on COD in the WPO can be negative or positive but dominated by a negative effect in the NAO. The effect on CCF is positive in the WPO but can be positive or negative in the NAO. The effect on CTT can be negative or positive in the WPO but dominated by a negative effect in the NAO. These differences may indicate that COD and CTT in the WPO are more sensitive to the changes in aerosols, while CCF is more sensitive to the change of aerosols in the NAO. Further in-depth studies are needed in the future to find out the causes of different sensitivity of COD, CCF, and CTT to the aerosol changes in two DCC regions.

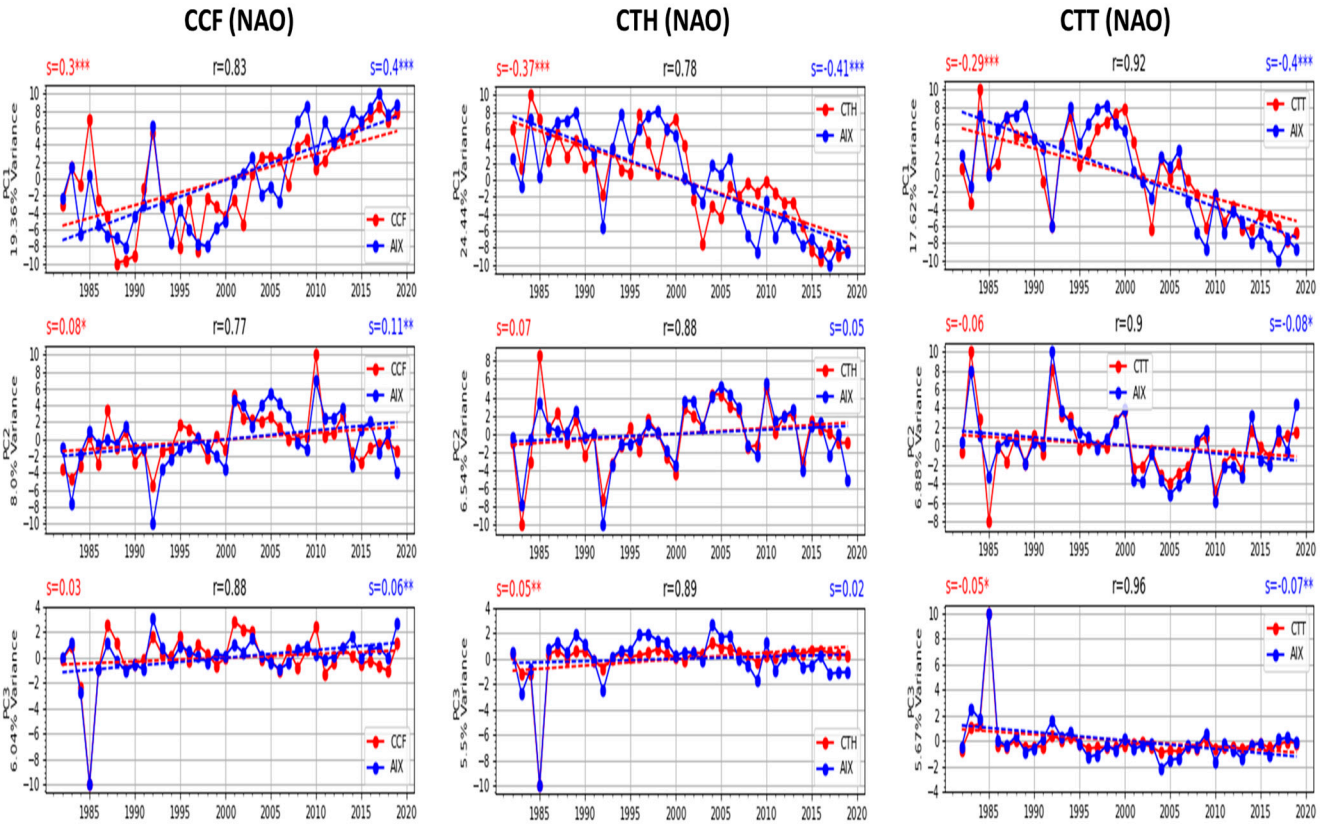

Figure 13. Similar to Figure 13 but for aerosol index (AIX) and three cloud macro-physical variables: cloud cover fraction (CCF, **left column**), cloud top height (CTH, **middle column**), and cloud top temperature (CTT, **right column**). The corresponding trend values are given by *s* with three, two, one, and no asterisks to indicate 99%, 95%, 90%.

Table 4. Similar to Table 3 but for the North Atlantic Ocean (NAO).

No	AIX/DCC PC1 Properties	6 DCC Variables						Comments
		CPER	COD	IWP	CCF	CTH	CTT	
1	AIX PC1 Trend	(+) N (−)	(−) P (+)	(+) N (−)	(−) P (+)	(+) N (−)	(+) N (−)	(−) P (+): (from negative to positive) Positive trend (+) N (−): (from positive to negative) Negative trend
2	AIX PC1 Eigenvector	−	+	−	+	−	−	+: positive value −: negative value
3	AIX Mode1 Variance	(−) P (+)	(−) P (+)	(−) P (+)	(−) P (+)	(−) P (+)	(−) P (+)	row1 × row2
4	DCC PC1 Trend	(+) N (−)	(−) P (+)	(+) N (−)	(−) P (+)	(+) N (−)	(+) N (−)	(−) P (+): (from negative to positive) Positive trend (+) N (−): (from positive to negative) Negative trend
5	DCC PC1 Eigenvector	−	−	−	+/−	−	+	+: positive value −: negative value +/−: positive or negative

Table 4. *Cont.*

No	AIX/DCC PC1 Properties	6 DCC Variables						Comments
		CPER	COD	IWP	CCF	CTH	CTT	
6	DCC Mode1 Variance	(−) P (+)	(+) N (−)	(−) P (+)	(−) P (+)/(+) N (−)	(−) P (+)	(+) N (−)	Row4 × row5
7	Summary	P-AIE	N-AIE	P-AIE	P-/N-AIE	P-AIE	N-AIE	P/N-AIE: positive/negative AIE

Notes: (1) (−) P (+) means changing from negative value to positive value with a positive trend and (+) N (−) is reversed. (2) +/−, (−) P (+)/(+) N (−), and P-/N-AIE mean can be positive or negative but positive is somewhat dominant in the studying region. (3) Rows 1 and 4 are the PC1 trend of AIX and cloud variables extracted from the linear regression as in Table 3. (4) Rows 2 and 5 are determined by examining the distribution of PC1 Eigenvector for AIX and cloud variables as in Table 3. (5) Sign and trend tendency of Mode 1 variance for AIX (row 3) is determined from multiplying AIX PC1 Eigenvalues (row 1) and AIX PC1 eigenvectors (row 2). Similarly, sign and trend tendency of mode 1 variance for the DCC variables (row 6) is determined from multiplying their PC1 eigenvalues (row 4) and PC1 eigenvectors (row 5).

5. Summary and Conclusions

We performed studies of aerosol effects on DCCs over the global oceans from a climatological perspective by using machine learning techniques and long-term (1982–2019) satellite CDRs of aerosols and clouds and CFSR reanalysis of meteorological fields. We combined the Python XGBoost nonlinear regression model and Python SHAP analysis package to identify the potential AIE of DCCs in three ocean latitude belts (NML, TRL, and SML) where DCCs appear more frequently in the climatology. It was found that AIE of DCCs may be detected only in NML from long-term averaged satellite aerosol and cloud observations. Specifically, the cloud micro-physical variable CPER is more susceptible to the aerosol effect than the other two cloud micro-physical variables (COD and IWP) in NML. At the same time, AIE signature is more easily obscured by meteorological covariances for the three macro-physical cloud variables (CCF, CTH, and CTT) in NML. Wind and moisture fields in the upper troposphere are most important for the prediction of cloud coverage. Temperature, wind, and moisture fields in the UT are critical for the prediction of CTH. Temperature field in the UT, convective energy, and moisture supplies near the surface are critical for the CTT prediction.

Statistical relationships of cloud variables with AIX were computed to find the sensitive regime of AIX for the AIE of DCCs, which is from AIX~0.05 to ~0.32. With the help of the BPNN machine learning model, we tried to separate the aerosol effect from the covariances of the meteorological feedbacks in the AIX sensitive regime. It was found that the primary aerosol effect (the aerosol effect without meteorological feedback) only partially contributes to the observed variations in the three cloud micro-physical variables when AIX increases. The meteorological feedback and covariance need to be included in order to accurately capture the observed changing tendency of the cloud micro-physical variables. There are two peaks in the AIX sensitive regime for the cloud macro-physical variable CCF. The first peak at AIX~0.06 is probably due to meteorological covariance or a spurious feature due to limited sample size, whereas the second peak at AIX~0.21 is due to entangled influence of the primary aerosol effect and the corresponding meteorological feedback and covariance. For the cloud macro-physical variables CTH and CTT, the convective invigoration peaked at AIX~0.20 can only be partially captured by the primary aerosol effect. The corresponding meteorological feedback and covariance need to be included for a more complete capture of the CTH and CTT variations. The meteorological variables CAPE, PW, RH_{clm}, T_{850}, ω_{400}, and VSHW along with aerosol variable AIX are the minimum needs to accurately capture the observed AIE and associated meteorological feedback and covariance of DCCs.

From the SVD analysis for the six cloud variables versus AIX in the WPO and NAO regions, we found that the aerosol effects in the three leading PC modes may explain about one third of the variance of the six cloud variables. Significant positive or negative trends are only observed in the PC1 of cloud and aerosol variables. In the Western Pacific Ocean region, aerosol effect on DCCs should manifest more easily in the three PC modes of IWP

than in the corresponding modes of CPER and COD, especially in PC3 mode, even though the aerosol effect on DCCs accounts for more variance of CPER (~44%) than IWP (~37%) and COD (~33%). IWP is a more effective cloud micro-physical variable for detecting aerosol effects on DCCs in the PC modes of satellite observation. Aerosol effect on DCC may manifest somewhat more easily in the three PC modes of CTT (or CTH) than in the corresponding PC modes of CCF. For the Northern Atlantic Ocean region, the aerosol effect on DCC should be manifest more easily in the three leading PC modes of IWP and CPER than that of COD. Moreover, the aerosol effect on DCC manifests more easily in the three leading PC modes of CTT than those of CTH (or CCF), and CTT is a more effective cloud macro-physical variable for detecting aerosol invigoration on DCCs in the PC modes of satellite observation. Aerosol DCC invigoration over the Western Pacific Ocean (CTH and CCF increase) and over the Northern Atlantic Ocean (CTH increase but CTT decrease) is enhanced in the PC1 component due to the increase of AIX PC1 component from 1982 to 2019, and the PC1 component is an effective mode for detecting the aerosol effect on DCCs.

Understanding the climate effect of aerosols on deep convective clouds is challenging due to the nonlinear relationship between aerosols and cloud variables, as well as the influence of meteorological factors. Existing climatological studies on the aerosol effects on DCCs primarily rely on climate model simulations that compare a control scenario (excluding aerosol effects) with an aerosol scenario (including aerosol effects). In contrast, observational studies typically focus on regional, local, and instantaneous interactions between aerosols and clouds. Assessing how these regional/local and instantaneous interactions impact the long-term mean state (or climate) of clouds from an observational perspective remains a significant challenge. In this paper, we employ three different approaches—SHAP analysis, nonlinear fitting with a BPNN model, and SVD analysis—to identify the signatures of aerosol effects on DCCs in long-term averaged satellite observations. This multifaceted approach is necessary because no single method is sufficient to capture the complex picture of aerosol impacts on DCCs. By combining the insights from these three methods, we aim to gain a better understanding of aerosol effects on DCCs from a climatological observational perspective. We chose to focus our study on the analysis and results rather than the three analytical tools, thus omitting many of the optimization procedures performed for the BPNN model, SHAP analysis, and SVD analysis. We acknowledge that there is still a long way to go to achieve a comprehensive and integrated understanding of aerosol impacts on DCCs based on long-term satellite observations. However, our current analysis represents an initial step towards this goal. Our results are valuable for the evaluation and improvement of aerosol-cloud interactions in the long-term climate simulations of global climate models.

Author Contributions: Conceptualization, X.Z.; methodology, X.Z. and J.F.; software, X.Z. and J.F.; validation, X.Z., J.F., M.J.F., and A.K.H.; formal analysis, X.Z., J.F., M.J.F., and A.K.H.; writing—original draft preparation, X.Z.; writing—review and editing, J.F., M.J.F., and A.K.H.; visualization, X.Z. and J.F.; supervision, X.Z.; All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by NOAA/NESDIS/NCEI Climate Data Record (CDR) program and NCEI base fund. J.F. is supported by NCEI through a contract with the Cooperative Institute for Satellite Earth System Studies (CISESS)—Maryland at the University of Maryland in College Park.

Data Availability Statement: The data sources used in this study have been provided in Section 2 with the data weblinks. For the data produced in this study, please email to the lead author.

Acknowledgments: Proofreading the paper by NCEI English editor Lukas Noguchi is greatly appreciated. We would like to acknowledge the comments and suggestions provided by three anonymous reviewers, which are highly constructive for the improvement of our paper. We also acknowledge the science editor for organizing this special issue and the MDPI section managing editor for assisting the reviewing process. The views, opinions, and findings contained in this paper are those of the author (s) and should not be construed as an official National Oceanic and Atmospheric Administration or U.S. Government position, policy, or decision.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

ACI(s)	aerosol cloud interaction(s)
AI	artificial intelligence
AIE	aerosol indirect effect
AIX	aerosol index
AOT	aerosol optical thickness
AVHRR	Advanced Very High-Resolution Radiometer
BPNN	back-propagation neural network
CAPE	convective available potential energy
CCF	cloud cover fraction
CDR(s)	climate data record(s)
CFSR	climate forecast system reanalysis
COD	cloud optical depth
CPER	cloud particle effective radius
CTH	cloud top height
CTT	cloud top temperature
DCC(s)	deep convective cloud(s)
EOF	Empirical Orthogonal Function
EUMETSAT	European Organization for the Exploitation of Meteorological Satellites
GAC	global area coverage
HIRS	High-resolution Infra-Red Sounder
IWP	ice water path
NASA	National Aeronautics and Space Administration
MCS	mesoscale convective systems
MetOp	Meteorological Operational Satellites
MAI	covarying meteorology-aerosol invigoration
ML	machine learning
MODIS	Moderate-resolution Imaging Spectroradiometer
MSE	mean squared error
NAO	Northern Atlantic Ocean
NCEI	National Centers for Environmental Information
NCEP	National Centers for Environmental Prediction
NESDIS	National Environmental Satellite, Data, and Information Service
NH	northern hemisphere
NML	northern middle latitude
NOAA	National Oceanic and Atmospheric Administration
PAI	primary aerosol convective invigoration
PATMOS-x	Pathfinder Atmospheres-Extended
PC(s)	principal component(s)
POES	Polar Operational Environmental Satellites
PW	precipitable water
RH	relative humidity
SHAP	SHapley Additive exPlanation
SML	southern middle latitude
STAR	Center for Satellite Applications and Research
SVD	singular value decomposition
TRL	tropical latitude
UT	upper troposphere
VSHW	vertical shear of horizontal wind
WPO	Western Pacific Ocean
XGBoost	extreme gradient boosting (XGBoost)

References

1. Arakawa, A. The cumulus parameterization problem: Past, present, and future. *J. Clim.* **2004**, *17*, 2493–2525. [CrossRef]
2. Futyan, J.M.; Del Genio, A.D. Deep convective system evolution over Africa and the tropical Atlantic. *J. Clim.* **2007**, *20*, 5041–5060. [CrossRef]
3. Khain, A.; Rosenfeld, D.; Pokrovsky, A. Aerosol impact on the dynamics and microphysics of deep convective clouds. *Q. J. Roy. Meteor. Soc.* **2005**, *131*, 2639–2663. [CrossRef]

4. Lohmann, U.; Feichter, J. Global indirect aerosol effects: A review. *Atmos. Chem. Phys.* **2005**, *5*, 715–737. [CrossRef]
5. Stevens, B.; Feingold, G. Untangling aerosol effects on clouds and precipitation in a buffered system. *Nature* **2009**, *461*, 607–613. [CrossRef] [PubMed]
6. Tao, W.K.; Chen, J.P.; Li, Z.Q.; Wang, C.; Zhang, C.D. Impact of Aerosols on Convective Clouds and Precipitation. *Rev. Geophys.* **2012**, *50*, 1–62. [CrossRef]
7. Fan, J.W.; Wang, Y.; Rosenfeld, D.; Liu, X.H. Review of Aerosol-Cloud Interactions: Mechanisms, Significance, and Challenges. *J. Atmos. Sci.* **2016**, *73*, 4221–4252. [CrossRef]
8. Seinfeld, J.H.; Bretherton, C.; Carslaw, K.S.; Coe, H.; DeMott, P.J.; Dunlea, E.J.; Feingold, G.; Ghan, S.; Guenther, A.B.; Kahn, R.; et al. Improving our fundamental understanding of the role of aerosol-cloud interactions in the climate system. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 5781–5790. [CrossRef] [PubMed]
9. Li, Z.Q.; Rosenfeld, D.; Fan, J.W. Aerosols and their impact on radiation, clouds, precipitation, and severe weather events. *Oxf. Res. Encycl. Environ. Sci.* **2017**, 1–36. [CrossRef]
10. Fan, J.W.; Yuan, T.L.; Comstock, J.M.; Ghan, S.; Khain, A.; Leung, L.R.; Li, Z.Q.; Martins, V.J.; Ovchinnikov, M. Dominant role by vertical wind shear in regulating aerosol effects on deep convective clouds. *J. Geophys. Res.-Atmos.* **2009**, *114*, 1–9. [CrossRef]
11. Grabowski, W.W. Untangling Microphysical Impacts on Deep Convection Applying a Novel Modeling Methodology. *J. Atmos. Sci.* **2015**, *72*, 2446–2464. [CrossRef]
12. Grabowski, W.W.; Morrison, H. Untangling Microphysical Impacts on Deep Convection Applying a Novel Modeling Methodology. Part II: Double-Moment Microphysics. *J. Atmos. Sci.* **2016**, *73*, 3749–3770. [CrossRef]
13. Zang, L.; Rosenfeld, D.; Pan, Z.X.; Mao, F.Y.; Zhu, Y.N.; Lu, X.; Gong, W. Observing Aerosol Primary Convective Invigoration and Its Meteorological Feedback. *Geophys. Res. Lett.* **2023**, *50*, 1–10. [CrossRef]
14. Zhao, X.-P.; Dubovik, O.; Smirnov, A.; Holben, B.N.; Sapper, J.; Pietras, C.; Voss, K.J.; Frouin, R. Regional evaluation of an advanced very high resolution radiometer (AVHRR) two-channel aerosol retrieval algorithm. *J. Geophys. Res.* **2004**, *109*, D02204. [CrossRef]
15. Zhao, T.X.P.; Chan, P.K.; Heidinger, A.K. A global survey of the effect of cloud contamination on the aerosol optical thickness and its long-term trend derived from operational AVHRR satellite observations. *J. Geophys. Res.-Atmos.* **2013**, *118*, 2849–2857. [CrossRef]
16. Nakajima, T.; Higurashi, A.; Kawamoto, K.; Penner, J.E. A possible correlation between satellite-derived cloud and aerosol microphysical parameters. *Geophys. Res. Lett.* **2001**, *28*, 1171–1174. [CrossRef]
17. Liu, J.J.; Li, Z.Q. Estimation of cloud condensation nuclei concentration from aerosol optical quantities: Influential factors and uncertainties. *Atmos. Chem. Phys.* **2014**, *14*, 471–483. [CrossRef]
18. Stier, P. Limitations of passive remote sensing to constrain global cloud condensation nuclei. *Atmos. Chem. Phys.* **2016**, *16*, 6595–6607. [CrossRef]
19. Heidinger, A.K.; Foster, M.J.; Walther, A.; Zhao, X.P. The Pathfinder Atmospheres-Extended Avhrr Climate Dataset. *Bull. Am. Meteorol. Soc.* **2014**, *95*, 909–922. [CrossRef]
20. Foster, M.J.; Phillips, C.; Heidinger, A.K.; Borbas, E.E.; Li, Y.; Menzel, W.P.; Walther, A.; Weisz, E. PATMOS-x Version 6.0: 40 Years of Merged AVHRR and HIRS Global Cloud Data. *J. Clim.* **2023**, *36*, 1143–1160. [CrossRef]
21. Heidinger, A.K.; Straka, W.C.; Molling, C.C.; Sullivan, J.T.; Wu, X.Q. Deriving an inter-sensor consistent calibration for the AVHRR solar reflectance data record. *Int. J. Remote Sens.* **2010**, *31*, 6493–6517. [CrossRef]
22. Heidinger, A.K.; Evan, A.T.; Foster, M.J. A Naive Bayesian Cloud-detection Scheme Derived from CALIPSO and Applied withing PATMOS-x. *J. Appl. Meteoro. Climat.* **2012**, *51*, 1129–1144. [CrossRef]
23. Walther, A.; Heidinger, A.K. Implementation of the Daytime Cloud Optical and Microphysical Properties Algorithm (DCOMP) in PATMOS-x. *J. Appl. Meteorol. Clim.* **2012**, *51*, 1371–1390. [CrossRef]
24. Heidinger, A.K.; Cao, C.; Sullivan, J.T. Using Moderate Resolution Imaging Spectrometer (MODIS) to calibrate advanced very high resolution radiometer reflectance channels. *J. Geophys. Res.* **2002**, *107*, 4702. [CrossRef]
25. Cao, C.; Weinreb, M.; Xu, H. Predicting Simultaneous Nadir Overpasses among Polar-Orbiting Meteorological Satellites for the Intersatellite Calibration of Radiometers. *J. Atmos. Ocean. Technol.* **2004**, *21*, 537–542. [CrossRef]
26. Cao, C.Y.; Xiong, X.X.; Wu, A.H.; Wu, X.Q. Assessing the consistency of AVHRR and MODIS L1B reflectance for generating fundamental climate data records. *J. Geophys. Res.-Atmos.* **2008**, *113*, 1–10. [CrossRef]
27. Pavolonis, M.J.; Heidinger, A.K.; Uttal, T. Daytime global cloud typing from AVHRR and VIIRS: Algorithm description, validation, and comparisons. *J. Appl. Meteorol.* **2005**, *44*, 804–826. [CrossRef]
28. Zhao, X.; Foster, M.J. Analyzing Sensitive Aerosol Regimes and Active Geolocations of Aerosol Effects on Deep Convective Clouds over the Global Oceans by Using Long-Term Operational Satellite Observations. *Climate* **2022**, *10*, 167. [CrossRef]
29. Saha, S.; Moorthi, S.; Pan, H.L.; Wu, X.R.; Wang, J.D.; Nadiga, S.; Tripp, P.; Kistler, R.; Woollen, J.; Behringer, D.; et al. The Ncep Climate Forecast System Reanalysis. *Bull. Am. Meteorol. Soc.* **2010**, *91*, 1015–1057. [CrossRef]
30. Lundberg, S.M.; Erion, G.; Chen, H.; DeGrave, A.; Prutkin, J.M.; Nair, B.; Katz, R.; Himmelfarb, J.; Bansal, N.; Lee, S.I. From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* **2020**, *2*, 56–67. [CrossRef]
31. Hou, L.L.; Dai, Q.L.; Song, C.B.; Liu, B.W.; Guo, F.Z.; Dai, T.J.; Li, L.X.; Liu, B.S.; Bi, X.H.; Zhang, Y.F.; et al. Revealing Drivers of Haze Pollution by Explainable Machine Learning. *Environ. Sci. Tech. Let.* **2022**, *9*, 112–119. [CrossRef]

32. Song, C.B.; Becagli, S.; Beddows, D.C.S.; Brean, J.; Browse, J.; Dai, Q.L.; Dall'Osto, M.; Ferracci, V.; Harrison, R.M.; Harris, N.; et al. Understanding Sources and Drivers of Size-Resolved Aerosol in the High Arctic Islands of Svalbard Using a Receptor Model Coupled with Machine Learning. *Environ. Sci. Technol.* **2022**, *56*, A–J. [CrossRef] [PubMed]
33. Choudhary, I. A Step-By-Step Guide to Understand and Learn Shap (Shapley Additive Explaination) and How to Interpret ML Models Using the Shap Library. Available online: https://medium.com/@shahooda637/all-you-need-to-know-about-shap-for-explainable-ai-8ad35a05e6ec (accessed on 22 April 2023).
34. Lundberg, S.M. SHAP Python Package. Available online: https://github.com/slundberg/shap (accessed on 1 May 2023).
35. Molnar, C. Interpretable Machine Learning. Available online: https://christophm.github.io/interpretable-ml-book/shap.html (accessed on 14 April 2023).
36. Atkinson, P.M.; Tatnall, A.R.L. Neural networks in remote sensing—Introduction. *Int. J. Remote Sens.* **1997**, *18*, 699–709. [CrossRef]
37. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, *5*, 1097–1105. [CrossRef]
38. Zang, L.; Mao, F.Y.; Guo, J.P.; Wang, W.; Pan, Z.X.; Shen, H.F.; Zhu, B.; Wang, Z.M. Estimation of spatiotemporal PM, distributions in China by combining PM observations with satellite aerosol optical depth. *Sci. Total Environ.* **2019**, *658*, 1256–1264. [CrossRef]
39. Bretherton, C.S.; Smith, C.; Wallace, J.M. An Intercomparison of Methods for Finding Coupled Patterns in Climate Data. *J. Clim.* **1992**, *5*, 541–560. [CrossRef]
40. Wallace, J.M.; Smith, C.; Bretherton, C.S. Singular Value Decomposition of Wintertime Sea-Surface Temperature and 500-Mb Height Anomalies. *J. Clim.* **1992**, *5*, 561–576. [CrossRef]
41. Newman, M.; Sardeshmukh, P.D. A Caveat Concerning Singular-Value Decomposition. *J. Clim.* **1995**, *8*, 352–360. [CrossRef]