Supplemental Information for

Pan-Arctic Methanesulfonic Acid Aerosol: Source regions, atmospheric drivers, and future projections

Jakob Boyd Pernov[1*], Eliza Harris[2], Michele Volpi[2], Tamara Baumgartner[1], Benjamin Hohermuth[3], Stephan Henne[4], William H. Aeberhard[2], Silvia Becagli[5,6], Patricia K. Quinn[7], Rita Traversi[5,6], Lucia M. Upchurch[7,8], Julia Schmale[1*]

[1]Extreme Environments Research Laboratory, École Polytechnique Fédérale de Lausanne, Sion, Switzerland
[2]Swiss Data Science Center, ETH Zurich and École Polytechnique Fédérale de Lausanne, Switzerland
[3]Schroders Capital ILS, Zurich, Switzerland
[4]Empa, Swiss Federal Laboratories for Materials Science and Technology, Dübendorf, Switzerland
[5]Department of Chemistry, University of Florence, Florence, Italy
[6]Institute of Polar Sciences, National Research Council, Venice, Italy
[7]Pacific Marine Environmental Laboratory, National Oceanic and Atmospheric Administration, Seattle, WA, USA
[8]Cooperative Institute for Climate, Ocean, and Ecosystem Studies, University of Washington, Seattle, WA, USA

**Correspondence to*: Jakob Boyd Pernov (jakob.pernov@epfl.ch) and Julia Schmale (julia.schmale@epfl.ch)

Jakob Pernov (https://orcid.org/0000-0003-1906-2589)

Julia Schmale (https://orcid.org/0000-0002-1048-7962)

Supplementary Notes

Supplementary Note 1

MSA concentrations during May

The highest in situ Pan-Arctic MSA concentrations were observed during May (Fig. 1), with major source regions being the Greenland/Barents Seas (Fig. 2b and c). The phytoplankton blooms in the Arctic peak during spring[1], causing a peak in the amount of biogenically derived sulfur observed at coastal Arctic stations[2–6]. It is likely that atmospheric processes, in addition to marine biological activity, are affecting the MSA peak during May[7]. Positive SHAP values for BLH, wind speed, mean sea level pressure, cloud water content, SW, and LW radiation were observed in the Greenland/Barents Seas during May (Supplementary Fig. 9). High wind speeds can induce vertical mixing thereby transporting nutrients to the surface and enhancing phytoplankton activity[8], but also increase the oceanic mixed layer depth[9] which decreases the abundance of phytoplankton and incident light exposure in the surface waters. Inverse correlations between the North Atlantic Oscillation (NAO) and MSA have been detected at Thule, indicating strong westerly winds are deepening the oceanic mixed layer and thus decreasing DMS emission, however, no correlation was found at Gruvebadet, highlighting the complex relationship between wind speed and DMS emission[7]. A high BLH dilutes the atmospheric levels of DMS thus favoring ocean-atmospheric fluxes but also dilutes oxidants necessary for MSA production. These complex processes will likely have competing effects on oceanic DMS emission depending on the prevailing conditions[7]. Here our results suggest that high wind speeds and BLH are conducive towards DMS emission. High pressure over the Canadian Archipelago and low pressure over the Atlantic would promote the transport of air masses from southerly latitudes (Supplementary Fig. 1d). High amounts of cloud liquid water content and LW radiation are indicative of aqueous-phase processes aiding in the oxidation of DMS to MSA and SW radiation is necessary for the formation of oxidants, and especially for halogen production[10]. This demonstrates how the interplay between ocean biological activity (strong DMS source strength) and meteorological processes (efficient transport from southerly latitudes) affect MSA concentrations in the Arctic during peak MSA concentrations in May.

Supplementary Note 2

Model uncertainty and assumptions

Sources of uncertainty in the ML modeling include uncertainty due to in situ MSA measurements, ERA5 data, and FLEXPART simulations. MSA measurement uncertainties at the different stations are all listed below 13 %[11–14]. ERA5 has been shown to be one of the best reanalysis products in the Arctic, with correlation coefficients exceeding 0.83 against observations of temperature, RH, wind speed, and specific humidty[15]. Uncertainties in FLEXPART simulations arise from uncertainty in meteorological data (ERA5) and numerical errors from the discretization of stochastic differential equations[16], although FLEXPART has been shown to simulate long-range transport rather well[17,18]. Overall, we expect uncertainties arising from input data to be minimal. Sources of uncertainty in model projections include the accuracy of the model results, CIs of the trends, and no sources/sinks being included in the model. This approach makes several assumptions including (1) the relationships between the MSA and the atmospheric variables will not change in the future, (2) only SS grid cells will change in the future, (3) the linear trends will continue at the same rate, (4) transport patterns will not substantially change, and (5) the distribution of the extrapolated explanatory variables does not materially differ from the training set's distribution. The first assumption is likely valid since the relationships learned by our model are based on the input data and the underlying principles are physically interpretable and in line with the current understanding of MSA formation[19–22]. While the magnitude of the relationships displayed in Fig. 5 and Supplementary Fig. 7 might change slightly with new input data the overall shape of the relationship is expected to remain constant. The second assumption may not hold since changes will likely occur in the future that are not reflected in the observed trends in ERA5. To evaluate the effect of the confidence level (CL) for SS trend detection on the projected total relative change in MSA, we tested different CLs (75^{th} – 95^{th} % CL). Supplementary Figure 13 shows that while the magnitude of the total relative changes is affected by the CL for trend detection, the overall pattern is similar regardless of the CL. This indicates that the absolute value of the projected changes will be affected but our interpretations will not change. The third assumption may not hold since the slope of the trends reflects linear changes observed during the past four decades and non-linear changes in this slope cannot be quantified nor projected using the applied methods. Given the current rate of climate change in the Arctic and the possibility of the Arctic environment crossing tipping points[23,24], time-dependent changes in the slopes are likely although we are currently unable to quantify or test this assumption. The fourth assumption may also not hold given recent research[25,26] that shows transport patterns in the Arctic are changing albeit over shorter timescales (< 2 decades) than those analyzed in the current study (5 decades). The similarity between the monthly means of ERA5 and the MWT values (Supplementary Figs. 1 and 2) suggests that changes in transport patterns will have an overall minimal effect on the results of this study. The fifth assumption was tested by calculating the distribution of the extrapolated explanatory variables for each decadal interval (after imputing physically unrealistic values, see Table S4 for details). The distribution of the extrapolated explanatory variables for each decadal interval is displayed in Supplementary Fig. 14. This analysis shows that the extrapolated data distribution for any decadal interval is not greatly exceeding the original data distribution (year 0) and therefore the model is not being used to predict outside of the training set's distribution and that the fifth assumption is valid.

The use of the lower/upper CIs allows us to gauge the sensitivity of the total relative changes in Pan-Arctic MSA to changes in the slope of the trend in the ERA5 data and weigh the confidence to be placed in the projected changes for each month-decadal interval combination. The sensitivity of total relative changes is largest for the 50-year interval in each month especially in April. During August, the range of total relative changes for each decadal interval spans positive and negative values,

indicating that depending on the magnitude of the slope used to make future projections, decreases and increases in MSA are possible, although the majority of the CI range shows positive values. Similar observations are made for April and May for the 50-year projection. The sub-optimal values of the evaluation metrics for April and September (Fig. 3) and the large CIs (Fig. 7b) for the projected changes in MSA suggest that these months carry added uncertainty. Conversely, the excellent values of the evaluation metrics and small CIs during June and July give us confidence in the direction and magnitude of the projected changes in MSA during these months. Based on the limitations and assumptions of our model and analysis, the projected changes of MSA during the early spring and late summer months should be interpreted with caution, however, we can be confident in the projected changes of MSA during the peak summer months. Overall, this modeling approach makes several assumptions and has some limitations (excluding oceanic variables and sinks), but it produces robust results that show Pan-Arctic MSA concentrations will change in the future.

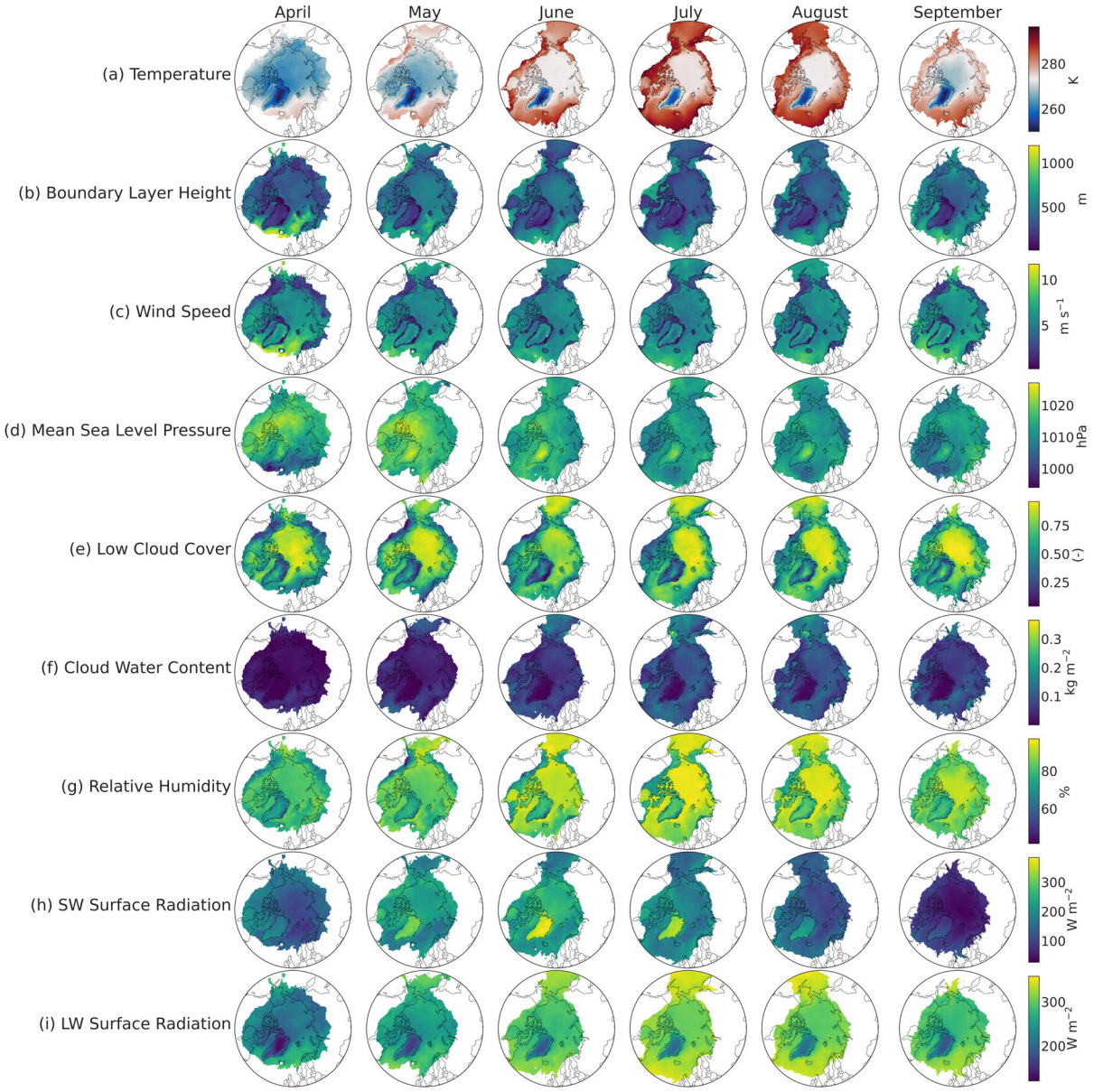

Figure S1. Maps of the meteorological weighted trajectory (MWT) for each feature (rows) for each month (columns).

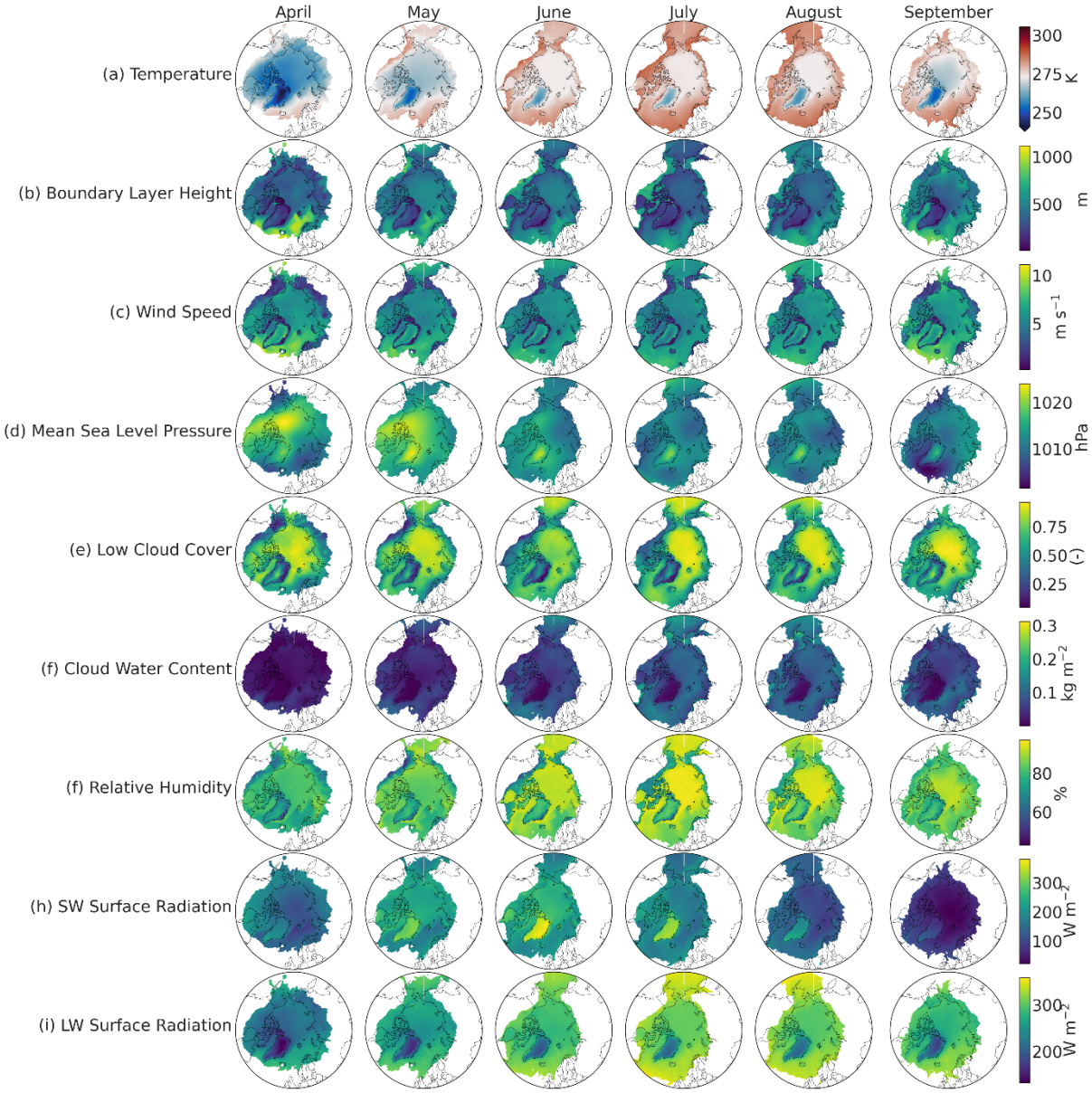

Figure S2. Maps of monthly mean ERA5 data for each feature (rows) for each month (columns). Only grid cells that are included in the extent of the CWT are shown for comparison to Figure S1.

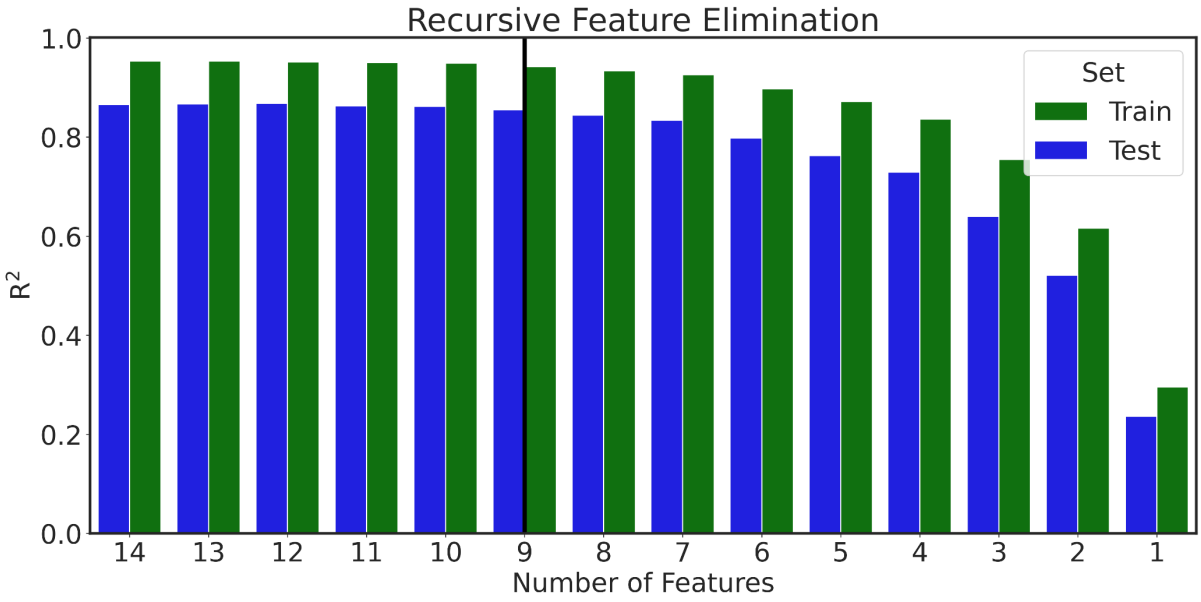

Figure S3. Input variable selection results.

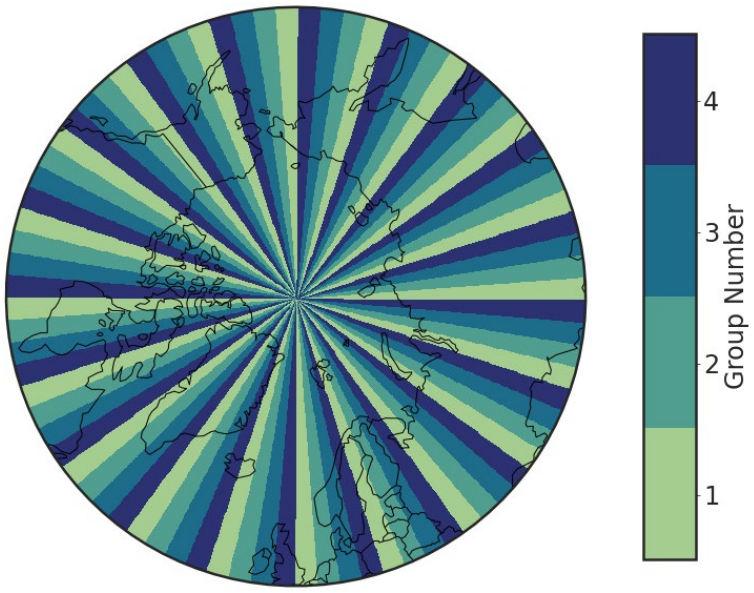

Figure S4. Map of the spatially stratified cross-validation groups.

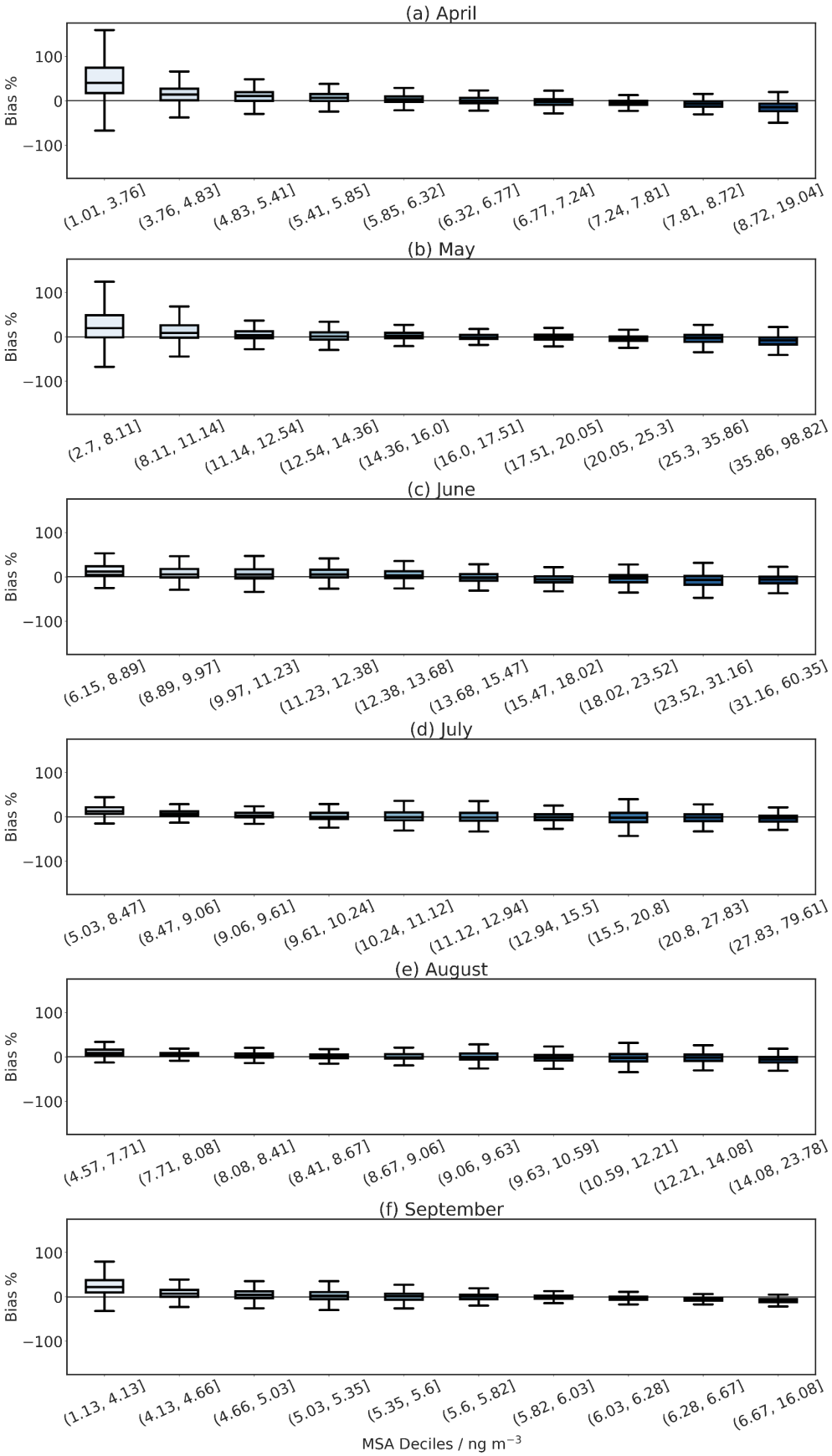

Figure S5. Evaluation of the percent bias for each month across deciles of CWT MSA concentrations.

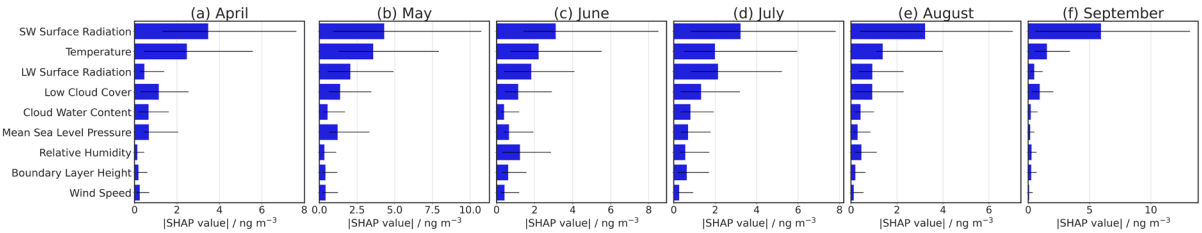

Figure S6. Monthly analysis of the median (bars) and interquartile range (lines) of the absolute SHAP values. The feature order is retained from Fig. 3.

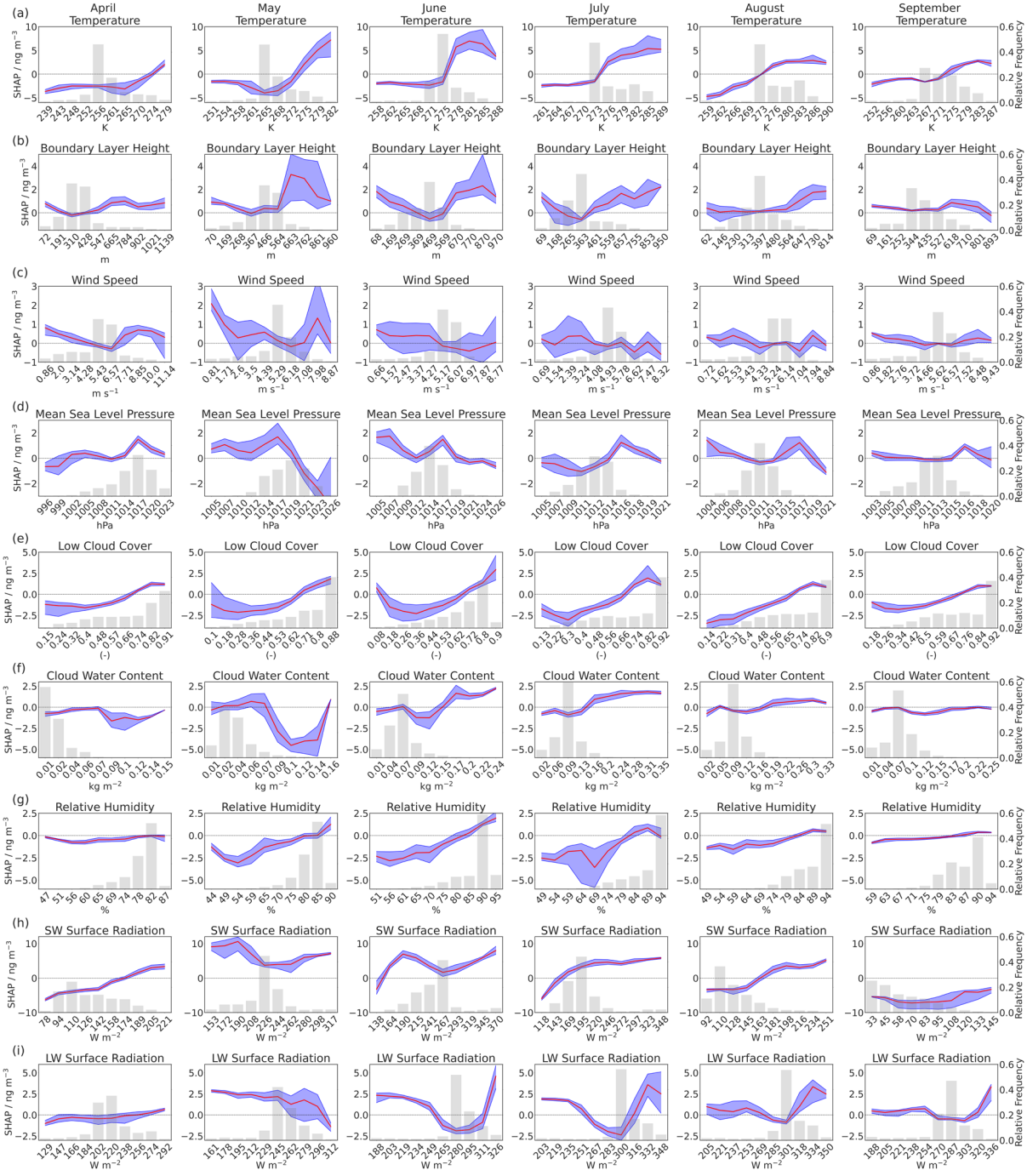

Figure S7. Relationships between SHAP values and MWT values for each feature (rows) and each month (columns). The MWT values were binned into 10 equally spaced bins and the median (red line) and interquartile range (blue shading) were calculated for each bin. The light gray bars represent the relative frequency of the input MWT values for each bin (right y-axis).

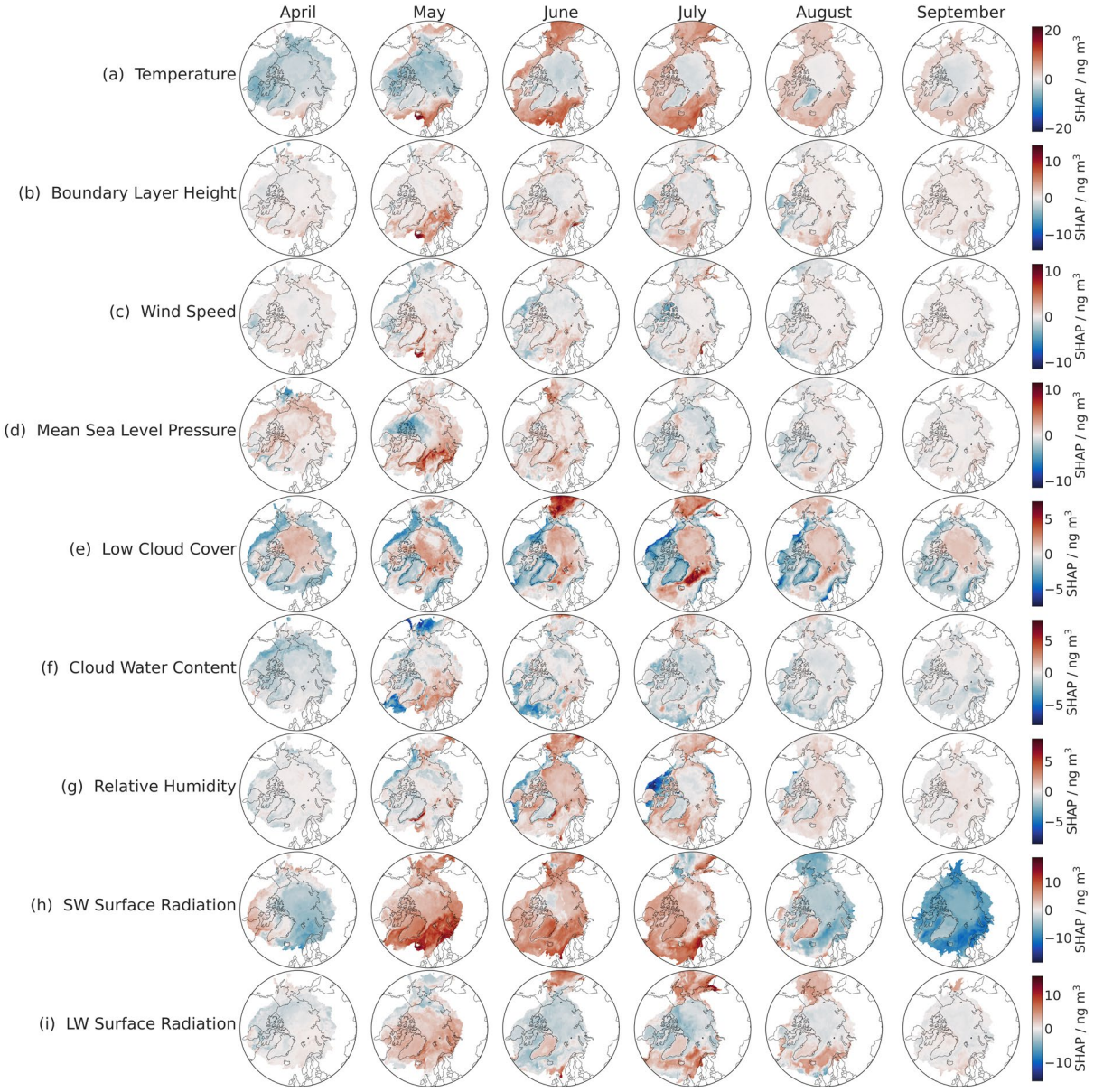

Figure S8. Maps of the SHAP values for each feature (rows) and each month (columns).

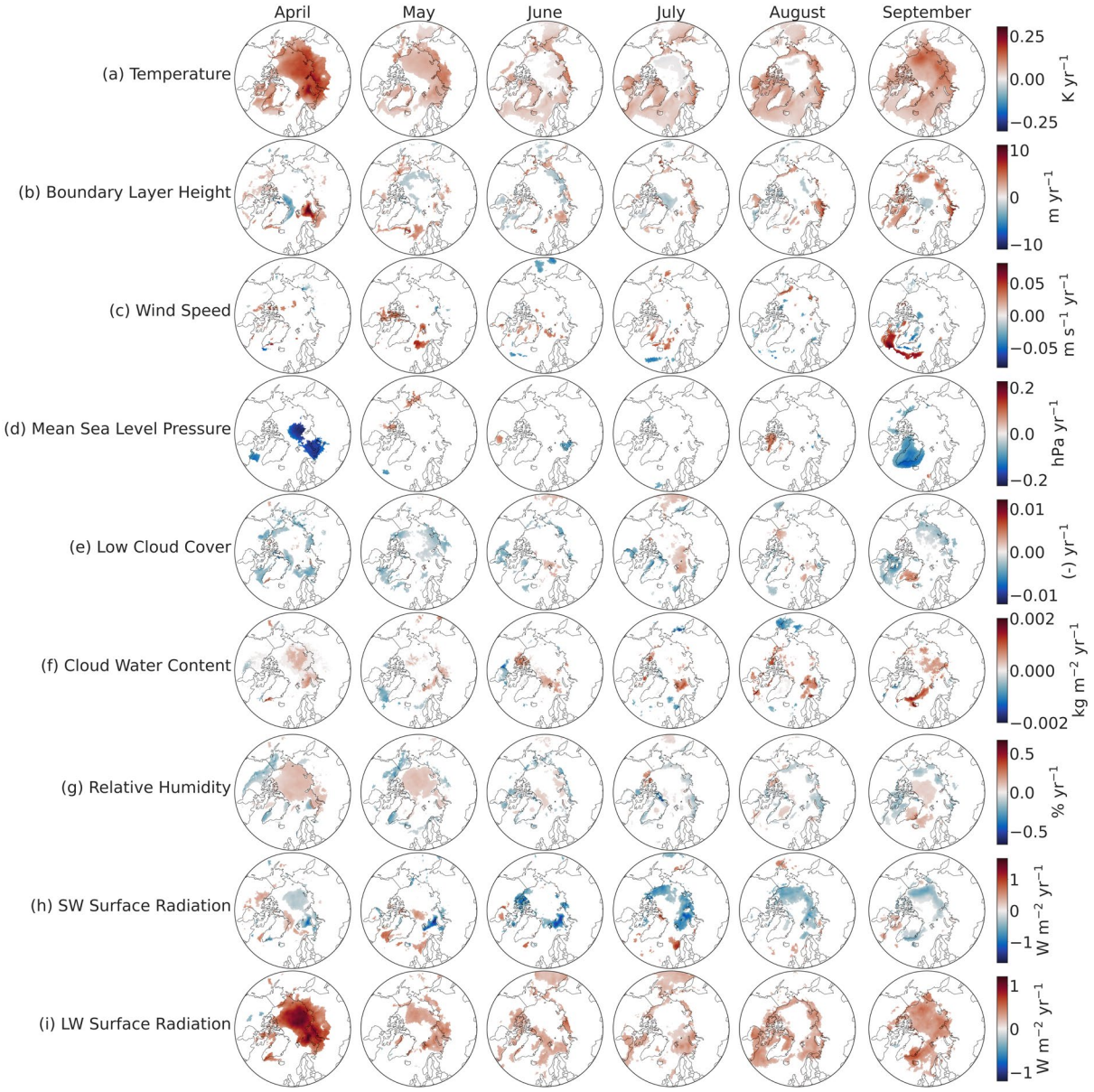

Figure S9. Maps of the trends in the monthly mean ERA5 data (1979-2021). Trends are only displayed for SS ($p < 0.05$) grid cells and grid cells that are included in the extent of the CWT.

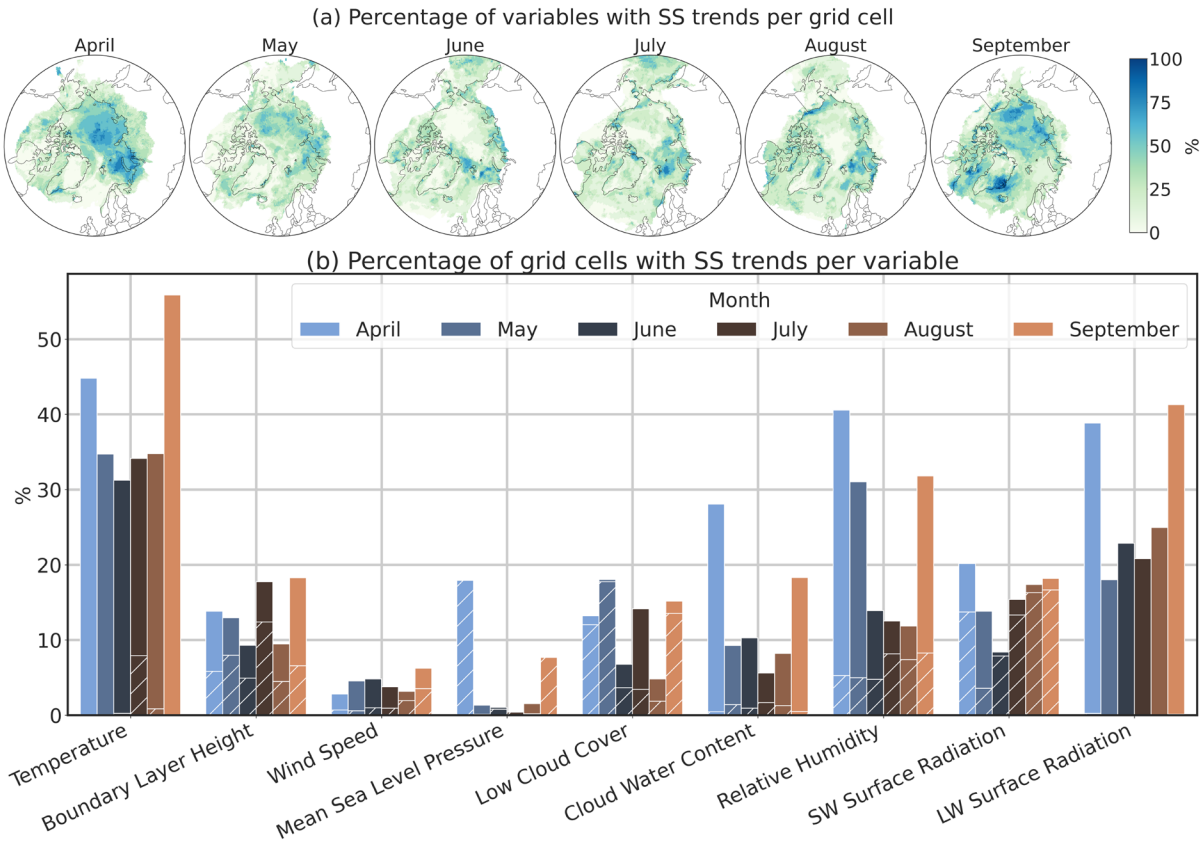

Figure S10. Overview of the ERA5 trends. (a) Maps of the percentage of variables with SS trends in each grid cell. Only grid cells included in the CWT analysis are included. (b) Percentage of grid cells with SS trends for each variable. The total number of grid cells with SS trends is expressed as a percentage of the number of grid cells included in the CWT analysis. The hatched bars represent the percentage of grid cells with SS trends that are negative.

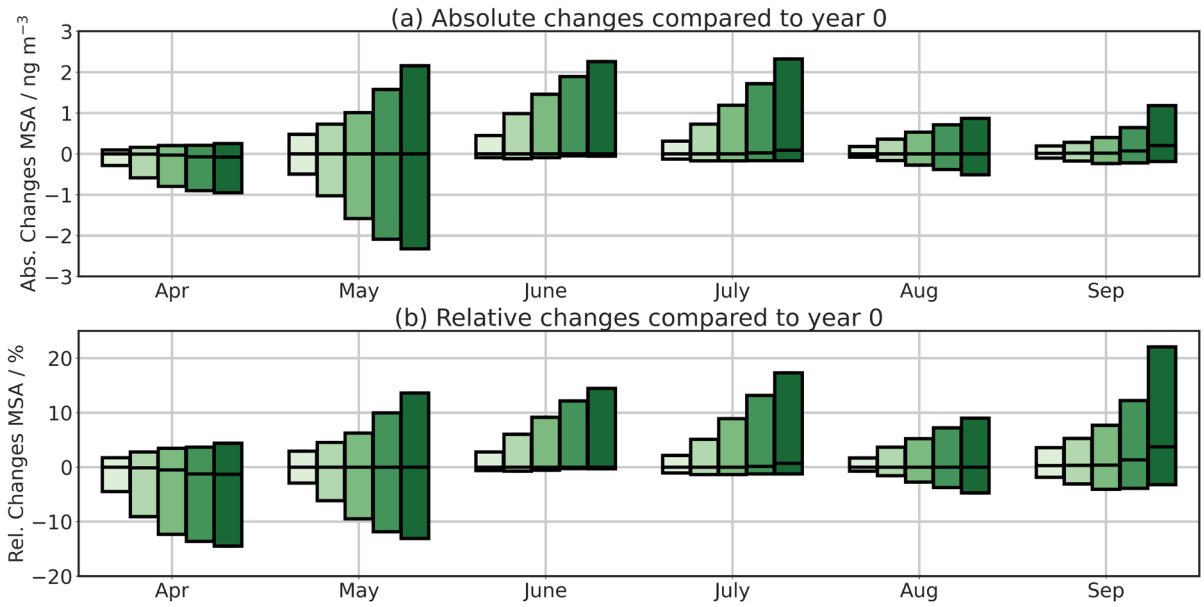

Figure S11. Projected changes in the seasonal cycle of Pan-Arctic MSA. Changes in the distribution of projected Pan-Arctic MSA values for each decadal interval for (**a**) absolute changes and (**b**) relative changes. The middle line represents the median while the upper and lower limits of the box represent the 25th/75th percentiles.

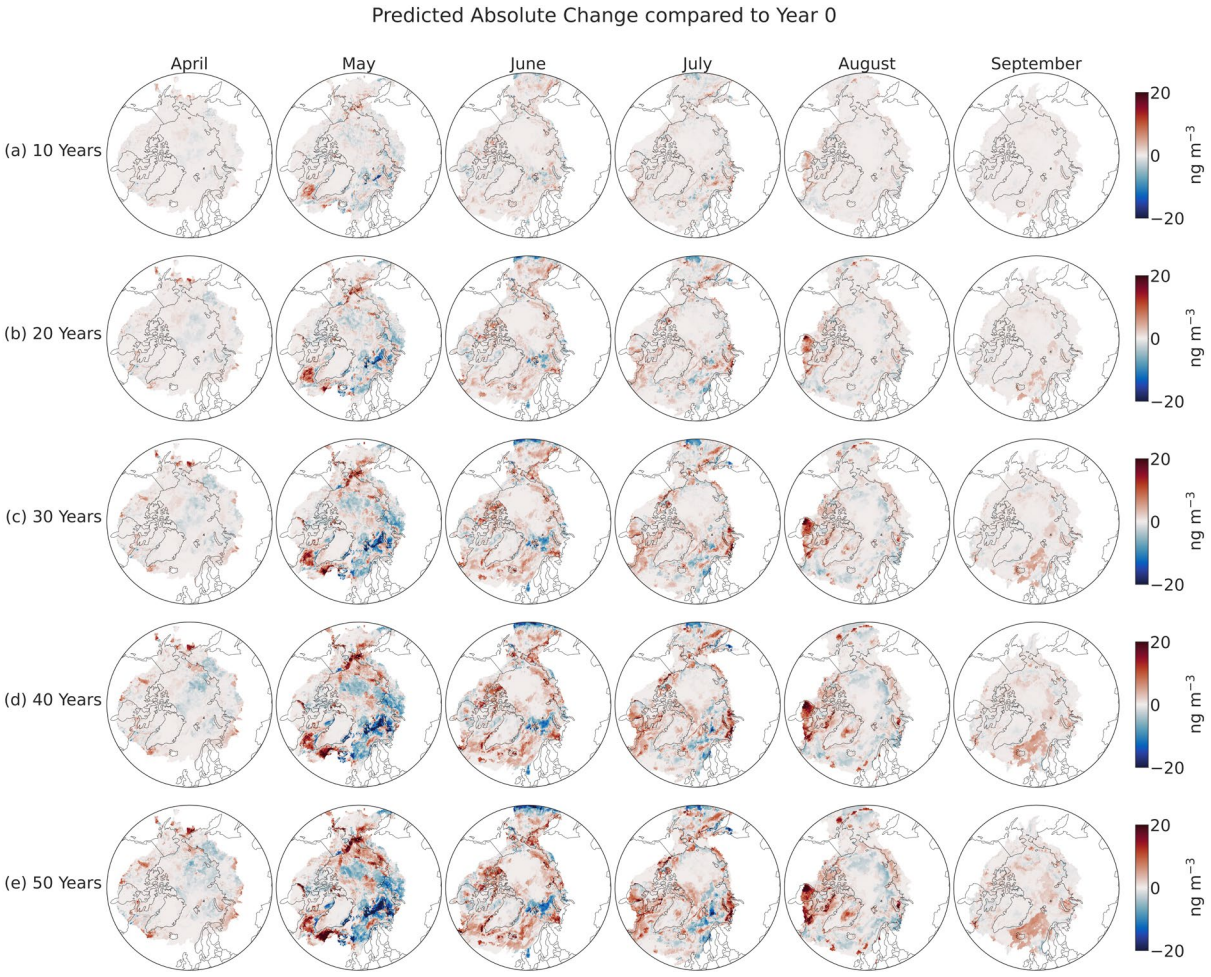

Figure S12. Maps of the projected absolute changes in MSA for each decadal interval compared to year 0 from our machine learning model.

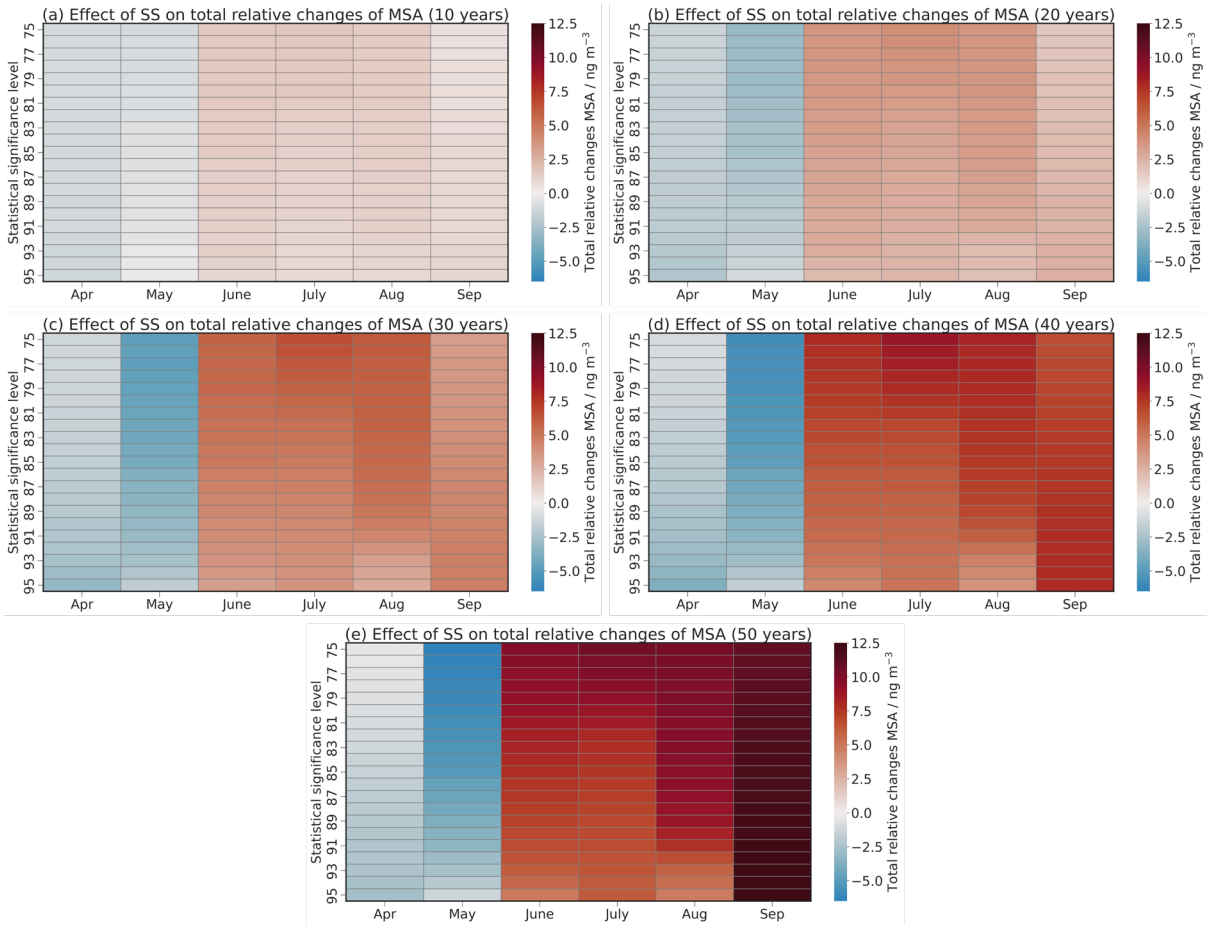

Figure S13. The effect of changing the SS level on the total relative changes of MSA for each decadal interval.

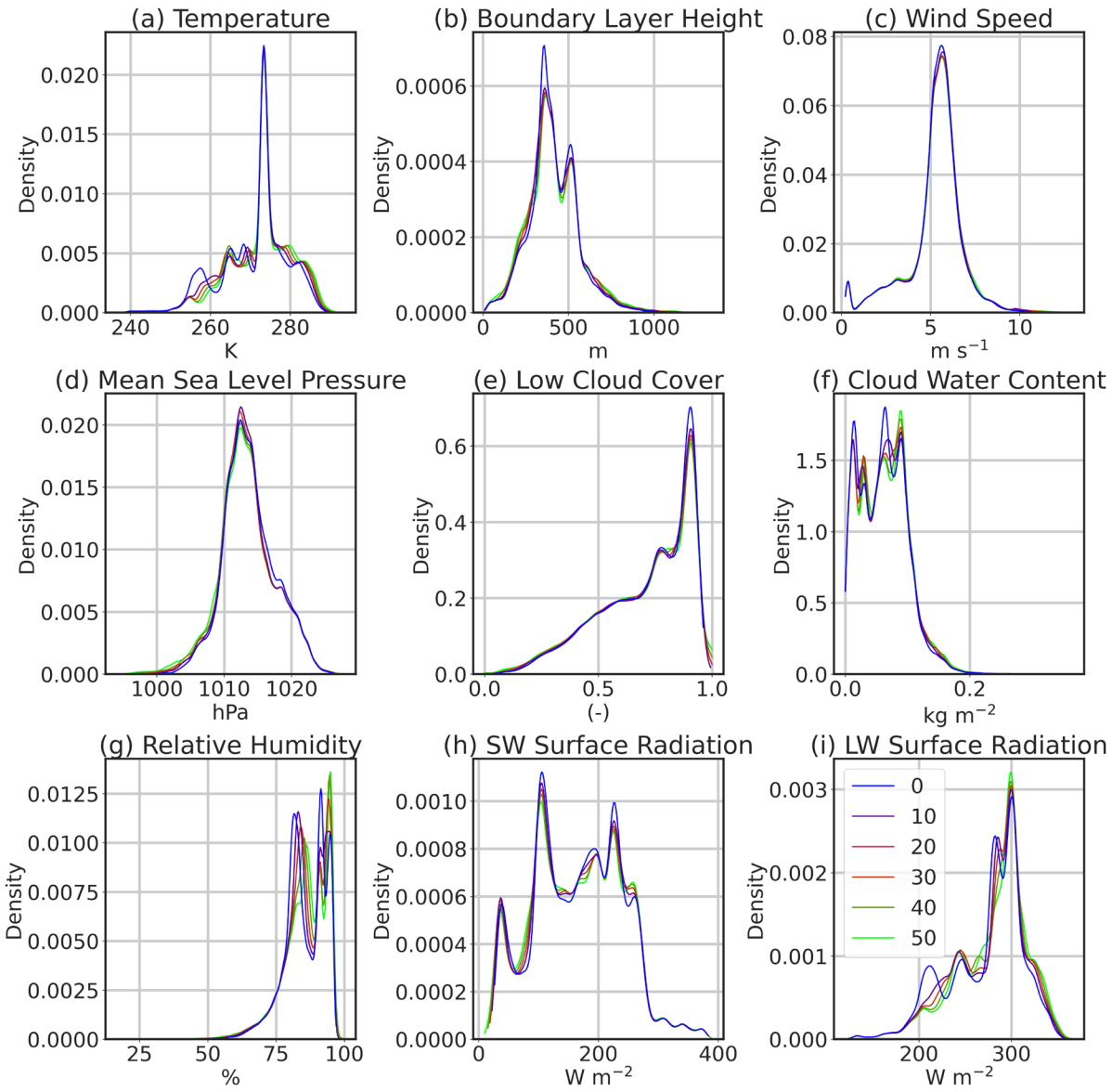

Figure S14. Kernel density estimates of the feature distribution for different decadal intervals (as represented by the different colored lines).

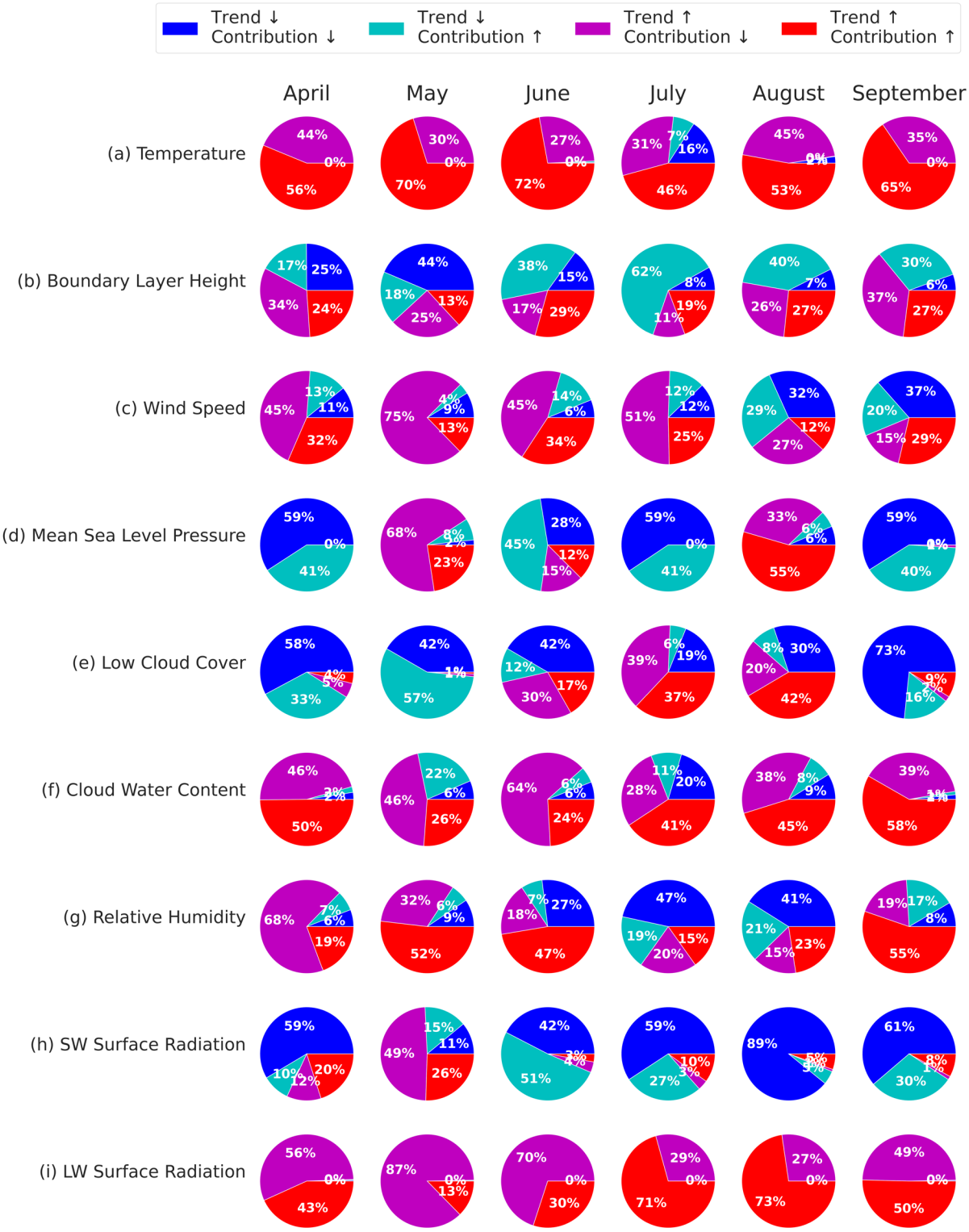

Figure S15. Results of the classification of each grid cell with an SS trend for each feature (rows) and each month (columns) for the year 50 projection.

Table S1. Features included in initial feature selection via recursive feature elimination.

Temperature	**Boundary Layer Height**	Convective Rain Rate	**Low Cloud Cover**
Mean Sea Level Pressure	Total Cloud Cover	Total Column Cloud Ice Water	**Total Column Cloud Liquid Water**
Total Column Supercooled Liquid Water	Total Column Snow Water	**Wind Speed**	**Relative Humidity**
Large Scale Rain Rate	**SW Radiation**	**LW Radiation**	

Bold indicates the feature was selected and included in the model.

Table S2. Overview of the hyperparameter optimization.

Hyperparameter	Short name	Range	Optimal Value
Number of estimators	n_estimators	100-250	250
Maximum tree depth	max_depth	3-7	7
L1 regularization	reg_alpha	0-7	7
L2 regularization	reg_lambda	0-7	2
Minimum child weight	min_child_weight	1-7	3
Gamma	gamma	0-7	0
Learning rate	learning_rate	0.01-0.3	0.116
Subsample fraction	subsample	0.5-1.0	0.5
Column sample by tree fraction	colsample_bytree	0.5-1.0	1.0

Table S3. Spearman rank correlation coefficients for the explanatory variables. The correlation coefficient was calculated using all data from each month.

	Temperature	Boundary Layer Height	Wind Speed	Mean Sea Level Pressure	Low Cloud Cover	Cloud Water Content	Relative Humidity	SW Solar Radiation
Boundary Layer Height	0.31							
Wind Speed	0.028	0.32						
Mean Sea Level Pressure	-0.51	-0.056	-0.12					
Low Cloud Cover	-0.28	0.075	0.16	0.075				
Cloud Water Content	0.77	0.2	0.086	-0.53	0.23			
Relative Humidity	0.25	-0.19	0.069	-0.24	0.58	0.6		
SW Solar Radiation	0.027	**0.00093**	-0.23	0.42	-0.34	-0.23	-0.076	
LW Surface Radiation	0.9	0.24	0.087	-0.53	0.057	0.93	0.52	-0.11

The values in bold indicate p-values below 0.05.

Table S4. Physical constraints and imputed values applied to the extrapolated explanatory variables for the future projections. The percentage below and above the lower and upper constraint, respectively, are for the year 50 projection only to display the maximum percentage of imputed values during any month.

Explanatory variable	Lower constraint	Upper constraint	Lower imputed value	Upper imputed value	% < lower constraint	% > upper constraint
Boundary Layer Height	<20	N/A	20	N/A	0.29	N/A
Cloud Water Content	<0	N/A	0	N/A	0.12	N/A
Relative Humidity	<0	>100	0	100	0	0.06
Low Cloud Cover	<0	>1	0	1	0.15	4.76

References

1. Sverdrup, H. U. On Conditions for the Vernal Blooming of Phytoplankton. *ICES Journal of Marine Science* **18**, 287–295 (1953).
2. Sharma, S. *et al.* Influence of transport and ocean ice extent on biogenic aerosol sulfur in the Arctic atmosphere. *Journal of Geophysical Research: Atmospheres* **117**, (2012).
3. Sharma, S. *et al.* A Factor and Trends Analysis of Multidecadal Lower Tropospheric Observations of Arctic Aerosol Composition, Black Carbon, Ozone, and Mercury at Alert, Canada. *Journal of Geophysical Research: Atmospheres* **124**, 14133–14161 (2019).
4. Nielsen, I. E. *et al.* Biogenic and anthropogenic sources of aerosols at the High Arctic site Villum Research Station. *Atmos. Chem. Phys.* **19**, (2019).
5. Jang, S. *et al.* Large seasonal and interannual variations of biogenic sulfur compounds in the Arctic atmosphere (Svalbard; 78.9° N, 11.9° E). *Atmos. Chem. Phys.* **21**, 9761–9777 (2021).
6. Nøjgaard, J. K. *et al.* A local marine source of atmospheric particles in the High Arctic. *Atmospheric Environment* **285**, 119241 (2022).
7. Becagli, S. *et al.* Biogenic Aerosol in the Arctic from Eight Years of MSA Data from Ny Ålesund (Svalbard Islands) and Thule (Greenland). *Atmosphere* **10**, 349 (2019).
8. Park, K.-T. *et al.* Linking atmospheric dimethyl sulfide and the Arctic Ocean spring bloom. *Geophysical Research Letters* **40**, 155–160 (2013).
9. Henson, S. A., Dunne, J. P. & Sarmiento, J. L. Decadal variability in North Atlantic phytoplankton blooms. *Journal of Geophysical Research: Oceans* **114**, (2009).
10. Simpson, W. R., Brown, S. S., Saiz-Lopez, A., Thornton, J. A. & Glasow, R. Tropospheric halogen chemistry: sources, cycling, and impacts. *Chem Rev* **115**, 4035–62 (2015).
11. Barrie, L. A., den Hartog, G., Bottenheim, J. W. & Landsberger, S. Anthropogenic aerosols and gases in the lower troposphere at Alert Canada in April 1986. *J Atmos Chem* **9**, 101–127 (1989).
12. Sirois, A. & Barrie, L. A. Arctic lower tropospheric aerosol trends and composition at Alert, Canada: 1980–1995. *Journal of Geophysical Research: Atmospheres* **104**, 11599–11618 (1999).
13. Quinn, P. K. *et al.* A 3-year record of simultaneously measured aerosol chemical and optical properties at Barrow, Alaska. *Journal of Geophysical Research: Atmospheres* **107**, AAC 8-1 (2002).
14. Becagli, S. *et al.* Relationships linking primary production, sea ice melting, and biogenic aerosol in the Arctic. *Atmospheric Environment* **136**, 1–15 (2016).
15. Graham, R. M., Hudson, S. R. & Maturilli, M. Improved Performance of ERA5 in Arctic Gateway Relative to Four Global Atmospheric Reanalyses. *Geophysical Research Letters* **46**, 6138–6147 (2019).
16. Ramli, H. M. & Esler, J. G. Quantitative evaluation of numerical integration schemes for Lagrangian particle dispersion models. *Geoscientific Model Development* **9**, 2441–2457 (2016).
17. Stohl, A. Characteristics of atmospheric transport into the Arctic troposphere. *Journal of Geophysical Research* **111**, (2006).
18. Stohl, A. *et al.* Black carbon in the Arctic: the underestimated role of gas flaring and residential combustion emissions. *Atmos. Chem. Phys.* **13**, 8833–8855 (2013).
19. Barnes, I., Hjorth, J. & Mihalopoulos, N. Dimethyl Sulfide and Dimethyl Sulfoxide and Their Oxidation in the Atmosphere. *Chem. Rev.* **106**, 940–975 (2006).

20. von Glasow, R. & Crutzen, P. J. Model study of multiphase DMS oxidation with a focus on halogens. *Atmospheric Chemistry and Physics* **4**, 589–608 (2004).

21. Hoffmann, E. H. *et al.* An advanced modeling study on the impacts and atmospheric implications of multiphase dimethyl sulfide chemistry. *Proceedings of the National Academy of Sciences* **113**, 11776–11781 (2016).

22. Chen, Q., Sherwen, T., Evans, M. & Alexander, B. DMS oxidation and sulfur aerosol formation in the marine troposphere: a focus on reactive halogen and multiphase chemistry. *Atmospheric Chemistry and Physics* **18**, 13617–13637 (2018).

23. Lenton, T. M. Arctic Climate Tipping Points. *AMBIO* **41**, 10–22 (2012).

24. Armstrong McKay, D. I. *et al.* Exceeding 1.5°C global warming could trigger multiple climate tipping points. *Science* **377**, eabn7950 (2022).

25. Heslin-Rees, D. *et al.* From a polar to a marine environment: has the changing Arctic led to a shift in aerosol light scattering properties? *Atmos. Chem. Phys.* **20**, 13671–13686 (2020).

26. Pernov, J. B. *et al.* Increased aerosol concentrations in the High Arctic attributable to changing atmospheric transport patterns. *npj Clim Atmos Sci* **5**, 1–13 (2022).