# Supplementary Materials

## *Texts*

**Text S1.** The interdependency among LAI, surface properties (Albedo, FSR and SW), environmental conditions (T2M and RH) and heat fluxes (SSH and SLH) under greening background.

In the context of vegetation greening, LAI increases significantly. The increased LAI could directly alter the surface properties: (1) Surface albedo decreases, favoring an increase in surface net radiation. (2) Surface roughness increases, enhancing turbulent mixing and facilitating heat transfer. (3) SW decreases, for vegetation growth and evapotranspiration (transpiration and soil evaporation) both consume SW, although soil evaporation may be slightly reduced due to the radiation interception by leaves (Supplementary Figure S6).

The increase in LAI also increases stomatal conductance, improving transpiration efficiency. The combination of evapotranspiration, increased radiation and turbulence changes the way energy is allocated. Since transpiration is the main component of evapotranspiration, the SLH usually increases in non-arid areas such as the EEP, WSP and NEP region. As both net surface radiation and SLH increase, it is difficult to determine whether the SSH

is increasing or decreasing. The increased SLH indicates that more moisture is being transferred to the atmosphere, which increases RH but has a cooling effect on T2M. Due to the warming effect from the increased surface radiation, the change in T2M is also difficult to determine. The T2M and RH also change the SSH and SLH by affecting the atmospheric evaporative demand.

Vegetation greening can induce a range of changes in surface properties, heat fluxes and environmental conditions, and in turn there are complex feedbacks from T2M, RH, SW and surface radiation to the vegetation.

Note that the main dependencies between LAI and several types of variables in the context of greening are briefly listed here. In the real world, the relationships between the above variables will be more complex.

**Text S2.** The $R^2$ of partial least squares regression models.

The models obtained by PLSR (combined with 10-fold cross validation) were evaluated according to $R^2$. It was found that EEP (0.77- 0.80) and WSP (0.68-0.76) had the highest $R^2$, followed by NEP (0.70-0.76), and MGP had a slightly lower $R^2$ (0.51-0.64) (Supplementary Figure S7). The above indicates that vegetation (LAI), surface properties (Albedo, FSR, SW), environmental conditions (T2M, RH), and heat fluxes (SSH, SLH) can explain 58%-80% of the causes of TCC variability except for NDF in MGP, which we believe is reasonable because other factors could affect clouds, such as aerosols. The study reported that aerosols from the 2014 Icelandic volcanic eruption increased cloud cover by 10% [72], indicating that the contribution of aerosols

to TCC is not negligible. All four regions in this study are densely populated areas with relatively high aerosol emissions from anthropogenic activities, and the dusty weather in the MGP is also an important source of aerosols, thus the cloud cover must be disturbed by aerosols. Anyway, the eight factors in this study can already explain most of the cloud variability, thus it is reasonable to analyze using the results of PLSR models.

**Text S3.** Multicollinearity diagnostic results for four regions.

In regression modelling, a diagnosis of multicollinearity is usually performed on the independent variables. Multicollinearity can increase the variance of the regression coefficients and lead to deceptive results [73,74]. The variance inflation factor (VIF) is a commonly used metric in statistical analysis, and it is generally assumed that factors with VIF > 5 or VIF > 10 have a large contribution to multicollinearity [75,76]. Here, we consider independent variables with a VIF greater than 10 to have severe multicollinearity with other independent variables. We diagnosed multicollinearity for each pixel in the study area, counted the number of independent variables with a VIF>10, and plotted histograms (Supplementary Figure S16). The histograms show that multicollinearity was present in almost all pixels, making traditional multiple regression equations unsuitable for our study. Therefore, we opted for PLSR, which can overcome the multicollinearity problem.

**Text S4.** Linear trend after accounting for time autocorrelation.

Most current studies have examined the time trends of factors based on the

slopes of one-dimensional regression equations, and this method is used in the main text to facilitate comparisons with the current results. However, if there is autocorrelation in the time series, the slope of the one-dimensional regression equation alone may not accurately reflect the trend of the factor. To eliminate the effect of temporal autocorrelation, we recalculated the trends of TCC and LAI. The first step is to determine whether there is autocorrelation in the time series. If there is no autocorrelation, the linear trend of the time series is obtained directly from the one-dimensional regression equation. If there is autocorrelation, the Autoregressive Integrated Moving Average (ARIMA) model is fitted with the autocorrelation order to obtain the residual series of the original time series. Finally, based on the residual series, the linear trend of the time series is obtained using the Generalized Least Squares (GLS) model. These calculations are carried out in Matlab R2023b. The presence of autocorrelation in the time series is determined using the "autocorr" function. The "arima" and "lscov" functions are used to fit the ARIMA and GLS models, respectively.

Supplementary Figures S17-18 show the new computational results. The new results show that after accounting for temporal autocorrelation, the trends of TCC and LAI are very close to those in the main text and do not affect our main conclusions.

**Text S5.** Calculation of lifting condensation height (LCH).

Referring to previous studies [66,77], we calculated the LCH based on the equation: $LCH = 125(T - T_d)$, where T and $T_d$ represent the air temperature

and dew point temperature at 2 m (unit: K), respectively, both from the ERA5 reanalysis data. The unit of LCH is m.

## *Tables*

**Table S1.** Range and elevation information for 4 regions.

| Area | Longitude Span (°E) | Latitude Span (°N) | Average Elevation (m) |
|---|---|---|---|
| EEP | 32-52 | 55-63 | 146 |
| WSP | 66-86 | 56-64 | 90 |
| MGP | 94-114 | 43-51 | 1464 |
| NEP | 120-129 | 43-49 | 355 |

In determining the scope of EEP and WSP, the following points were considered: (1) they are divided by the Ural Mountains, and to avoid the influence of topographic clouds, they must keep a certain distance from the Ural Mountains; (2) they both belong to high latitudes, and the northern part is adjacent to the Arctic Ocean, where vegetation is scarce, so the northernmost region is not concerned. (3) This study aims to compare the differences between different regions and therefore tries to keep the latitudinal range of them as close as possible.

**Table S2.** Details of datasets used in this study.

| Dataset | Time Span | Resolution | Source | Variables |
|---|---|---|---|---|
| Land cover classification gridded maps from 1992 to the present derived from satellite observations | 1992-2020 | yearly, 300m | European Centre for Medium-Range Weather Forecasts | Land cover class, Change count |
| GLASS LAI (version6) dataset | 2000-2021 | monthly, 0.25°×0.25° | Beijing Normal University | LAI |
| ERA5 monthly averaged data on single levels from 1940 to present | 1940 to present | monthly, 0.25°×0.25° | European Centre for Medium-Range Weather Forecasts | TCC, albedo, T2M, SSH, SLH, SW, $T_d$ |
| ERA5 monthly averaged data on pressure levels from 1940 to present | 1940 to present | monthly, 0.25°×0.25° | European Centre for Medium-Range | RH (1000hpa) |

| | | | | |
|---|---|---|---|---|
| | | | Weather Forecasts | |
| Elevation - Global version - Version 1 | — — | 30 arcseconds | Geospatial Information Authority of Japan | Elevation |

**Table S3.** Abbreviation of terms.

| Abbreviations | Long Name |
|---|---|
| EEP | Eastern European Plain |
| WSP | Western Siberian Plain |
| MGP | Mongolian Plateau |
| NEP | Northeastern China Plain |
| BDF | Deciduous broadleaf forest |
| NEF | Evergreen needle leaved forest |
| MF | Mixed forest |
| NDF | Deciduous needle leaved forest |
| Sparse | Sparse vegetation |
| TCC | Total cloud cover |
| LAI | Leaf area index ($m^2$ $m^{-2}$) |
| T2M | Surface air temperature at 2 m (°C) |
| RH | Relative humidity (%) |
| Albedo | Surface albedo |

| | |
|---|---|
| FSR | Forecast surface roughness (m) |
| SW | Soil water content ($m^3$ $m^{-3}$) |
| SSH | Surface sensible heat flux (W $m^{-2}$) |
| SLH | Surface latent heat flux (W $m^{-2}$) |
| $C_{environment}$ | The contribution of environmental conditions to TCC |
| $C_{surface}$ | The contribution of surface properties to TCC |
| $C_{heat}$ | The contribution of heat fluxes to TCC |

# *Figures*

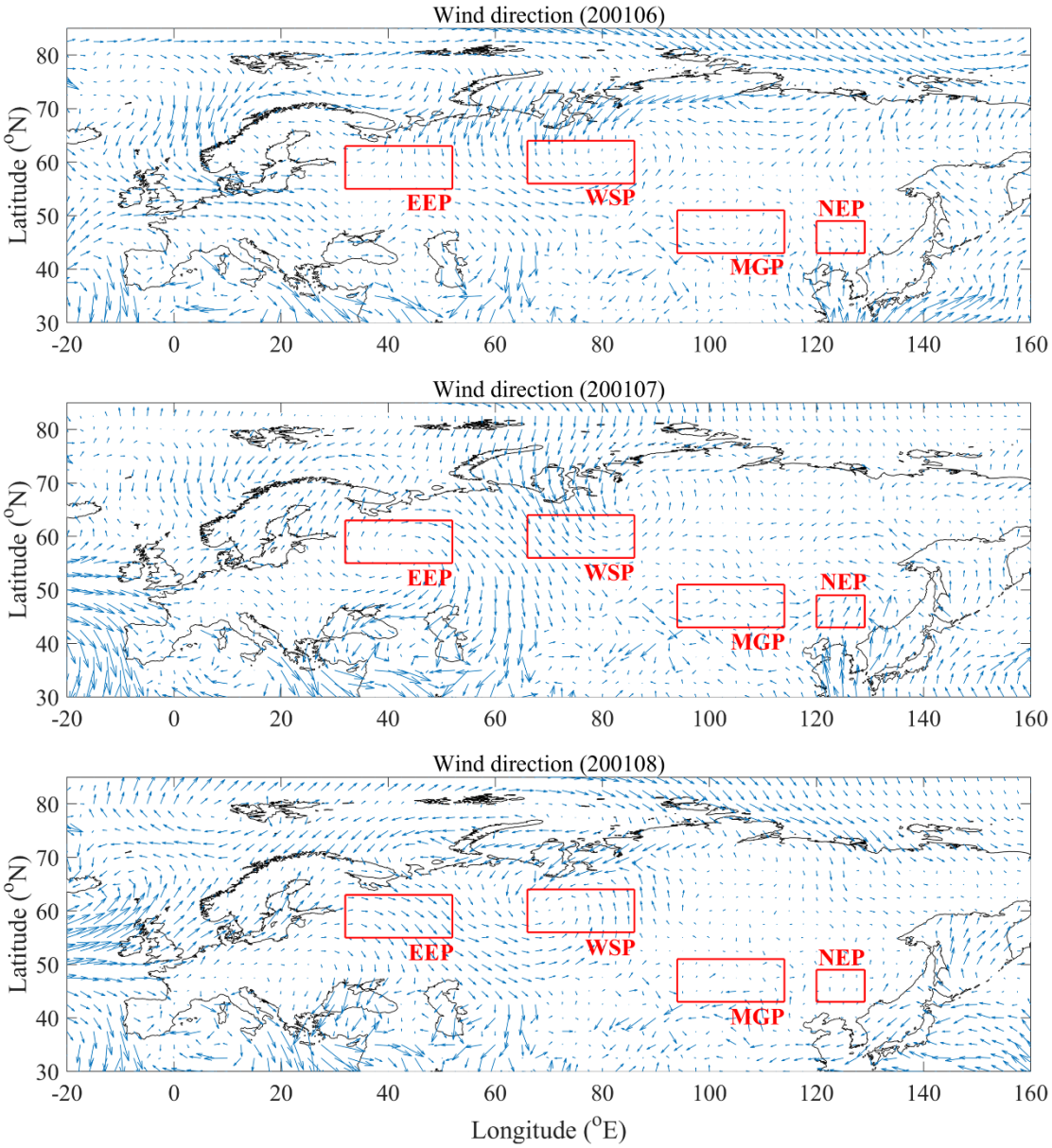



**Figure S1.** The wind direction map of the study area in 2001 (JJA).

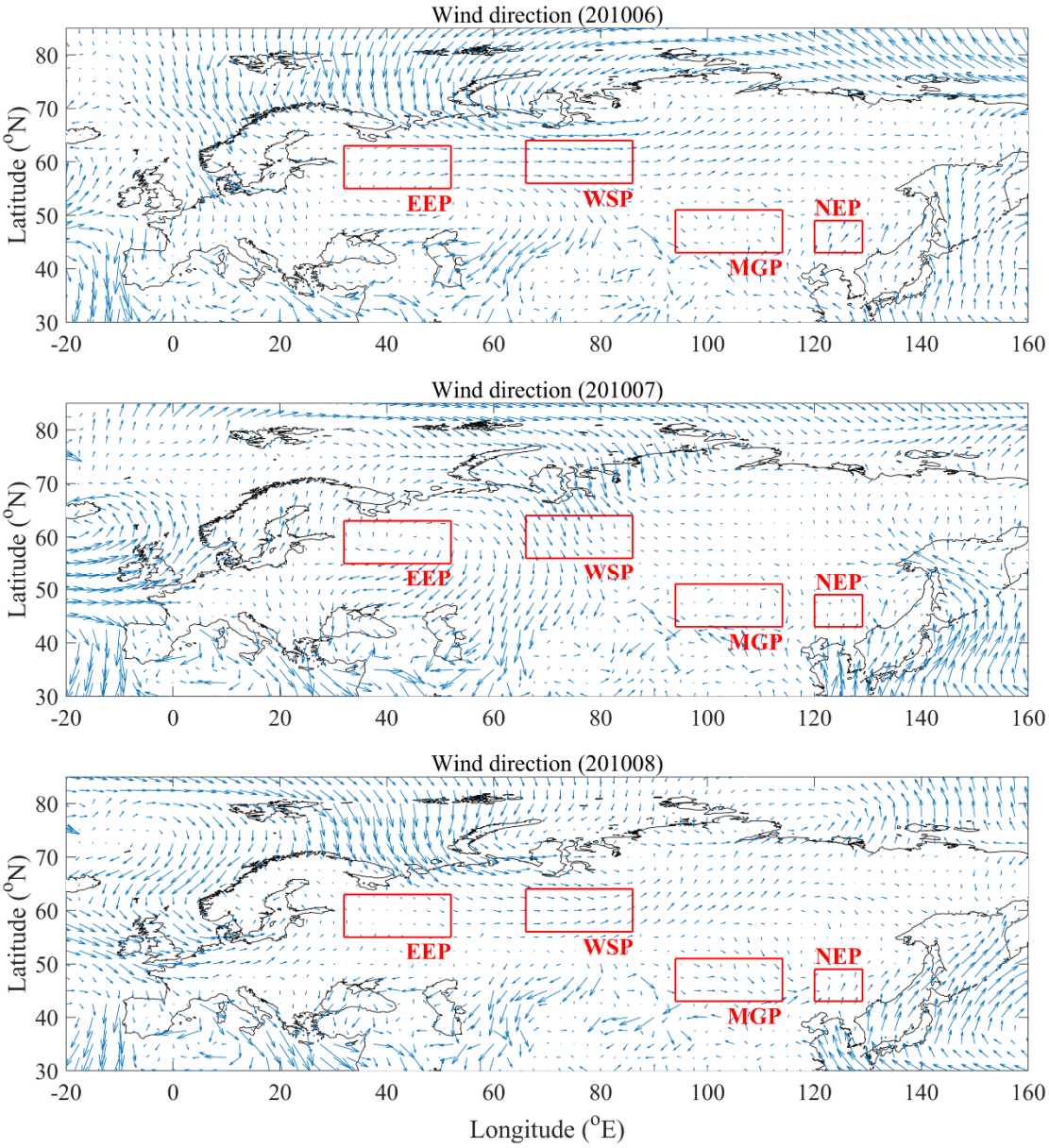



**Figure S2.** The wind direction map of the study area in 2010 (JJA).

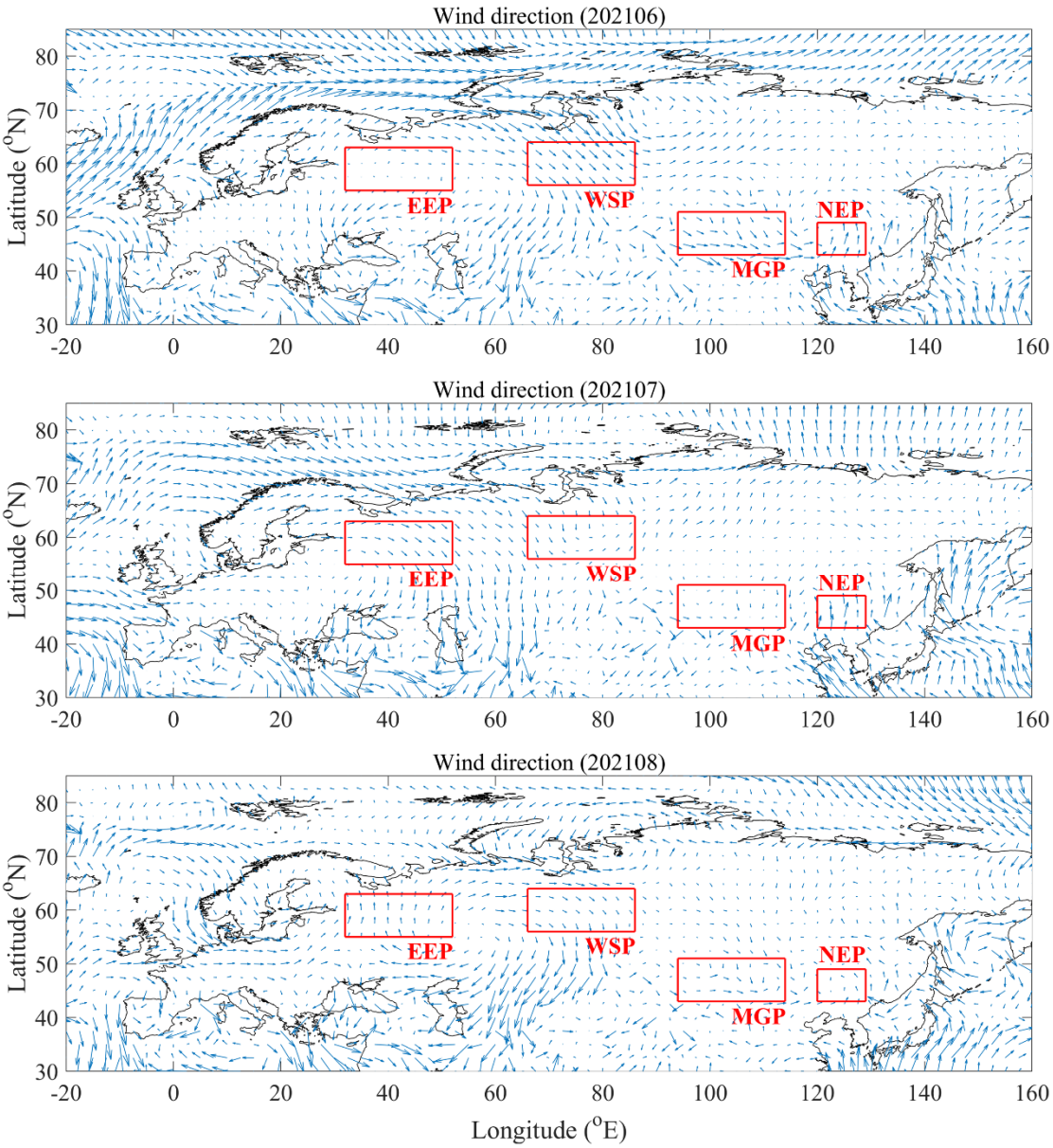



**Figure S3.** The wind direction map of the study area in 2021 (JJA).

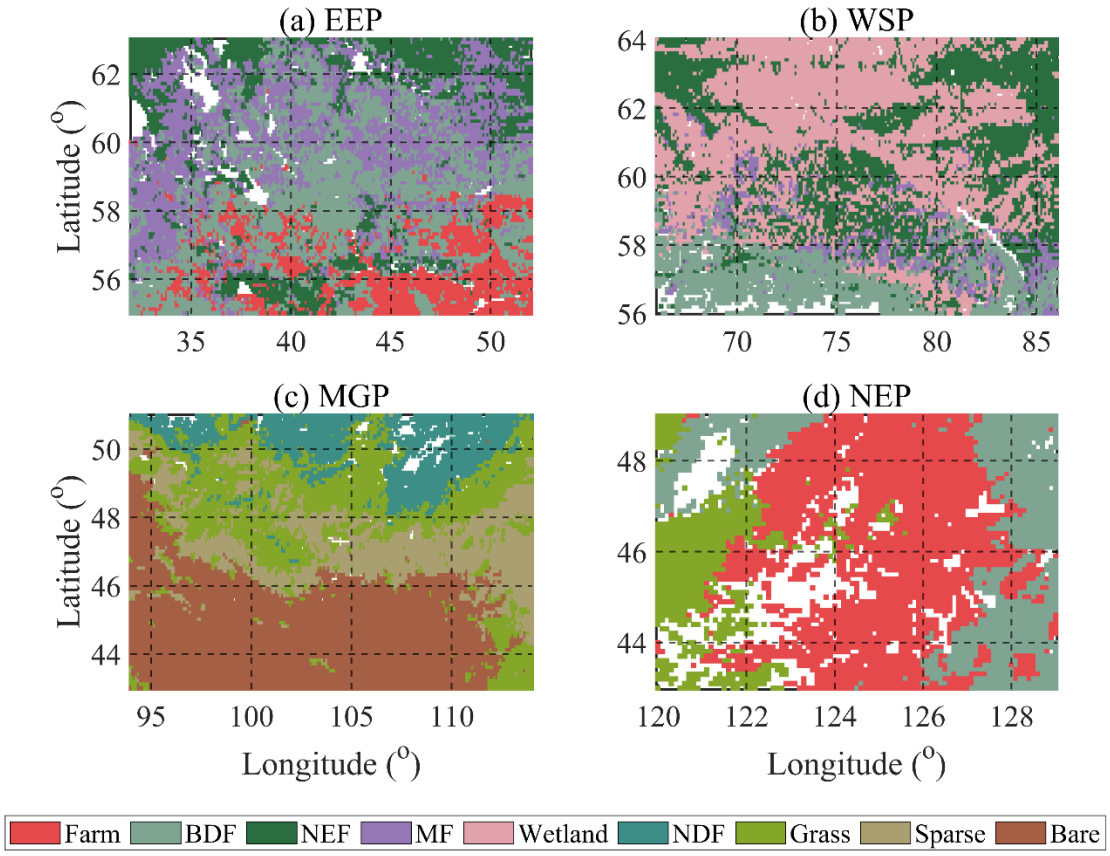



**Figure S4.** Distribution of main ecosystem types in the EEP, WSP MGP and NEP. The blank areas represent the types of ecosystems not concerned in this study. The NEP has many white pixels due to the presence of extensive irrigated farmland.

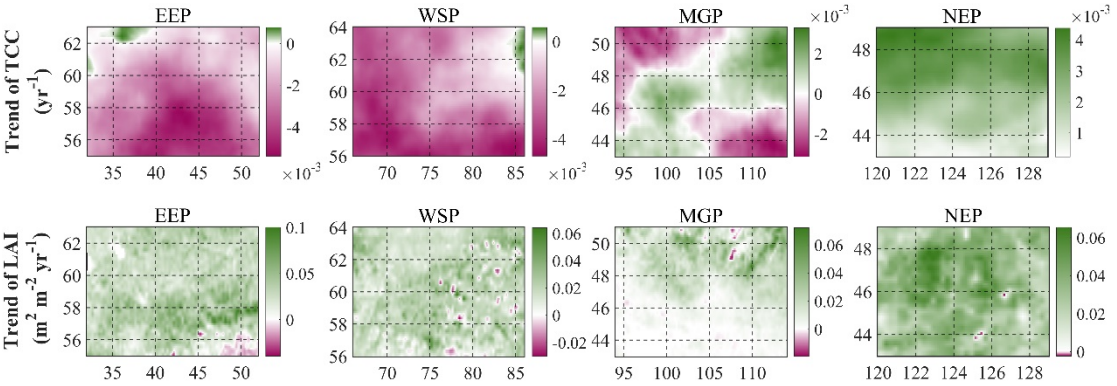



**Figure S5.** The spatial distribution of TCC and LAI trends from 2001 to 2021(JJA).

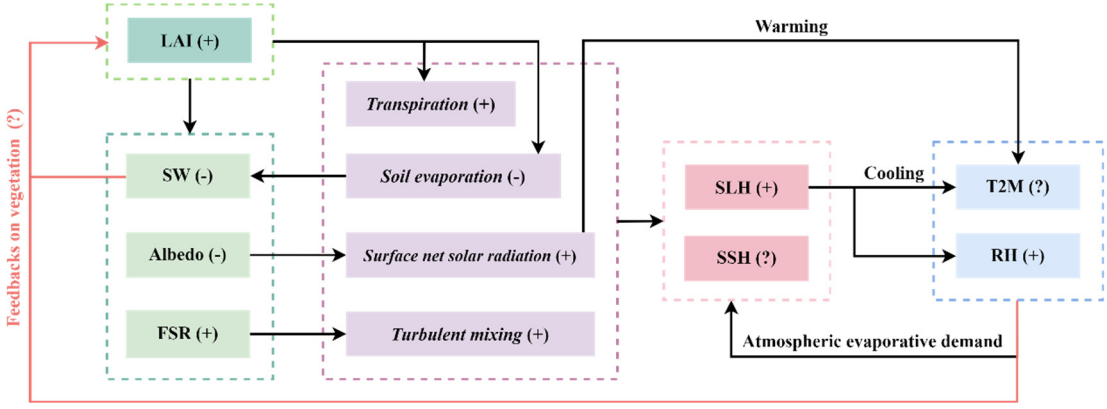



**Figure S6.** The interdependency among LAI, surface properties (Albedo, FSR and SW), environmental conditions (T2M and RH) and heat fluxes (SSH and SLH) under greening background. The black line represents a series of responses triggered by the increased LAI, and the red line represents feedbacks from other variables to LAI. The (+) symbol indicates that variable is strengthened, the (-) symbol indicates that variable is weakened, and the (?) symbol indicates that variable is affected both positively and negatively, with the exact change unknown.

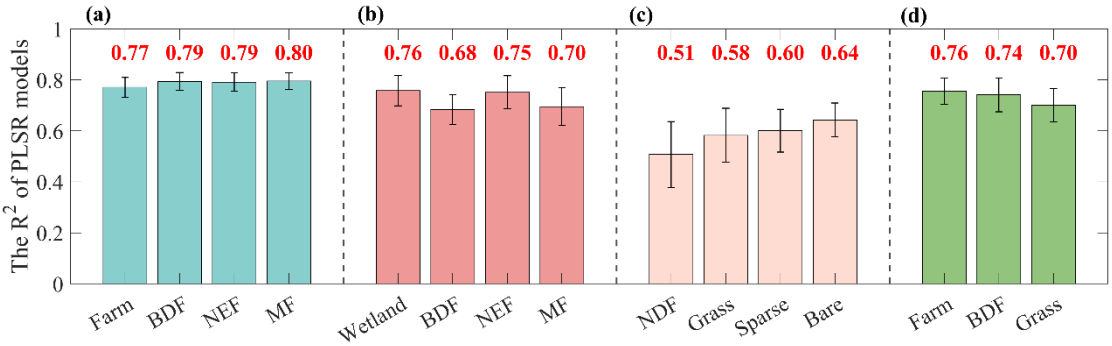



**Figure S7.** The $R^2$ of partial least squares regression models. Where (a) - (d) represent EEP, WSP, MGP and NEP, respectively. The bar chart and red numbers represent the means of $R^2$ in different ecosystems, while the black line represents a standard deviation.

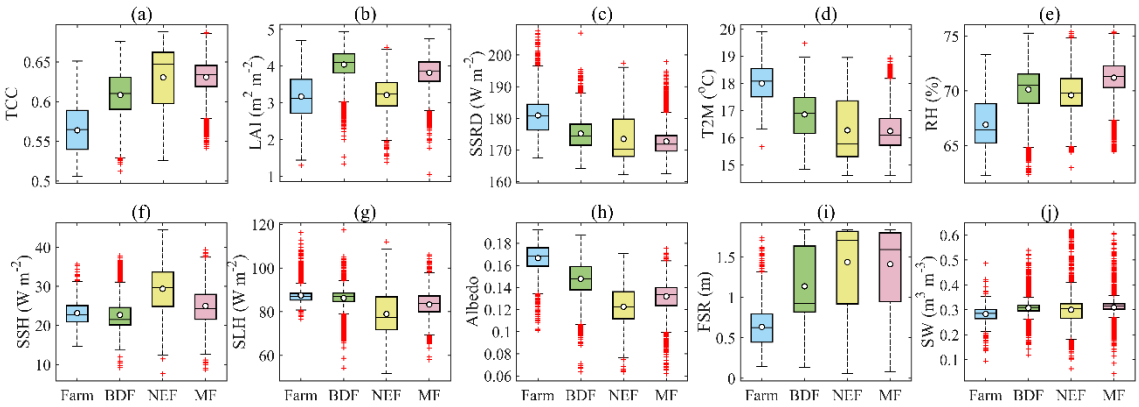



**Figure S8.** Box plots of different factors in EEP (JJA).

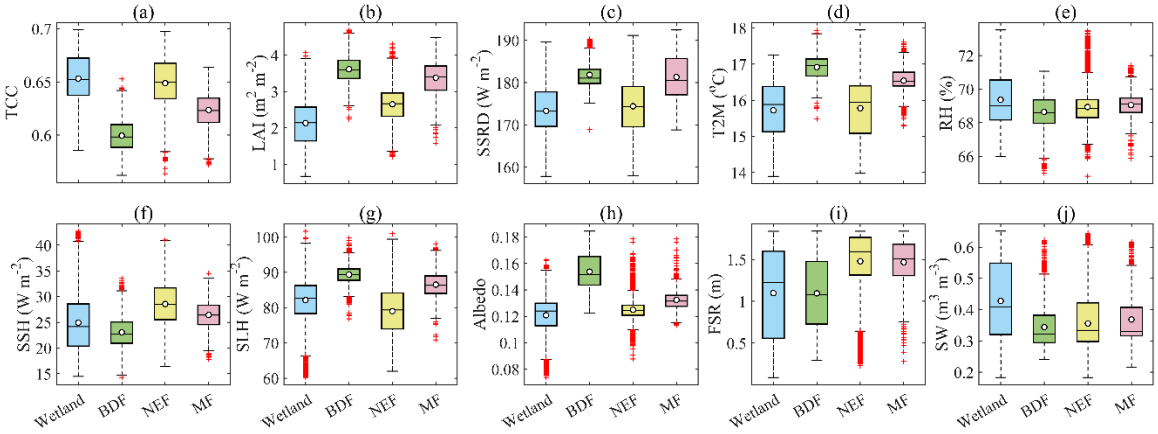



**Figure S9.** Box plots of different factors in WSP (JJA).

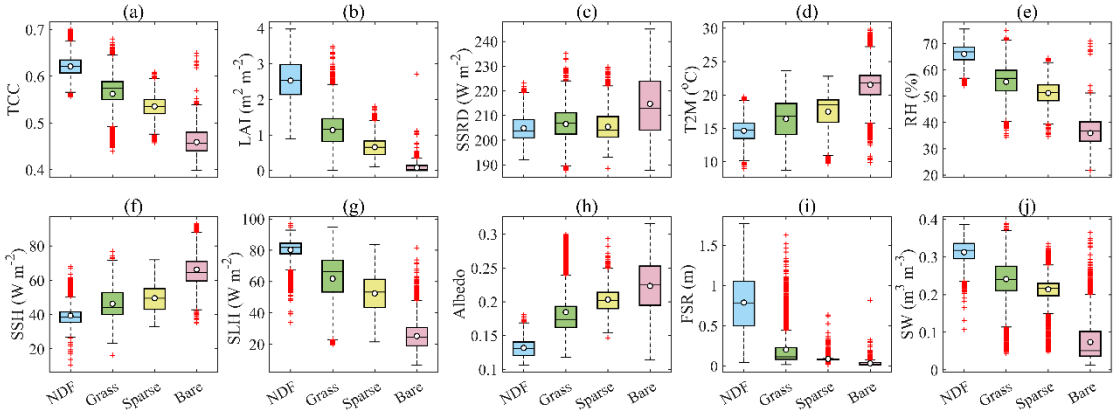



**Figure S10.** Box plots of different factors in MGP (JJA).

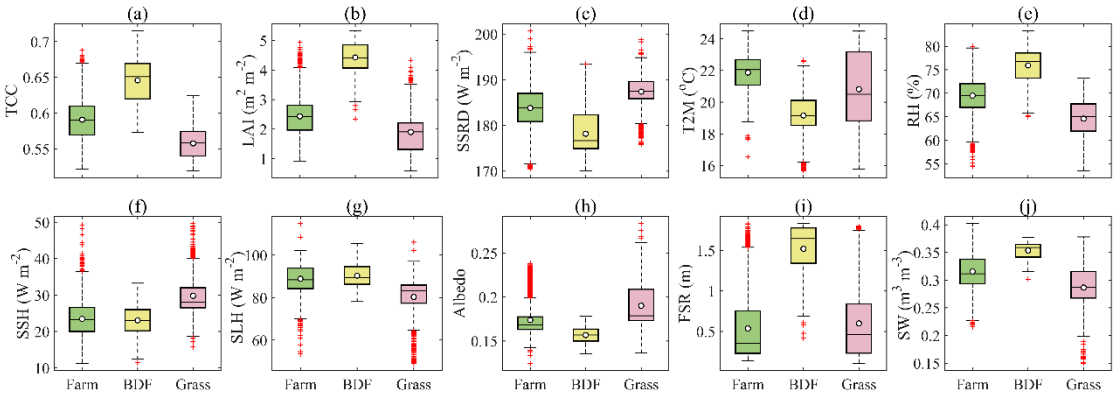



**Figure S11.** Box plots of different factors in NEP (JJA).

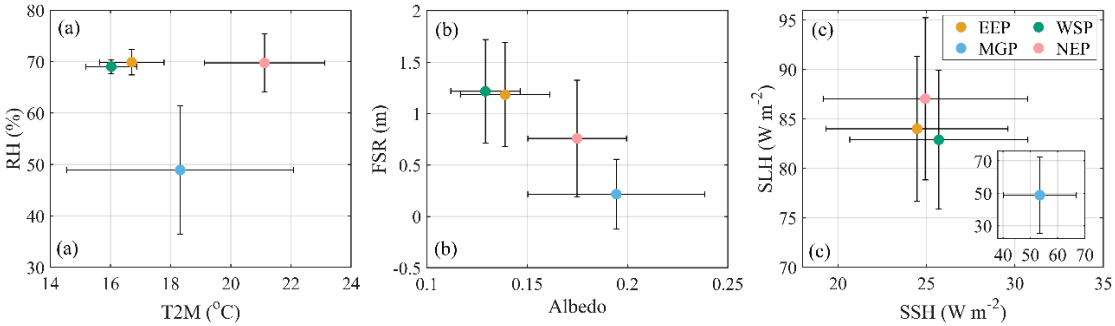



**Figure S12.** Multi-year summer averages of (a) environmental conditions, (b) surface properties, and (c) heat fluxes in four regions. Black line represents 1 standard deviation.

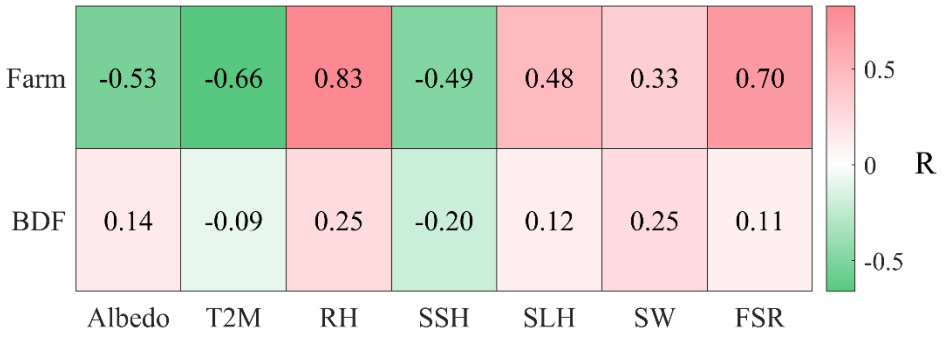



**Figure S13.** Pearson correlation coefficients between LAI and other factors. The first row represents NEP's farm ecosystem and the second row represents EEP's BDF ecosystem. Note: All the above correlation coefficients passed the significance test, i.e., $p<0.05$.

For each pixel, we calculated the multi-year mean values (JJA) of different factors. Then, we used all the pixels in one ecosystem as sample for correlation analysis.

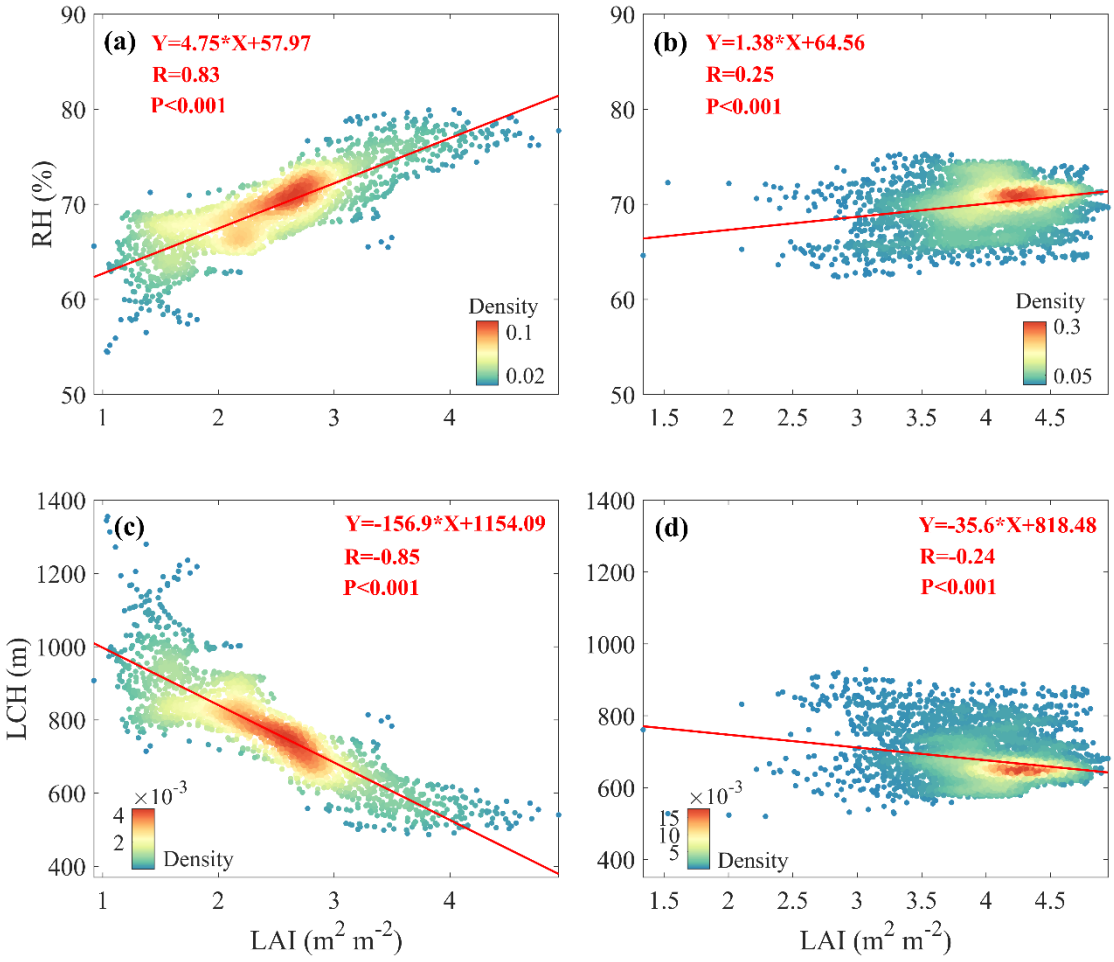



**Figure S14.** Scatter plots of relative humidity (RH) and lifting condensation height (LCH) as a function of leaf area index (LAI). Panels (a) and (c) represent the farmland in the NEP, while panels (b) and (d) represent the forest ecosystem in the EEP.

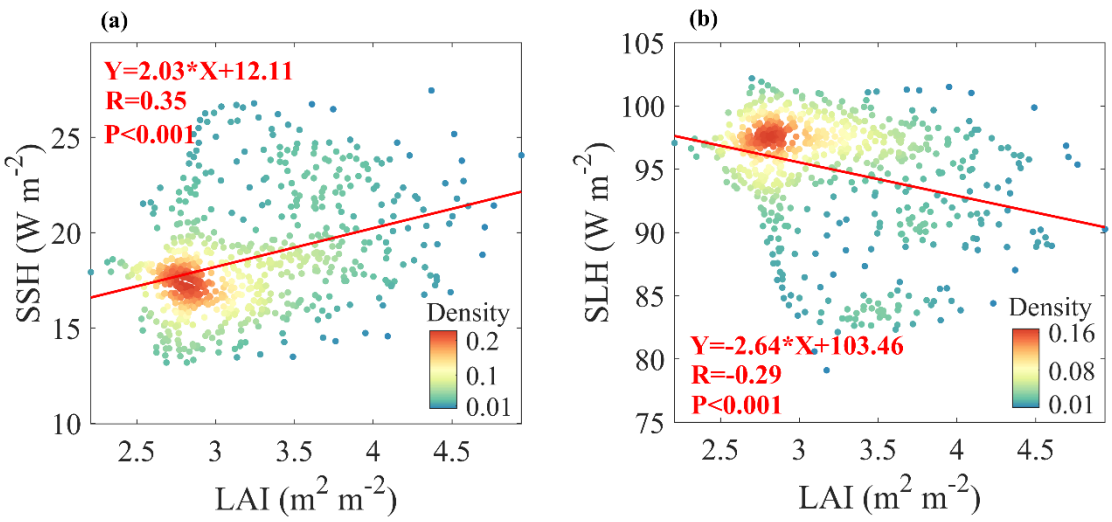



**Figure S15.** The changes in (a) SSH and (b) SLH with LAI in low temperature farmland (NEP).

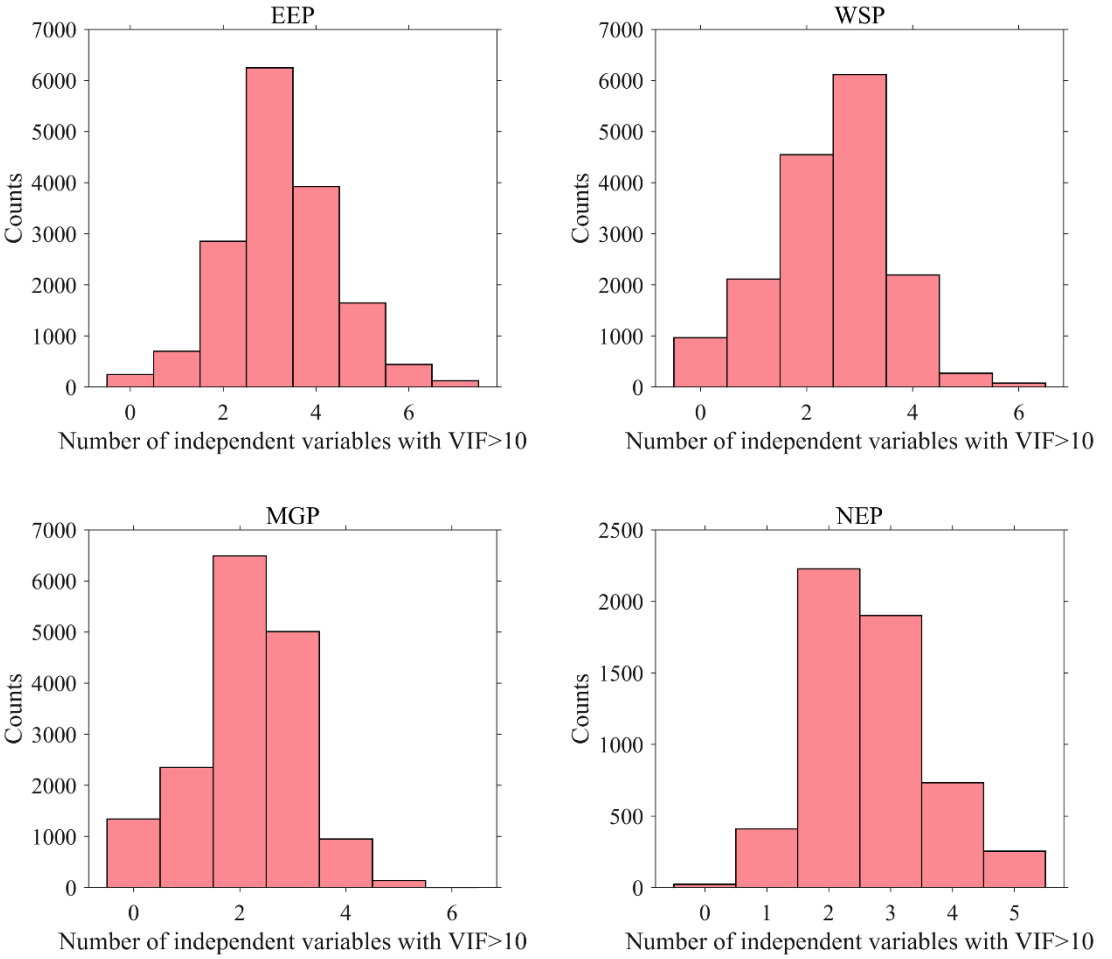



**Figure S16.** Multicollinearity diagnostic results for four regions. The x-axis represents the number of factors with VIF >10 and the y-axis represents the

corresponding number of pixels.

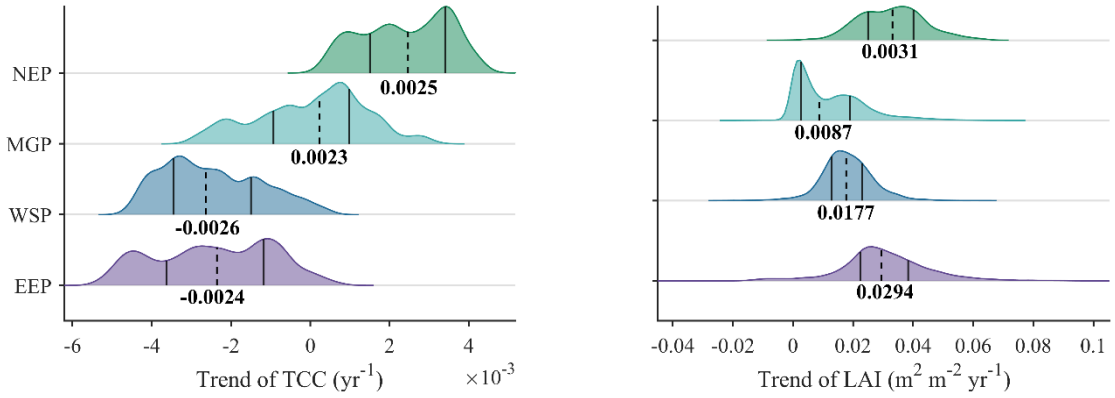



**Figure S17.** The trend distributions of TCC and LAI (JJA), where the black dashed lines and numbers represent the median of the trends, and the black solid lines represent the 25% and 75% quantiles of the trend, respectively. Note that the linear trend here takes into account the time autocorrelation.

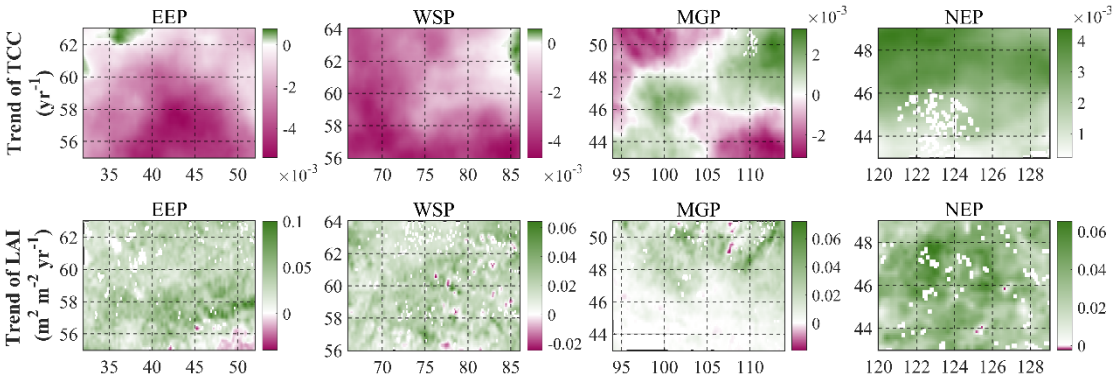



**Figure S18.** The spatial distribution of TCC and LAI trend from 2001 to 2021(JJA). Note that the linear trend here takes into account the time autocorrelation.