




Please cite the Published Version

Li, Hao , Zhou, Yuyu , Zhao, Xiang, Zhang, Xin  and Liang, Shunlin (2024) A dataset of 0.05-degree leaf area index in China during 1983–2100 based on deep learning network. *Scientific Data*, 11 (1). 1122 ISSN 2052-4463

DOI: <https://doi.org/10.1038/s41597-024-03948-z>

Publisher: Nature Publishing Group UK

Version: Published Version

Downloaded from: <https://e-space.mmu.ac.uk/636139/>

Usage rights:  [Creative Commons: Attribution-Noncommercial-No Derivative Works 4.0](https://creativecommons.org/licenses/by-nc-nd/4.0/)

Additional Information: This is an open access article which first appeared in *Scientific Data*

Data Access Statement: The code that calibrates the model and predicts the future LAI is based on Python 3.9.12, working with the key libraries numpy, gdal, and Pytorch. The code can be found on the GitHub repository (https://github.com/NoSleepTonight-0712/LAI_downscaling_upload).

Enquiries:

If you have questions about this document, contact openresearch@mmu.ac.uk. Please include the URL of the record in e-space. If you believe that your, or a third party's rights have been compromised through this document please see our Take Down policy (available from <https://www.mmu.ac.uk/library/using-the-library/policies-and-guidelines>)



OPEN

DATA DESCRIPTOR

A dataset of 0.05-degree leaf area index in China during 1983–2100 based on deep learning network

Hao Li^{1,2}, Yuyu Zhou^{3,4}, Xiang Zhao^{1,2}✉, Xin Zhang⁵ & Shunlin Liang⁶

Leaf Area Index (LAI) is a critical parameter in terrestrial ecosystems, with high spatial resolution data being extensively utilized in various research studies. However, LAI data under future scenarios are typically only available at 1° or coarser spatial resolutions. In this study, we generated a dataset of 0.05° LAI (F0.05D-LAI) from 1983–2100 in a high spatial resolution using the LAI Downscaling Network (LAIDN) model driven by inputs including air temperature, relative humidity, precipitation, and topography data. The dataset spans the historical period (1983–2014) and future scenarios (2015–2100, including SSP-126, SSP-245, SSP-370, and SSP-585) with a monthly interval. It achieves high accuracy ($R^2 = 0.887$, $RMSE = 0.340$) and captures fine spatial details across various climate zones and terrain types, indicating a slightly greening trend under future scenarios. F0.05D-LAI is the first high-resolution LAI dataset and reveals the potential vegetation variation under future scenarios in China, which benefits vegetation studies and model development in earth and environmental sciences across present and future periods.

Background & Summary

Leaf Area Index (LAI) is an important indicator of vegetation defined as one half of the total green leaf area per unit horizontal ground surface area^{1,2}. It is a critical parameter for depicting vegetation situation and has been used in assessing vegetation dynamics³, building land surface models⁴, assisting crop yield estimation⁵ and supporting forest management^{6,7}. Projecting future vegetation changes under different climate scenarios requires using model simulations. The Coupled Model Intercomparison Project (CMIP) has tried to use climate models to make such projections, and the sixth phase – usually known as CMIP 6 – started years before⁸. These models which predict global variables including precipitation, air temperature, humidity, and LAI are called Global Climate Models (GCMs). GCMs contribute significantly to future climate research on a global scale and are an important basis for analyzing global change⁹. LAI data from GCMs has different resolutions, usually 100 or 250 km, such as 100 km data from BCC-CSM2-MR¹⁰ and 250 km data from UKESM1-0-LL¹¹. Some higher resolution GCMs like HighResMIP¹² can provide LAI at 25 km resolution, but it only simulates future vegetation under one scenario (highres-future). This limitation in resolution or scenario coverage restricts the potential usage of GCMs. Therefore, downscaling is necessary to be applied to GCMs to improve the applicability of GCM data for local studies.

There are mainly two kinds of downscaling methods: dynamic downscaling and statistical downscaling¹³. A good example of dynamic downscaling is Regional Climate Models (RCMs). However, generating an RCM is computationally expensive, and its accuracy is restricted by the original GCM¹⁴. On the contrary, statistical downscaling is easily applicable, and it can further remove bias in GCMs¹³, allowing the simulation to be close to the truth. Perfect Prognosis (PP) is one of the structures to fulfill the statistical downscaling. It links the large-scale observed predictors to local-scale observed predictands by a statistical model, and then applies the

¹State Key Laboratory of Remote Sensing Science, Faculty of Geographical Science, Beijing Normal University, Beijing, 100875, China. ²Beijing Engineering Research Center for Global Land Remote Sensing Products, Faculty of Geographical Science, Beijing Normal University, Beijing, 100875, China. ³Institute for Climate and Carbon Neutrality, The University of Hong Kong, Hong Kong SAR, 999077, China. ⁴Department of Geography, The University of Hong Kong, Hong Kong, 999077, China. ⁵School of Computing, Mathematics and Digital Technology, Manchester Metropolitan University, Manchester, M1 5GD, UK. ⁶Jockey Club STEM Laboratory of Quantitative Remote Sensing, Department of Geography, The University of Hong Kong, Hong Kong, 999077, China. ✉e-mail: zhaoxiang@bnu.edu.cn

model to predictors from GCMs to generate downscaled results¹³. Various statistical models are used in the PP downscaling procedure such as the General Linear Model (GLM)¹³, Convolution Neural Network (CNN)¹⁵, and Quantile Mapping¹⁶. Recent research has explored the integration of super-resolution techniques, like applying Skip connections and Upsampling in Super Resolution Deep Residual Network (SRDRN)^{17,18} to enhance downscaling performance. Other architectures may also be suitable for downscaling, such as stacking residual blocks in Super Resolution Residual Network (SR-ResNet)¹⁹ and the usage of PixelShuffle in Enhanced Deep Super-resolution Network (EDSR)²⁰.

Existing downscaling methods designed for precipitation and temperature are not fully applicable for downscaling LAI because of the different target resolutions. Much existing research downscaled precipitation or temperature by a factor of 4, which means from 100 km to 25 km or 200 km to 50 km^{15,21}. Nevertheless, LAI needs a higher resolution to be useful. A review⁷ of LAI indicated that the dataset with the lowest spatial resolution among major global long-term LAI products is GIMMS LAI (1/12°) which is used widely in global and regional vegetation research^{2,22}, while other moderate-resolution data like GLOBMAP²³ have even higher resolution. Focusing on application, research aiming at grasslands²⁴, droughts²⁵ and streamflow²⁶ in provinces, countries and continents used LAI data with a resolution of about 0.1°. Some other applications like urban green space planning need even higher resolution²⁷.

The purpose of this study is to provide Future 0.05° LAI (F0.05D-LAI) data across historical and future scenarios in a high resolution in China, allowing researchers to understand possible vegetation patterns in the future, expanding the application of GCMs from the global scale to the regional scale. China is a large country with an intricate climate and vegetation structure. The climate in China influenced by the monsoon and the Qinghai-Tibet Plateau, exhibits significant variation from one region to another, resulting in a diverse range of vegetation. Nowadays, the Chinese government is making an effort to environmental protection and vegetation restoration, launching multiple ecological programs like the ‘Three-North Shelter Forest Program’, and LAI is used in detecting fragile regions²⁸. Our research declared and applied a deep learning network to downscale the LAI under a scaling ratio of 20 (from 1° to 0.05°) and did a comparison with other benchmarks. We also used our data to do a simple analysis of future vegetation and its degradation risk, showing an example of the data usage. Although the data represents potential LAI only under four scenarios and may be influenced by future human activities, it remains a valuable support dataset for analyzing future vegetation, assessing degradation risks and growth potential.

Methods

Data source. The data used in this study can be summarized as low-resolution data (climate data) and high-resolution data (topography and LAI reference data).

The main low-resolution data in this work include climate data (air temperature, relative humidity and precipitation) from CMIP 6¹⁰ and European Centre of Medium-Range Weather Forecasts Reanalysis v5 (ERA-5)²⁹.

The CMIP 6 climate simulation data are derived from the BCC-CSM2-MR experiment¹⁰. BCC-CSM2-MR is a medium-resolution climate system model developed by the National (Beijing) Climate Center with a time range of 1850–2100, with a temporal resolution of 1 month and a spatial resolution of 1°. BCC provides a detailed validation of climate factors especially in East Asia, making it a widely used model in China or East Asia climate projection. In the future period covered by CMIP6 (2015–2100), we selected four Shared Socio-economic Pathways (SSPs): SSP-126, SSP-245, SSP-370, and SSP-585 scenarios to represent the future possibilities since they are the high-priority scenarios among SSPs, and most climate models usually simulate these four scenarios^{30,31}. We used air temperature and relative humidity at different pressure levels (1000hpa, 850hpa, 700hpa, and 500hpa) along with ground precipitation data as model input, while the LAI was utilized for result comparison purposes.

ERA-5 reanalysis dataset was selected as the source of climate data, with a 0.25° horizontal resolution for the period from 1940 onwards²⁹. The used variables are the same as CMIP 6 climate data except for LAI. ERA-5 reanalysis data is widely used in downscaling research^{18,32}, and it is also a recommended data source in the VALUE framework²¹.

Two main high-resolution data in this work include Global Land Surface Satellite (GLASS) LAI³³ and ALOS World 3D – 30 m Digital Elevation Model (AW3D30 DEM)³⁴.

GLASS LAI was selected as a high-resolution product in this work to represent the reference LAI. GLASS LAI product is a widely used LAI dataset with 0.05° horizontal resolution and 8 days temporal resolution from 1981 to 2018³³. Compared with MODIS LAI, which only ranges from 2000 to 2023, GLASS LAI can provide better data coverage on the time range, increasing the sample size that can be used in training. The monthly LAI data was calculated from the average LAI in the month.

AW3D30 DEM³⁴ data derived from the Japan Aerospace Exploration Agency was also used in this study since topography has a clear impact on vegetation. It can provide more accurate topography data across latitudes with a 30-meter horizontal resolution³⁵. It is also regarded as a constant across years since humanity has not been able to drastically change the terrain.

Downscaling procedure. The downscaling process includes three primary parts: Data Acquisition, LAI Downscaling Network (LAIDN) Model Training, and Downscaling LAI. ‘Data Acquisition’ comprises downloading data and data preprocessing. ‘LAIDN Model Training’ contains model fitting and model validation. Finally, ‘Downscaling LAI’ includes bias-adjustment, producing downscaled data, data post-processing, and its practical application.

Data acquisition. This part is shown in the top frame of Fig. 1. We first preprocessed our data, which included clipping them to the range of China and resampling. To avoid potential biases caused by convolution or window-based algorithms, data were all clipped to a rectangular region spanning from 5.50°N, 55.50°E to 71.75°N, 155.75°E, ensuring a buffer zone along China’s borders.

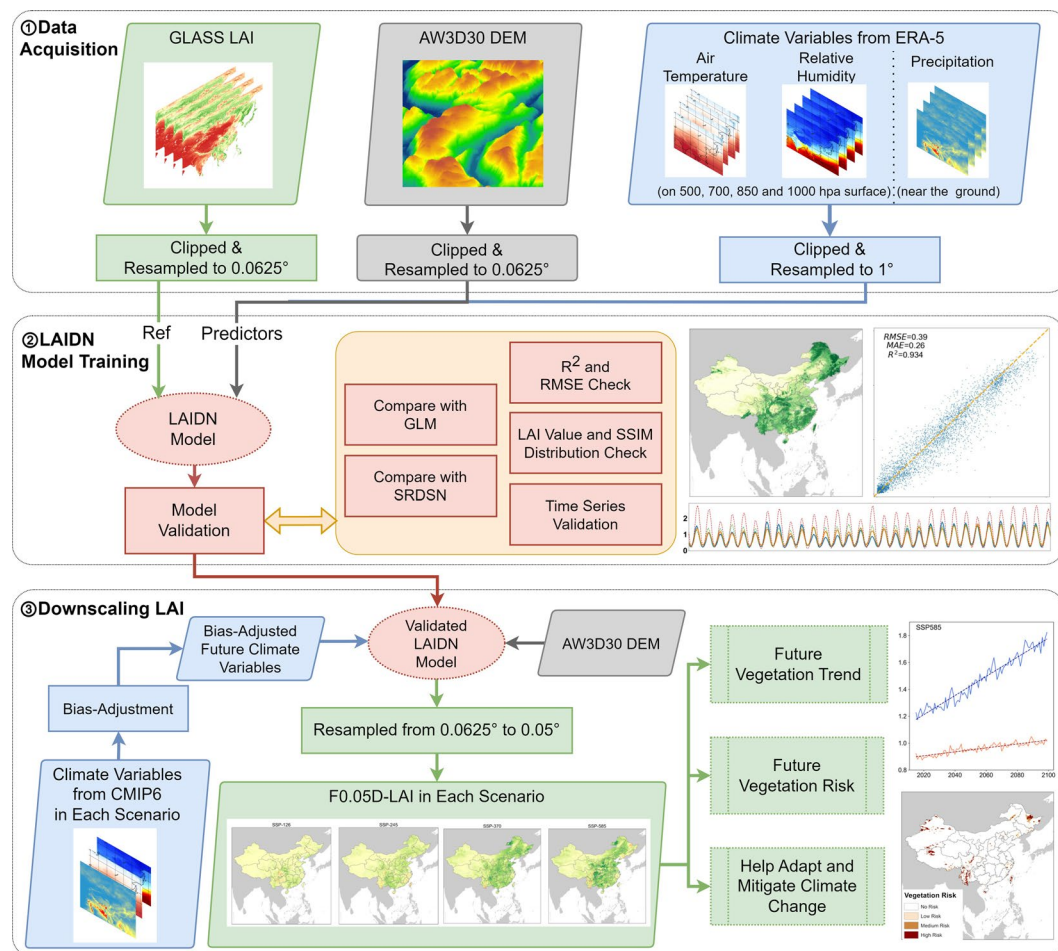


Fig. 1 Downscaling procedure.

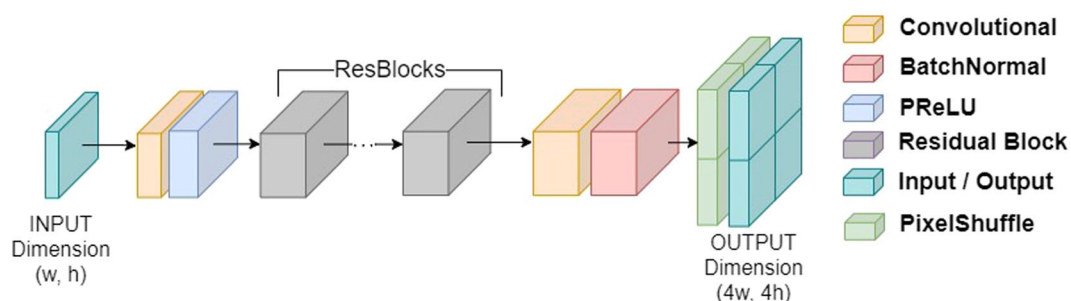


Fig. 2 The structure of the downscaling block.

Then we did some resampling to bridge the resolution gap between different data. Low-resolution data, including air temperature, relative humidity, and precipitation derived from ERA-5, were resampled to a horizontal resolution of 1°. High-resolution data including GLASS LAI and DEM, were resampled to a horizontal resolution of 0.0625° (1/16°).

LAIDN model training. The model claimed and used in our study is the LAIDN, utilizing residual blocks³⁶, PixelShuffle³⁷, and the Parametric ReLU (PReLU)³⁸. It is a deep learning network receiving climate data and topography constant data as input, GLASS LAI as a reference and yields the downscaled high-resolution LAI data as output.

Figure 2 shows the Downscaling Block which is the core structure of LAIDN. It utilized the PixelShuffle block to effectively extend the resolution of low-resolution climate data by compressing their channels. By applying the Downscaling Block twice to the low-resolution climate data, it enhanced their resolution to match that of high-resolution DEM data and GLASS LAI. Subsequently, the concatenated climate and DEM data were processed through additional convolutions and residual blocks to reach the model output with a resolution of 0.0625°. To ensure ease of use, the data was finally interpolated to 0.05° using cubic interpolation.

Mean Square Error (MSE) and the optimizer Adam³⁹ are selected as the loss function and optimizer in model training. We ended the training when the loss function converged. The initial learning rate was 0.02 and was halved when training reached the 5, 25, 50 and 100 epochs. The hardware platform was RTX 3060, the software platform was CUDA12.1, and the training was performed on Windows Subsystem Linux 2 Ubuntu operating system through Pytorch deep learning framework.

After training the model, we validated our model to prove that our model is stable and robust in different situations. To quantitatively assess the downscaling performance of each method, we employed three validation techniques: (1) calculation and comparison of R^2 and Root Mean Square Error (RMSE) between GLASS LAI and the downscaled results, (2) analysis of LAI value distribution charts and Structural Similarity Index Measure (SSIM) distribution charts, and (3) time series validation in different regions of China.

RMSE and R^2 are both generously used metrics in evaluating the correctness. In historical periods, we calculated these metrics between GLASS LAI and downscaled results among training, validation, and test sets. In plotting the LAI distribution charts, the Kernel Density Estimate (KDE) curve of downscaled LAI data from different methods and the GLASS LAI data were painted into the same figure. The result performance will be better if there is less discrepancy between downscaled LAI's and GLASS LAI's KDE curves. Besides, to compare the reservation of spatial textures in results, we used SSIM as the metric to compare the similarity of textures between GLASS LAI and downscaled LAI. SSIM is a widely used metric to compare structural similarity, and it also provides additional information for distinguishing model performance⁴⁰. It calculates the brightness, contrasts, and structures in patches of the image clipped by a sliding window, ranging from -1 to 1 . The larger the value, the more effective the downscaling method is. To show the distribution of spatial similarity, we showed the KDE plot of SSIM calculated in each sliding window rather than an average value of SSIM. The result performance will be better if the peak is near 1 . In the time series validation, we chose six points in different regions of China and painted the LAI time series in each point based on GLASS LAI, GCM LAI, F0.05D-LAI, and benchmarks downscaled LAI. The performance will be better if there is less divergence between the downscaled LAI and the GLASS LAI time series curves.

The two selected benchmarks are GLM and SRDSN. GLM is a simple regression method to do downscaling and is a widely used benchmark method in the VALUE framework⁴¹ and other downscaling research²¹. Recently, many models have shown a better performance than GLM, including deep-learning-based models like SRDSN¹⁷. SRDSN is a novel super-resolution network that can migrate from different variables and shows better downscaling results. Therefore, we selected GLM and SRDSN as benchmarks.

Downscaling LAI. This part is shown in the bottom frame of Fig. 1. Before we produce the final downscaled result, we should first do bias-adjustment on CMIP 6 climate data. The adjustment is for matching the PP assumption. The “PP assumption” assumes that the predictors used in our model can be correctly simulated by GCM and convincingly projected to the future^{13,15,42}. However, because of the systematic bias in GCMs, there are always significant differences between GCM and observation predictors⁴³. Thus, the bias-adjustment methods should be used to reduce or eliminate the differences to fulfill the PP assumption³². As the PP assumption should be satisfied in the whole-time range, the adjustment is both applied in the historical period and future scenarios.

Baño-Medina³² provided a simple strategy to do the adjustment in both historical periods and future scenarios.

$$x^{j,t}_{BA} = x^{j,t}_{GCM} - \bar{x}^j_{GCM,i} + \bar{x}^j_{ERA,i} \quad (1)$$

Equation 1 shows the algorithm of the bias-adjustment process. j refers to a particular variable of climate data, t refers to time, and $i = 1, 2, \dots, 12$ refers to the month of the year. $x^{j,t}_{BA}$ refers to the variable j in time t after the bias-adjustment procedure; $x^{j,t}_{GCM}$ refers to the variable j of climate data in time t simulated by GCM; $\bar{x}^j_{GCM,i}$ refers to the average value of variable j in i th month in historical period simulated by GCM; $\bar{x}^j_{ERA,i}$ refers to the average value of variable j in i th month from ERA-5.

After the process, the PP assumption can be fulfilled, and we can transfer the model to future scenarios and do the projection. The transferred model receives GCM data after bias-adjustment and DEM data then yields the downscaled LAI data. Data were all interpolated from 0.0625° to 0.05° to ensure ease of use. We also provided some examples of data applications such as detecting future vegetation degradation risk in local areas.

Data Records

This F0.05D-LAI dataset is available under the Open Science Framework (<https://doi.org/10.17605/OSF.IO/9QZ4K>), containing 5 data records in historical (1983–2014), SSP-126, SSP-245, SSP-370 and SSP-585 (2015–2100)⁴⁴. The data is multiplied by 100 and is stored in int32 data type and GeoTiff file format with the WGS 1984 coordinate system. The temporal resolution is 1 month, and the spatial resolution is 0.05° .

All data were named ‘lai_<experiment>_<year>_<month>.tif’. The ‘experiment’ field can be one of ‘historical’, ‘ssp126’, ‘ssp245’, ‘ssp370’ and ‘ssp585’, representing the scenario of the data. The ‘year’ field and ‘month’ field represent the time of the data.

Please note that the LAI data presented here is potential LAI data only. Due to the complexity and unpredictability of human activities in the future, the actual LAI in the future may differ significantly from predictions.

Technical Validation

Validation in historical period. In the “LAIDN Model Training” step, we calibrated our model after data collection and preprocessing. The metrics during the training process can be seen in Supplementary Figure 1. Since the input of the model was bias-adjusted GCM (BA-GCM) data, we downscaled in historical period using BA-GCM data and calculated their R^2 and RMSE to show that the model would keep convincing after changing its input. The metrics of downscaled LAI by different methods are shown in Table 1.

Method	Training		Validation		Test	
	R ²	RMSE	R ²	RMSE	R ²	RMSE
LAI DN	0.879	0.347	0.874	0.354	0.887	0.340
GLM	0.387	0.781	0.393	0.768	0.395	0.779
SRDSN	0.675	0.594	0.666	0.594	0.675	0.593

Table 1. Downscaled result metrics. The result is calculated using the BA-GCM data as input.

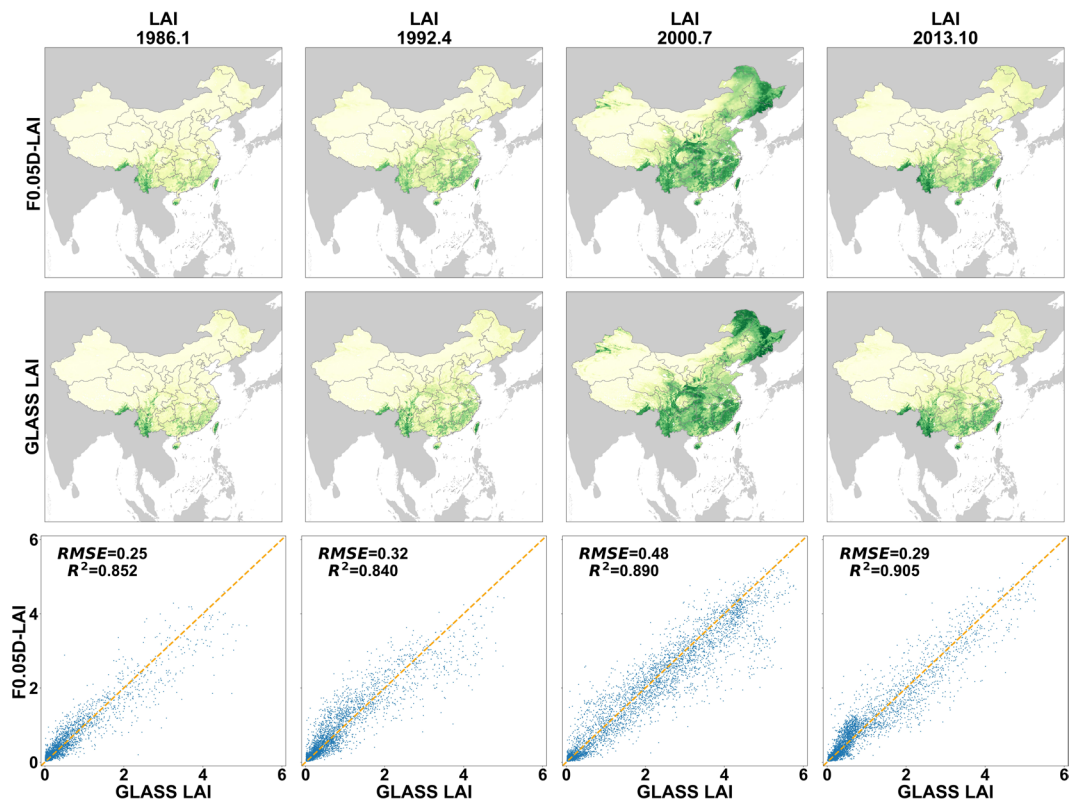


Fig. 3 Sample results of downscaling. All data are selected from the test set. The first row is the F0.05D-LAI data, while the second row is the GLASS LAI. The third row is the scatter plots which show the accuracy of downscaling. Each column is a representation of a season.

Table 1 shows that the LAI downscaled by LAIDN has the highest R² and lowest RMSE in general. Besides, despite changing the input of the LAIDN model from ERA-5 to BA-GCM, it kept a rather accurate downscaling performance. Therefore, we believe that our model has good downscaling ability.

The spatial distribution of downscaled LAI is illustrated in Fig. 3.

In general, our model has good downscaling performance. Regarding the value, the scatter plots in the bottom line also show that our model has convincing results in different seasons. Points are distributed around the 1:1 line, and metrics show the results are accurate. Regarding the spatial structure, the downscaled result closely resembles the GLASS LAI. The texture of the mountains in southwest China, the edge of the Sichuan basin in central China, and the high vegetation coverage in the northeast mountains are shown clearly in our result. Regarding the seasonal changes, the downscaled LAI shows the same seasonal vegetation changes, with the highest LAI in summer and the lowest LAI in winter.

To show the advantage of our model in expressing spatial details, we compared our model with benchmarks in a higher resolution. The comparison is in Fig. 4.

Figure 4 picks two sample areas in different regions of China. Among the three downscaled LAI, F0.05D-LAI shows higher accuracy, with clearer mountain and basin edges. Although some tiny vegetation texture still disappears, vegetation texture on medium-size mountains is still preserved correctly. GLM only produces a normal result, without intricate texture and edges. Another benchmark, SRDSN LAI is better than GLM LAI, but since SRDSN is not designed for downscaling LAI, its spatial pattern is not as clear as that of F0.05D-LAI. The KDE curves in the bottom row also indicate the advantages of our model. There is less discrepancy between F0.05D-LAI, SRDSN LAI and GLASS LAI while the difference is more significant in GLM LAI results, which means deep learning methods are good at projecting the LAI value. However, considering the SSIM curve, the peak of F0.05D-LAI reaches about 18, while SRDSN's reaches about 12, implying that F0.05D-LAI has more

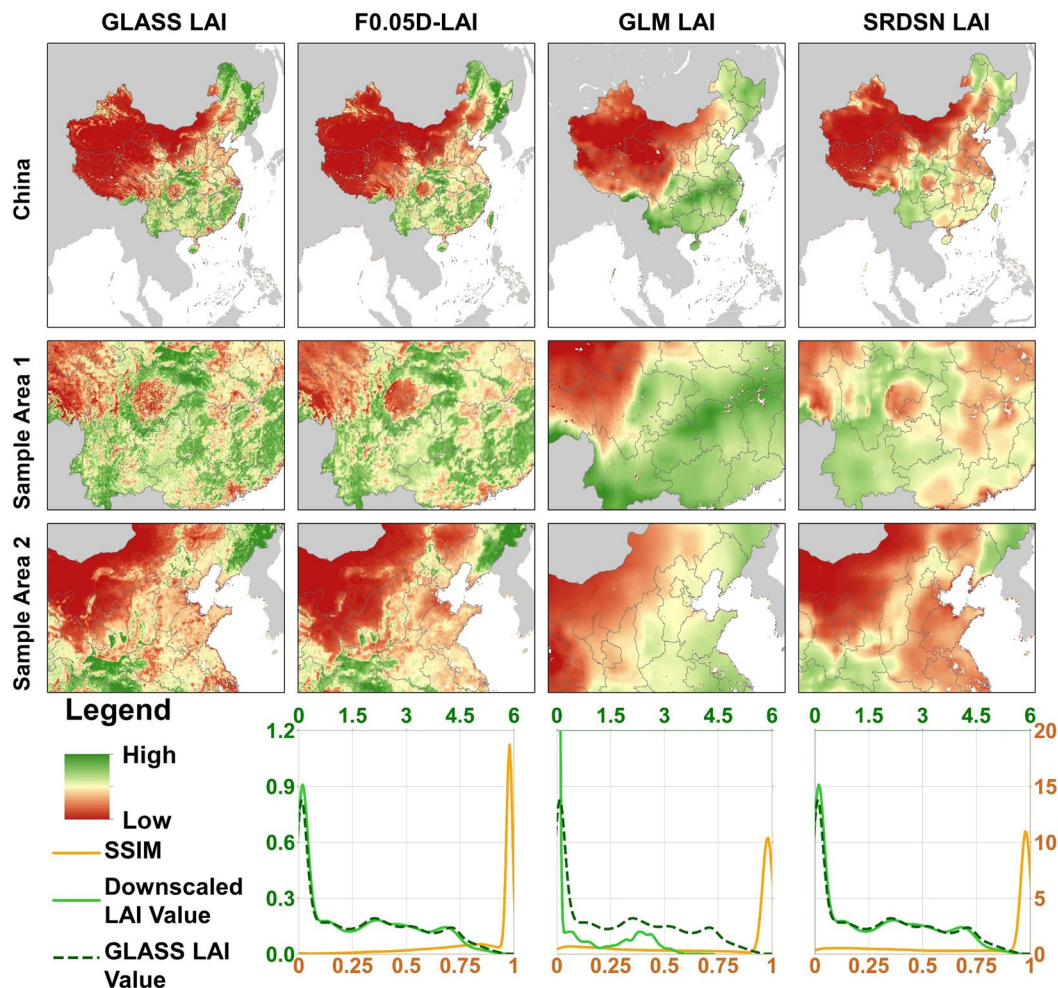


Fig. 4 Comparison between different downscaling methods. All data are selected from July 2014. The first row is the LAI in China. The second row (Sample Area 1) is the LAI in southwest China, and the third row (Sample Area 2) is in north China. The columns are GLASS LAI, F0.05D-LAI, GLM downscaled LAI and SRDSN downscaled LAI. The bottom row is the quantitative comparison between each downscaling method. The light green solid lines are the KDE curve of the downscaled LAI, and the dark green dashed lines are the KDE curve of the GLASS LAI. The axes of those two green lines are at the left and top. The orange lines are the KDE curve of the SSIM values whose axes are at the right and bottom.

regions with high spatial similarity. Therefore, we can further conclude that our LAIDN model shows a better performance in downscaling LAI, and F0.05D-LAI is a rather accurate LAI dataset.

Since time series analysis is an important usage of LAI⁴⁵, we depicted the time series curve of five kinds of LAI datasets in six locations in Fig. 5.

To clearly show the accuracy of the time series, we calculated two metrics to evaluate the similarity between F0.05D-LAI, SRDSN LAI, GCM LAI and GLASS LAI in Table 2. Here we use two widely-used time-series similarity measurements: Dynamic Time Warping (DTW)⁴⁶ and Euclidean Distance⁴⁷.

According to Fig. 5, in most regions, F0.05D-LAI has a rather high similarity to the GLASS LAI, especially in the Tibet Plateau, Northeast China, North China Plain, and the Loess Plateau. In South China and Mt. Hengduan which are filled with mountains and hills with complex vegetation patterns, there is a larger discrepancy between downscaled LAI and GLASS LAI, but our model can still produce a better result. Metrics in Table 2 show a similar conclusion: regardless of what metric has been used, F0.05D-LAI shows a higher similarity in time series comparison. Therefore, LAIDN can be concluded as a rather appropriate method for LAI downscaling, and F0.05D-LAI can be considered as a relatively accurate dataset for future LAI.

Comparison in future scenarios. After the downscaling method has been validated in the historical period, we can transfer the model and produce downscaled LAI data in future scenarios. The general trends of vegetation in future scenarios are shown in Fig. 6. To show the dispersion of LAI each year, we painted the shaded area within the range of mean value \pm standard error⁴⁸ of monthly averaged LAI.

From Fig. 6, the vegetation in China shows an increasing trend in all scenarios, and the greening phenomenon is more significant in higher emission scenarios. In SSP-126 which is the lowest emission scenario, the LAI in China shows a slightly increasing trend of about 3.7×10^{-4} per year. In a higher emission scenario, SSP-245,

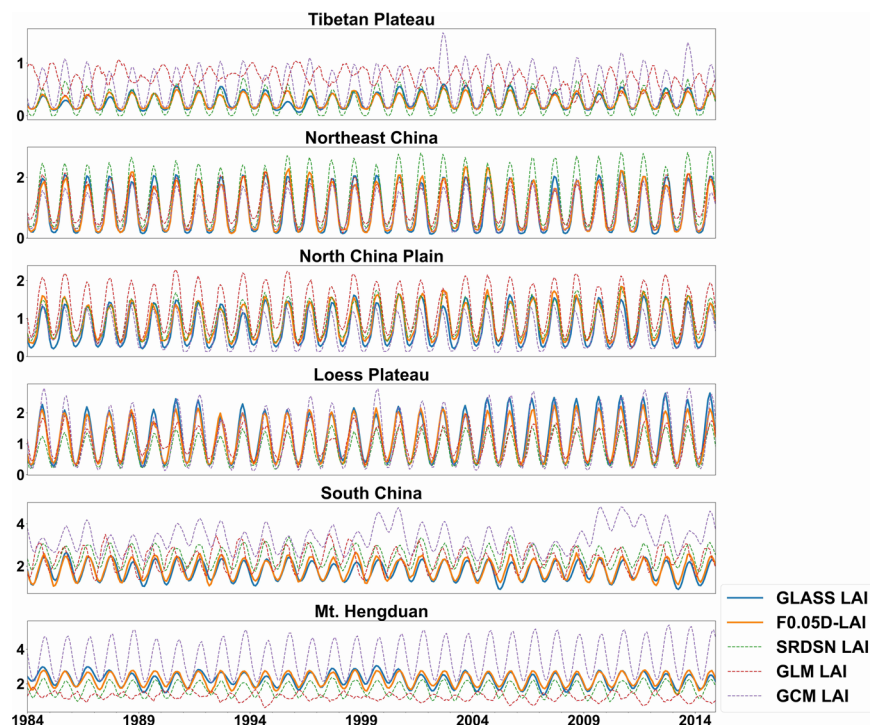


Fig. 5 Time series curve of GCM LAI, GLASS LAI and different downscaled LAI in different regions of China. The blue, bold line is the GLASS LAI, the orange bold lines are F0.05D-LAI, and the other dotted light lines are the SRDSN LAI, GLM LAI and GCM LAI.

Region	DTW				Euclidean Distance			
	F0.05D-LAI	SRDSN LAI	GLM LAI	GCM LAI	F0.05D-LAI	SRDSN LAI	GLM LAI	GCM LAI
Tibetan Plateau	0.781	1.726	5.218	5.576	0.957	1.924	9.91	6.603
Northeast China	2.849	5.596	4.255	3.73	3.5	7.971	5.642	4.532
North China Plain	2.631	2.591	6.369	2.628	3.739	3.911	9.408	5.451
Loess Plateau	2.697	7.032	5.766	3.575	3.266	9.523	7.789	6.539
South China	2.28	9.819	8.321	22.745	3.919	13.749	15.532	30.547
Mt. Hengduan	3.017	7.138	12.785	21.966	4.19	10.579	21.636	28.25

Table 2. Time series similarity between F0.05D-LAI/SRDSN LAI/GCM LAI and GLASS LAI.

the slope of average LAI is about 8.1×10^{-4} per year, much higher than SSP-126. High emission scenarios, SSP-370 and SSP-585 show the most significant greening trend around 15×10^{-4} per year, leading the average LAI in China to increase from about 0.85 to 1.0.

To fully understand the spatial and temporal distribution of greenness among scenarios, we painted the LAI increasing value in two periods: 2025 to 2050 (25 years) and 2050 to 2100 (50 years). Please refer to Supplementary Figures 2, 3 for more information.

In Fig. 6, there is a huge gap between GCM LAI and F0.05D-LAI. According to some related research, the future greening trend of the Earth may not be as large as previously thought⁴⁹, which means some systematic bias might exist in LAI projections. Similar overestimation may also exist in future scenarios according to the comparison in Fig. 7, and our data may be closer to the real situation in the future and have a rather high accuracy.

The subfigure (a.1) in Fig. 7 shows that the F0.05D-LAI value is much closer to the GLASS LAI value. Since our result has a higher accuracy in the historical period, it is reasonable to infer that our data can depict the vegetation more precisely in future scenarios.

The subfigure (a.2) of Fig. 7 shows the overestimation areas which are distributed mainly in the Hengduan Mountains, Sichuan Basin, Loess Plateau, and southeast China. The three sample areas are selected from these overestimation areas. Sample Area (b.1) located in Loess Plateau, with LAI around 1. F0.05D-LAI shows that the vegetation will remain stable until 2050, but GCM LAI indicates that the average LAI will increase to about 2.5 by 2050, which is equal to the vegetation situation in the sample (b.2) located in the Daba Mountains in Hubei Province, central south China. Vegetation in (b.1) is the temperate shrublands, growing *Vitex negundo* var. *heterophylla* and *Zizyphus jujuba* var. *spinosa*, while vegetation in (b.2) is the subtropical forest and shrublands growing *Quercus fabri* Hance and *Quercus glandulifera* var. *brevipetiolata* Nakai⁵⁰. Similarly, sample area (c.1)

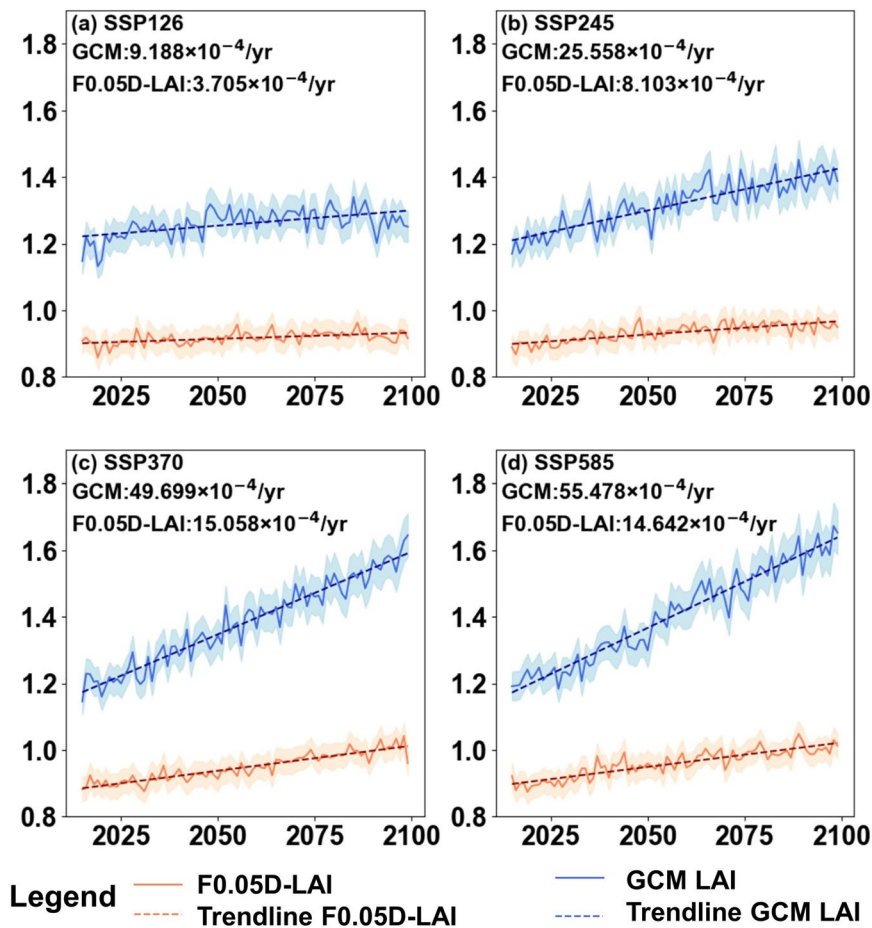


Fig. 6 The trend of average GCM LAI and F0.05D-LAI in China. The scenario and increasing rate of each chart is marked in the upper left of it. The shaded area depicts the mean value \pm standard error of the monthly averaged LAI in each year.

which is farmland growing wheat and rice will change to a subtropical Needleleaf Forests growing *Pinus masoniana Lamb* like Wuyi Mountain in (c.2)⁵⁰, and sample area (d.1) which is subalpine grassland will change to a rainforest like Sumatra in (d.2). Therefore, we infer that F0.05D-LAI is a rather precise LAI projection data in future scenarios, which shows a rather slightly greening trend.

In another aspect, take the sample area (d.1) as an example. The yearly average temperature of (d.1) is about 10°C, while the yearly average temperature of (d.2) is about 25°C; the precipitation of (d.1) is about 800 mm per year, while the precipitation in (d.2) is more than 2500 mm up to 6000 mm per year⁵¹. However, the precipitation will only increase about 15–20% in the (d.1) area, and the temperature will only increase about 3–4°C until the 2080s⁵². Likewise, the precipitation in the (b.1) area will increase from about 400 mm to 600 mm per year in the 2080s⁵², but there is still a large gap to the 1000 mm precipitation in (b.2).

Currently, many studies indicate that GCMs, which are designed for global-scale simulations, face challenges in accurately predicting future LAI at finer spatial resolutions^{49,53}. In the northern hemisphere, some models tend to overestimate the annual mean LAI and the rate of increase in LAI, which may reflect systematic biases due to the models' global focus⁵³. This overestimation aligns with research suggesting that vegetation growth may be slower than previously modeled⁴⁹. While GCMs excel in capturing large-scale climatic patterns⁹, they show greater variability and uncertainty when applied to fine-scale vegetation dynamics⁴⁹. Therefore, our dataset can help to improve the accuracy of climate models at regional scales, bridging the gap between GCMs and local environmental needs.

Limitation and future work. This study downscaled the LAI data in future scenarios, providing data that can support further research in the mitigation and adaptation of climate change. We provided high-resolution and high-accuracy F0.05D-LAI dataset in both historical and future scenarios, and we demonstrated that our result is better than other existing downscaling methods. However, there are still some limitations in our data. In Fig. 4, although our model depicts a better spatial distribution of vegetation, some details in hills and basins still disappear. Figure 5 also indicates that projection from our model shows a relatively low accuracy in mountains in southern China and Mountain Hengduan. Besides, city-level data is still not very convincing after downscaling. Metropolises in Fig. 4 Sample Area 1, Chongqing and Chengdu show a relatively indeterminate vegetation structure in the downscaled result because smaller vegetation such as green lands and parks are not shown. Therefore, using our data to

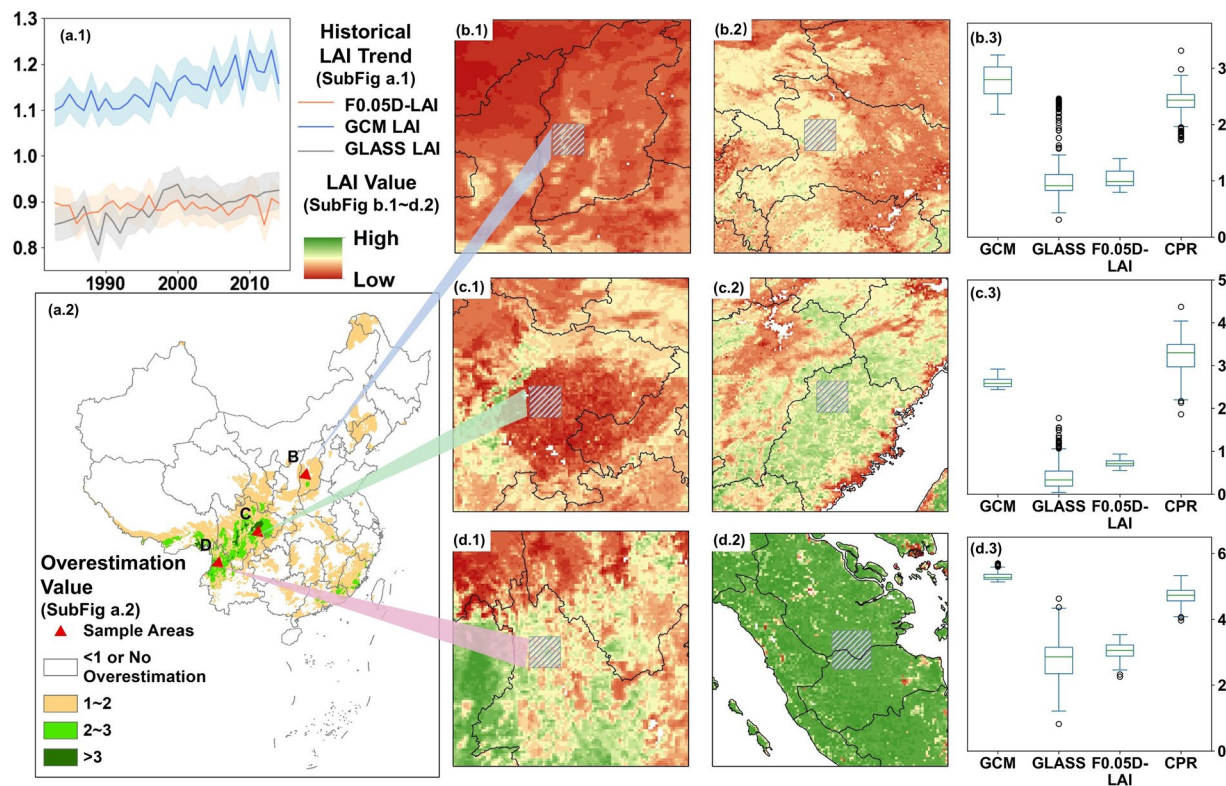


Fig. 7 Discussion of overestimation in GCM LAI. The GCM LAI and F0.05D-LAI are the average data in 2050 of SSP-370. The GLASS LAI data are the average data in 2014. The shaded area in chart a.1 depicts the mean value \pm standard error of the monthly averaged LAI in each year. Charts in the left column (b.1, c.1, d.1) are three sample areas in China with high overestimation. Charts on the center column (b.2, c.2, d.2) are other areas whose LAI is close to the GCM LAI of the left column. Charts on the right column (b.3, c.3, d.3) are the box plots of different sources of LAI. Column GCM is the box plot of GCM LAI in b.1, c.1 and d.1 in 2050. Column GLASS is the box plot of GLASS LAI in b.1, c.1 and d.1 in 2014. Column F0.05D-LAI is the box plot of F0.05D-LAI in b.1, c.1 and d.1 in 2050. Column CPR is the box plot of GLASS LAI in b.2, c.2, d.2 in 2014.

conduct inner-city research is still difficult. Besides, F0.05D-LAI data are only potential LAI, which are only based on four SSPs, and there are still many situations that cannot be covered by these scenarios. The complexity and unpredictability of future human activities may result in actual LAI being significantly different from predictions.

To minimize the limitations of the data, there are a few directions for future research that are worth considering. Firstly, using more variables with higher resolution may be useful to improve the final resolution. For example, we may consider using RCMs instead of GCMs to have a higher initial resolution and add other variables like wind speed or radiation to provide more information. Secondly, since China is a large country and the climate in different regions is controlled by different factors, we may split China into parts and do downscaling separately. For example, east China is mainly controlled by the monsoon, west-north China has a primarily continental climate, while the Qinghai-Tibet Plateau is influenced by its elevation. Splitting China into regions may allow the model to fit the regional climate with a higher accuracy. Besides, as different models have different advantages, comparison between models based on the PP method can provide more useful data and even correct some errors on some specific models. In addition, LAIDN might be generalizable to downscale other vegetation parameters. We have done a simple experimental downscaling to Gross Primary Productivity (GPP) and Net Primary Productivity (NPP) whose result is shown in Supplementary Figures 5, 6. We will conduct further research on downscaling vegetation parameters from different simulations and analyze future vegetation changes more accurately in the future.

Usage Notes

To show an application of our data, we evaluated the vegetation degradation risk in future scenarios in the Chengdu-Chongqing City Group. To distinguish the diverse levels of vegetation risk, we classify risk areas into three categories: High-Risk, Medium-Risk, and Low-Risk based on the significance level and decreasing range as factors. The classification method is shown in Table 3. We use the Mann-Kendall test to detect possible trends and use linear regression to calculate the slope. The decreasing range is calculated based on the average LAI in 2025–2035 and 2091–2100.

Figure 8 shows a resolution comparison between F0.05D-LAI and GCM LAI, highlighting the significance of downscaling. For example, some small risk areas in SSP-126 located in the southwest of the city group may not be detected through GCM data, since they are much smaller than a single pixel of GCM. The large risk

LAI decreasing range	Decreasing trend significance level		
	A significant decreasing trend on $p = 0.05$ level	A significant decreasing trend only on $p = 0.1$ level	No significant decreasing trend
>5%	High	Medium	Low
3%–5%	Medium	Medium	Low
<3%	Low	Low	No

Table 3. Vegetation degradation risk classification.

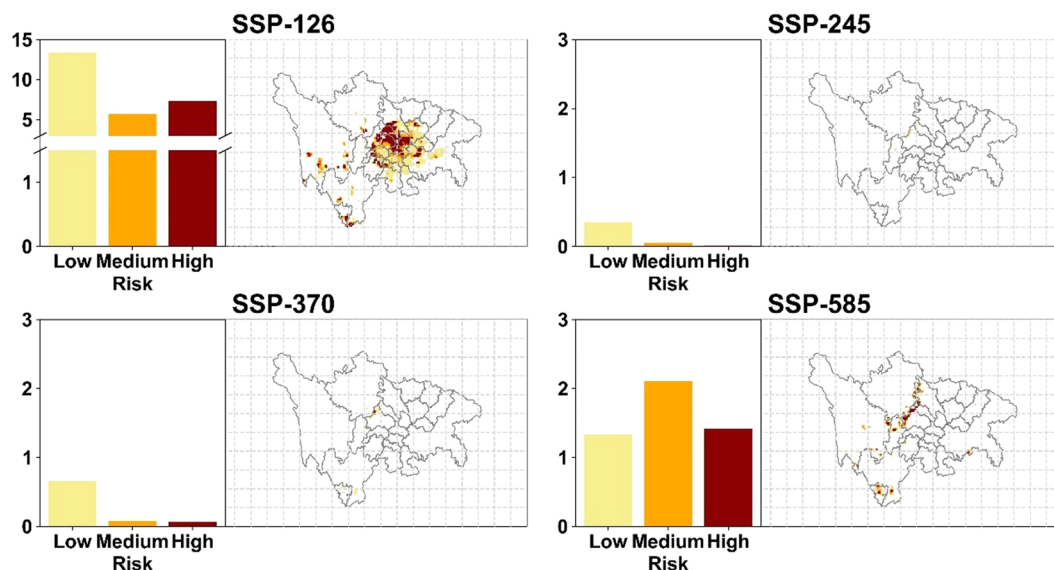


Fig. 8 Vegetation risk area distribution in Chengdu-Chongqing city group from 2025 to 2100. The figure indicates the vegetation degradation risk in Chengdu-Chongqing City Group from 2025 to 2100. The meaning of each color in the map is the same as that in the bar charts. The grid in the figure is a latitude/longitude grid with 1° intervals, which has the same size as GCM pixels.

area in SSP-126 has clear edges in F0.05D-LAI and different risk levels are shown. High-risk areas are at the Chengdu–Chongqing metropolis and the alluvial plain of Qingyi River, while low-risk areas are located at some mountains and hills. Another high-risk area under the SSP-585 scenario is also correctly detected. This area is on the edge of a very undulating mountain range, ecologically fragile, and shows a high vegetation risk under a high emission scenario with fast climate change. Therefore, F0.05D-LAI can detect smaller vegetation risk zones and may be helpful for urban planning and environmental assessment.

The vegetation degradation risk detection method is also available in the whole of China, and we also provide a risk assessment map. Please refer to Supplementary Figure 4 for more information.

Code availability

The code that calibrates the model and predicts the future LAI is based on Python 3.9.12, working with the key libraries *numpy*, *gdal*, and *Pytorch*. The code can be found on the GitHub repository (https://github.com/NoSleepTonight-0712/LAI_downscaling_upload).

Received: 9 May 2024; Accepted: 30 September 2024;

Published online: 11 October 2024

References

- Chen, J. M. & Black, T. A. Defining leaf area index for non-flat leaves. *Plant, Cell & Environment* **15**, 421–429 (1992).
- Cao, S. *et al.* Spatiotemporally consistent global dataset of the GIMMS leaf area index (GIMMS LAI4g) from 1982 to 2020. *Earth System Science Data* **15**, 4877–4899 (2023).
- Zhu, Z. *et al.* Greening of the Earth and its drivers. *Nature Clim Change* **6**, 791–795 (2016).
- Boussetta, S., Balsamo, G., Dutra, E., Beljaars, A. & Albergel, C. Assimilation of surface albedo and vegetation states from satellite observations and their impact on numerical weather prediction. *Remote Sensing of Environment* **163**, 111–126 (2015).
- Zhuo, W. *et al.* Crop yield prediction using MODIS LAI, TIGGE weather forecasts and WOFOST model: A case study for winter wheat in Hebei, China during 2009–2013. *International Journal of Applied Earth Observation and Geoinformation* **106**, 102668 (2022).
- Tian, A. *et al.* Water yield variation with elevation, tree age and density of larch plantation in the Liupan Mountains of the Loess Plateau and its forest management implications. *Science of The Total Environment* **752**, 141752 (2021).
- Fang, H., Baret, F., Plummer, S. & Schaepman-Strub, G. An Overview of Global Leaf Area Index (LAI): Methods, Products, Validation, and Applications. *Reviews of Geophysics* **57**, 739–799 (2019).

8. Eyring, V. *et al.* Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development* **9**, 1937–1958 (2016).
9. Flato, G. *et al.* Evaluation of climate models. in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* 741–866. <https://doi.org/10.1017/CBO9781107415324.020> (Cambridge University Press, 2014).
10. Wu, T. *et al.* The Beijing Climate Center Climate System Model (BCC-CSM): the main progress from CMIP5 to CMIP6. *Geoscientific Model Development* **12**, 1573–1600 (2019).
11. Mulcahy, J. P. *et al.* UKESM1.1: development and evaluation of an updated configuration of the UK Earth System Model. *Geoscientific Model Development* **16**, 1569–1600 (2023).
12. Haarsma, R. J. *et al.* High Resolution Model Intercomparison Project (HighResMIP v1.0) for CMIP6. *Geoscientific Model Development* **9**, 4185–4208 (2016).
13. Maraun, D. & Widmann, M. *Statistical Downscaling and Bias Correction for Climate Research*. (Cambridge University Press, 2018).
14. Gebrechorkos, S. *et al.* A high-resolution daily global dataset of statistically downscaled CMIP6 models for climate impact analyses. *Sci Data* **10**, 611 (2023).
15. Baño-Medina, J. *et al.* Downscaling multi-model climate projection ensembles with deep learning (DeepESD): contribution to CORDEX EUR-44. *Geoscientific Model Development* **15**, 6747–6758 (2022).
16. Wootten, A. M., Dixon, K. W., Adams-Smith, D. J. & McPherson, R. A. Statistically downscaled precipitation sensitivity to gridded observation data and downscaling technique. *International Journal of Climatology* **41**, 980–1001 (2021).
17. Wang, F., Tian, D., Lowe, L., Kalin, L. & Lehrter, J. Deep Learning for Daily Precipitation and Temperature Downscaling. *Water Resources Research* **57**, 3–7 (2021).
18. Wang, F. & Tian, D. On deep learning-based bias correction and downscaling of multiple climate models simulations. *Clim Dyn* **59**, 3451–3468 (2022).
19. Lim, B., Son, S., Kim, H., Nah, S. & Mu Lee, K. Enhanced Deep Residual Networks for Single Image Super-Resolution. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* 136–144 (2017).
20. Ledig, C. *et al.* Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 4681–4690 (2017).
21. Baño-Medina, J., Manzanar, R. & Gutiérrez, J. M. Configuration and intercomparison of deep learning neural models for statistical downscaling. *Geoscientific Model Development* **13**, 2109–2124 (2020).
22. Tian, F. *et al.* Satellite-observed increasing coupling between vegetation productivity and greenness in the semiarid Loess Plateau of China is not captured by process-based models. *Science of The Total Environment* **906**, 167664 (2024).
23. Liu, Y., Liu, R. & Chen, J. M. Retrospective retrieval of long-term consistent global leaf area index (1981–2011) from combined AVHRR and MODIS data. *Journal of Geophysical Research: Biogeosciences* **117**, (2012).
24. You, C. *et al.* Inner Mongolia grasslands act as a weak regional carbon sink: A new estimation based on upscaling eddy covariance observations. *Agricultural and Forest Meteorology* **342**, 109719 (2023).
25. Lawal, S. *et al.* Investigating the response of leaf area index to droughts in southern African vegetation using observations and model simulations. *Hydrology and Earth System Sciences* **26**, 2045–2071 (2022).
26. Shi, R., Wang, T., Yang, D. & Yang, Y. Streamflow decline threatens water security in the upper Yangtze river. *Journal of Hydrology* **606**, 127448 (2022).
27. Badach, J., Dymnicka, M. & Baranowski, A. Urban Vegetation in Air Quality Management: A Review and Policy Framework. *Sustainability* **12**, 1258 (2020).
28. Hu, Y. *et al.* LAI-indicated vegetation dynamic in ecologically fragile region: A case study in the Three-North Shelter Forest program region of China. *Ecological Indicators* **120**, 106932 (2021).
29. Copernicus Climate Change Service. ERA5 monthly averaged data on pressure levels from 1979 to present. ECMWF <https://doi.org/10.24381/CDS.6860A573> (2019).
30. Riahi, K. *et al.* The Shared Socioeconomic Pathways and their energy, land use, and greenhouse gas emissions implications: An overview. *Global Environmental Change* **42**, 153–168 (2017).
31. O'Neill, B. C. *et al.* The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6. *Geoscientific Model Development* **9**, 3461–3482 (2016).
32. Baño-Medina, J., Manzanar, R. & Gutiérrez, J. M. On the suitability of deep convolutional neural networks for continental-wide downscaling of climate change projections. *Clim Dyn* **57**, 2941–2951 (2021).
33. Liang, S. *et al.* The Global Land Surface Satellite (GLASS) Product Suite. *Bulletin of the American Meteorological Society* **102**, E323–E337 (2021).
34. Tadono, T. *et al.* Precise global DEM generation by ALOS PRISM. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* **2**, 71–76 (2014).
35. Florinsky, I. V., Skrypitsyna, T. N. & Luschikova, O. S. Comparative accuracy of the AW3D30 DSM, ASTER GDEM, and SRTM1 DEM: A case study on the Zaoksky testing ground, Central European Russia. *Remote Sensing Letters* **9**, 706–714 (2018).
36. He, K., Zhang, X., Ren, S. & Sun, J. Deep Residual Learning for Image Recognition. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 770–778 (2016).
37. Shi, W. *et al.* Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 1874–1883 (2016).
38. He, K., Zhang, X., Ren, S. & Sun, J. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. in *2015 IEEE International Conference on Computer Vision (ICCV)* 1026–1034. <https://doi.org/10.1109/ICCV.2015.123> (Santiago, Chile, 2015).
39. Kingma, D. P. & Ba, J. Adam: A Method for Stochastic Optimization. in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7–9, 2015, Conference Track Proceedings* (eds. Bengio, Y. & LeCun, Y.) (2015).
40. Wang, Z. & Bovik, A. C. Mean squared error: Love it or leave it? A new look at Signal Fidelity Measures. *IEEE Signal Processing Magazine* **26**, 98–117 (2009).
41. Gutiérrez, J. M. *et al.* An intercomparison of a large ensemble of statistical downscaling methods over Europe: Results from the VALUE perfect predictor cross-validation experiment. *International Journal of Climatology* **39**, 3750–3785 (2019).
42. Maraun, D., Widmann, M. & Gutiérrez, J. M. Statistical downscaling skill under present climate conditions: A synthesis of the VALUE perfect predictor experiment. *International Journal of Climatology* **39**, 3692–3703 (2019).
43. Ivanov, M. A. & Kotlarski, S. Assessing distribution-based climate model bias correction methods over an alpine domain: added value and limitations. *International Journal of Climatology* **37**, 2633–2653 (2017).
44. Li, H. A dataset of 0.05-degree LAI in China during 2015–2100 based on deep learning network. <https://doi.org/10.17605/OSF.IO/9QZ4K> (2024).
45. Park, H. & Jeong, S. Leaf area index in Earth system models: how the key variable of vegetation seasonality works in climate projections. *Environ. Res. Lett.* **16**, 034027 (2021).
46. Lhermitte, S., Verbesselt, J., Verstraeten, W. W. & Coppin, P. A comparison of time series similarity measures for classification and change detection of ecosystem dynamics. *Remote Sensing of Environment* **115**, 3129–3152 (2011).
47. Wang, X. *et al.* Experimental comparison of representation methods and distance measures for time series data. *Data Min Knowl Disc* **26**, 275–309 (2013).

48. Altman, D. G. & Bland, J. M. Standard deviations and standard errors. *BMJ* **331**, 903 (2005).
49. Zhao, Q., Zhu, Z., Zeng, H., Zhao, W. & Myneni, R. B. Future greening of the Earth may not be as large as previously predicted. *Agricultural and Forest Meteorology* **292–293**, 108111 (2020).
50. Zhang, X., Sun, S., Yong, S., Zhou, Z. & Wang, R. Vegetation map of the People's Republic of China (1: 1000000). *Geol. Publ. House* (2007).
51. Beck, H. E. *et al.* Present and future Köppen–Geiger climate classification maps at 1-km resolution. *Sci Data* **5**, 180214 (2018).
52. Yang, X., Zhou, B., Xu, Y. & Han, Z. CMIP6 Evaluation and Projection of Temperature and Precipitation over China. *Adv. Atmos. Sci.* **38**, 817–830 (2021).
53. Anav, A. *et al.* Evaluating the Land and Ocean Components of the Global Carbon Cycle in the CMIP5 Earth System Models. *Journal of Climate* **26**, 6801–6843 (2013).

Acknowledgements

This work was supported by the Third Xinjiang Scientific Expedition Program (Grant No. 2022xjkk0405) and the National Natural Science Foundation of China (No. 42090012). We thank the Open Science Framework (<https://osf.io/>) for publishing our dataset. We appreciate the support of ESGF, Copernicus and ALOS for providing climate simulations, ERA-5 and DEM data. We acknowledge the GLASS data support from “National Earth System Science Data Center, National Science & Technology Infrastructure of China (<http://www.geodata.cn>)”. We would like to thank the editors and two anonymous reviewers for their constructive comments which helped to improve the manuscript.

Author contributions

Hao Li: Methodology, Programming, Data Analysis, Writing the original draft, Manuscript Modification. Xiang Zhao: Conceptualization, Methodology, Funding Acquisition, Manuscript Modification. Yuyu Zhou, Xin Zhang and Shunlin Liang: Manuscript Modification. All authors have contributed to reviewing the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41597-024-03948-z>.

Correspondence and requests for materials should be addressed to Xiang Zhao.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024