Deep-learning-based harmonization and super-resolution of near-surface air temperature from CMIP6 models (1850–2100)

Xikun Wei | Guojie Wang | Donghan Feng | Zheng Duan
Daniel Fiifi Tawia Hagan | Liangliang Tao | Lijuan Miao | Buda Su | Tong Jiang

1Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters, Nanjing University of Information Science and Technology (NUIST), Nanjing, China
2Department of Physical Geography and Ecosystem Science, Lund University, Lund, Sweden
3Geographical Sciences College, Nanjing University of Information Science and Technology (NUIST), Nanjing, China

Abstract

Future global temperature change will have significant effects on society and ecosystems. Earth system models (ESM) are the primary tools to explore future climate change. However, ESMs have great uncertainty and often run at a coarse spatial resolution (usually about 2°). Accurate high-spatial-resolution temperature dataset are needed to improve our understanding of temperature variations and for many other applications. We apply Super resolution (SR) in computer vision using deep learning (DL) to merge 31 ESMs data. The proposed method performs data merging, bias-correction, and spatial downscaling simultaneously. The CRU TS (Climate Research Unit gridded Time Series) data is used as reference data in the model training process. To find a suitable DL method, we select five SR methodologies with different structures. We compare the performances of the methods based on mean square error (MSE), mean absolute error (MAE) and Pearson correlation coefficient ($R$). The best method is used to merge the projected monthly data (1850–1900), and monthly future scenarios data (2015–2100), at the high spatial resolution of 0.5°. Results show that the merged data have considerably improved performance compared with individual ESM data and their ensemble mean (EM), both spatially and temporally. The MAE shows significant improvement; the spatial distribution of the MAE widens along the latitudes in the Northern Hemisphere. The MAE of merged data is ranging from 0.60 to 1.50, the South American (SA) has the lowest error and the Europe has the highest error. The merged product has excellent performance when the observation data is smooth with few fluctuations in the time series. This work demonstrates the applicability and effectiveness of the DL methods in data merging, bias-correction and spatial...
downscaling when enough training data are available. Data can be accessed at https://doi.org/10.5281/zenodo.5746632.

KEYWORDS
climate change, CMIP6, data-merging, deep-learning, temperature

1 | INTRODUCTION

From the Sixth Assessment Report by the Intergovernmental Panel on Climate Change (IPCC), global air temperature increased by 1.09°C from 2011 to 2020 compared to 1850–1900 (Masson-Delmotte et al., 2021). Global warming has significant impact on the natural environment, mankind, the water cycle (Yamaguchi et al., 2020), food safety (Hasegawa et al., 2021), extreme disaster (Quante et al., 2021; Touma et al., 2021), socioeconomic (Naumann et al., 2021), diversity of species (Avaria-Llaurore et al., 2021) and other fields. Besides, the most direct influence is the decline of ice cover in the ocean. In past decades, the South Pole experienced an unprecedented warming speed (three times more than the whole earth) (Clem et al., 2020) and ice cover in Arctic gone through a decrease of average 12% every decade (Zhukov and Gushcha, 2020). The loss of mass ice will also bring an extra increase of temperature at a long time scale (Wunderling et al., 2020). If temperature have an extra increase 2°C will result in an 20 cm rise of the global ocean (Jevrejeva et al., 2016). In this background, the society and scientific research groups pay a lot of attention to the global temperature changes (Choi et al., 2020). Hence, in order to improve our understanding of the spatial and temporal changes of future temperature, a more precise future temperature dataset at high spatial resolution is needed.

The traditional meteorological stations can provide long time series point data. However, with the development of remote sensing technology, the number of satellites specially used for meteorological observation are increasing; this provides a large number of meteorological satellite observation data. The satellite observation data have relatively shorter time series which are not useful for future climate change prediction. The climate model and Earth system models (ESMs) that consider comprehensive physical and biological processes act as major tools for projecting future climate variations. The projected data of ESMs provide avenues to investigate climate change and the potential influence of those changes in global and regional contexts (Zhukov and Gushcha, 2020; Yuan et al., 2021). To address the increasing range of scientific questions posed by several research communities, the coupled model inter-comparison project (CMIP) team continues revising the experimental design and organization of CMIP (Eyring et al., 2016). The latest is the Phase6 CMIP (CMIP6) that has a substantial augmentation of model numbers and range of experiments. Compared with the previous Phase5 CMIP (CMIP5), ESMs of CMIP6 have spatial resolution improvements, making progress in physical parameterizations; it also includes extra Earth system processes and components (Eyring et al., 2019). The response of the Earth system to variations in radiative forcings, as well as humanity’s response through technology, lifestyle, economies and policy have implications for the environment and society through climate change. (Moss et al., 2010). Shared socioeconomic pathways (SSPs) (van Vuuren et al., 2014; Riahi et al., 2017) are new scenarios based on the anticipated challenges to adaptation and mitigation applied in CMIP6 (O’Neill et al., 2016). The SSPs include five alternative pathways for socioeconomic development, including sustainable development (SSP1) (Riahi et al., 2017), middle-of-the-road development (SSP2) (Fricko et al., 2017), regional rivalry (SSP3) (Fujimori et al., 2017), inequality (SSP4) (Calvin et al., 2017) and fossil-fueled development (SSP5) (Kriegler et al., 2017). These new scenarios can help to understand climate and socioeconomic futures better.

There are many studies analysing future climate changes based on the earth system model (ESM) data (Parsons et al., 2020; Gillett et al., 2021). These studies generally highlight three main problems using ESM data: (1) spatial downscaling of ESMs data to generate data at higher spatial resolutions; (2) bias reduction or correction of ESMs; (3) merging of the multi-ESMs to obtain improved accuracy. The output of ESMs have certain errors and raw low spatial resolutions due to limited computational resources, several simplified hypothesis and uncertainties in model structures and parameterizations (Phillips and Gleckler, 2006). Current ESMs are run at coarse spatial resolutions and can only provide outputs at spatial resolutions ranging from 0.5° to 3.75° (the majority are at about 2°); these spatial resolutions are not sufficient for elaborate studies (White and Toumi, 2013). Spatial downscaling methods are widely used to improve the resolution of ESMs (Shrestha et al., 2014; Baghanam et al., 2020). It can be divided into statistical and dynamical downscaling techniques (Kannan and Ghosh, 2013).
Dynamical downscaling is mainly based on the high-resolution regional climate model (RCM), using the forced lateral boundary from the global climate model (GCM) (Adachi and Tomita, 2020). Statistical downscaling builds the relationship between the ESM and the local-scale meteorological data to generate high-resolution data (Maraun and Widmann, 2018). Because the dynamical downscaling methods are complex and has high computational cost, statistical downscaling approaches are more commonly used. The most popular techniques of statistical downscaling are regression-based approaches such as the multiple and generalized linear regression models (Asong et al., 2016; Das and Akhter, 2019); they have low requirement of computational resources and realize simply. Despite the resolution enhancements, considerable bias still exists in both statistical and dynamical downscaling methods (Miao et al., 2015).

To reduce the bias, many bias correction approaches have been developed. These include statistical characteristics bias correction (Ho et al., 2012; Fang et al., 2015), probability distribution bias correction (Jakob Themßl et al., 2011) and non-stationary bias correction methods (Miao et al., 2016). The downscaling techniques get closer results to observation data through the bias correction methods (Fan et al., 2021). However, the correction and downscaling methods bring some uncertainties. Lafferty et al. (2021) found the loss of production from extreme disasters was underestimated by the bias-corrected and downscaled ESM data. In different applications of the ESM data bias correction, different bias correction methods get the best performance. For instance, Gudmundsson et al. (2012) found two nonparametric transformation approaches with better performance compared with nine other methods in precipitation bias correction in Norway. A regression technique outperforms 11 other methodologies in terms of correlation in temperature bias correction in Spain (Gutiérrez et al., 2013). Some limitations exist in current ESM bias correction approaches; the ESM mean bias correction only considers the ESM mean bias and quantile correction brings an extra bias to the spatial gradient of variables (Colette et al., 2012; Bruyère et al., 2014). Additionally, majority of bias correction methods are applied to individual model datasets, which introduces greater uncertainty in future climate projection scenarios.

Considering different ESMs have different strengths which can be potentially combined to complement each other, many efforts have been made to develop methods to merge multiple ESMs to obtain improved outputs. Due to the instability of individual ESMs, the ensemble mean (EM) method has been considered in previous works as a simple and effective way to merge ESMs (A. P. Weigel et al., 2008). The evaluation results of the CMIP6 model reveal that no single model performs best in all regions under different evaluation rules (Papalexiou et al., 2020). Because ESMs have varying resolutions and different advantages in different regions, EM method has better precision in analysing future temperature dataset both regionally and globally (Fan et al., 2020; Lovino et al., 2021; You et al., 2021; Tang et al., 2021a; 2021b). Although the EM method is easy and effective, there is still room for improvement.

With the rapid increase of “big earth data” such as earth observation and model simulation data in recent years, DL based methods play an increasingly important role in Earth Science because of their better capacity in processing big data (Reichstein et al., 2019). DL approaches have been successfully used in many fields of Earth Science (Toms et al., 2020; Yuan et al., 2020), such as data fusion and downscaling (Huang, 2020), land cover mapping (Tong et al., 2020), information reconstruction (Barth et al., 2020; Kadow et al., 2020; Tang et al., 2021a; 2021b), information prediction (Ham et al., 2019; Zheng et al., 2020) and environmental parameter retrieval (Tian et al., 2021; Zang et al., 2021). Those DL models all have the comparable result with traditional methods. For instances, Ham et al. (2019) used a deep learning model consisted of convolution network to predicate the ENSO events which has the better performance than the dynamical forecast systems. Kadow et al. (2020) used the DL model to reconstruct the missing temperature dataset and the result outperformed the principle component analysis (PCA) and Kringing interpolation methods. The DL algorithms also have great potential to improve the accuracy of ESM data (Liu et al., 2020; Feng et al., 2022).

The traditional methods need to solve downscaling, bias-correction and data merging separately in different processes, introducing extra uncertainties into future projection climate data. New DL methods provide new opportunities to address challenges with the use ESMs. Robust DL super-resolution (SR) algorithms in computer vision address these challenges simultaneously. Unlike the traditional methods, we can input various coarse resolution temperature datasets and get one high-resolution temperature dataset; these models can learn information from different ESMs and perform data merging, bias correction and data downscaling synchronously, and reduce uncertainty. An early deep learning-based approach named SRCNN proposed by Yoon et al. (2015) indicated that this method performed considerably better. Afterwards, many DL approaches have been proposed in SR applications and increasing efforts are being made to improve the performances of these algorithms (Shi et al., 2016; Seif and Androutsos, 2018; Blau et al., 2018; Chen et al., 2020). More different scenarios and historical
ESMs temperature from CMIP6 are now available. The output data from CMIP6 models will reach an estimated 30 petabytes (Stockhause and Lautenschlager, 2017). The increasing data quantity can provide enough training samples for the deep learning models.

In this work, we train five different DL networks based on the CRU TS temperature and 31 ESMs temperature over the period of 1901 to 2014. We compare the DL networks and identify the best model to merge the different land temperature under SSP1-2.6, SSP2-4.5, SSP3-7.0, and SSP5-8.5 scenarios during 1850–1900 and 2015–2100.

### TABLE 1 Summary of all ESMs used in this study

| Institution (country) | Model name | Resolution (lon × lat) | Used member |
|-----------------------|------------|------------------------|-------------|
| CSIRO (Australia)     | ACCESS-CM2 | 1.875° × 1.25°         |             |
|                       | ACCESS-ESM1-5 | 1.875° × 1.241°       |             |
| AWI (Germany)         | AWI-CM-1-1-MR | 1.875° × 1.875°       |             |
| BCC (China)           | BCC-CSM2-MR | 2.25° × 2.25°         |             |
| CAMS (China)          | CAMS-CSM1-0 | 1.125° × 1.125°       |             |
| NCAR (USA)            | CESM2-WACCM | 2.5° × 1.875°         |             |
| CMCC (Italy)          | CMCC-CM2-SR5 | 1.25° × 0.9375°       |             |
| CNRM-CERFACS (France) | CNRM-CM6-1-HR | 1.25° × 0.9375°       |             |
|                       | CNRM-CM6-1   | 1.40625° × 1.40625°   |             |
|                       | CNRM-ESM2-1  |                       |             |
| CCCMA (Canada)        | CanESM5-CanOE | 2.8125° × 2.8125°     |             |
|                       | CanESM5      |                       |             |
| EC-Earth-Consortium (EU) | EC-Earth3-Veg-LR | 1.125° × 1.125°       |             |
|                       | EC-Earth3-Veg | 0.7031° × 0.7031°     |             |
| CAS (China)           | FGOALS-f3-L  | 2.5° × 2°             |             |
|                       | FGOALS-g3    | 2° × 2.25°            |             |
| NOAA-GFDL (USA)       | GFDL-ESM4    | 2.5° × 2°             |             |
| NASA (USA)            | GISS-E2-1-G  | 2° × 2°               |             |
| INM (Russia)          | INM-CM4-8    | 2° × 1.5°             |             |
|                       | INM-CM5-0    |                       |             |
| IPSL (France)         | IPSL-CM6A-LR | 2.5° × 1.259°         |             |
| NIMS-KMA (Korea)      | KACE-1-0-G   | 1.875° × 1.25°        |             |
| UA (USA)              | MCM-UA-1-0   | 3.75° × 2.25°         |             |
| MIROC (Japan)         | MIROC-ES2L   | 2.8125° × 2.8125°     |             |
|                       | MIROC6       | 2.8125° × 0.703125°   |             |
| MPI-M (Germany)       | MPI-ESM1-2-HR | 0.9375° × 0.9375°     |             |
|                       | MPI-ESM1-2-LR | 1.875° × 1.875°       |             |
| MRI (Japan)           | MRI-ESM2-0   | 1.125° × 1.125°       |             |
| NCC (Norway)          | NorESM2-LM   | 5° × 3.75°            |             |
|                       | NorESM2-MM   | 2.5° × 1.875°         |             |
| AS-RCEC (China)       | TaiESM1      | 1.25° × 0.9375°       |             |

---

**DATA AND METHODS**

### 2.1 Model data and observation data

#### 2.1.1 CMIP6 model data

In this study, we retrieve global air temperature from 31 CMIP6 ESMs that provides both historical and future data. The most commonly used in previous studies is the first realization of each model (Norris et al., 2021); so, we also use the first realization simulations (r1i1p1f1, except where unavailable). These models are summarized in
Table 1. This data is open access from https://esgf-node.llnl.gov/projects/cmip6/ (Lu et al., 2021; Feng et al., 2022). In CMIP5 future climate simulations, the Representative Concentration Pathway (RCP), including RCP2.6, RCP4.5, RCP6.0 and RCP8.5, represent the radiative forcing which will reach about 2.6, 4.5, 6.0 and 8.5 W·m⁻² in 2100, respectively. In CMIP6, three new RCPs (RCP1.9, RCP3.4 and RCP7.0) have been proposed to meet more research needs. We design five sets of experiments: one historical simulation from 1850 to 2014 and four future projection scenarios (SSP1-2.6, SSP2-4.5, SSP3-7.0, SSP5-8.5) from 2015 to 2100 which are combined SSP and RCP. All GCM air temperature dataset are used as input data for the model. The raw resolutions of most ESMs are low. The DL SR model is effective in downsampling data of integer multiples such as 2, 4 or 8. The raw resolution of ESMs are very different; we therefore unify their resolutions before further analyses. About 1/3 of the total of ESMs used in this work have raw resolutions less than 2°. To retain more information in the original data, we resample all ESMs data to 2° × 2° resolution using bilinear interpolation. The input data are 31 ESM temperature dataset and the output is one merged data at 0.5° × 0.5° resolution. We split the data into 80% training set (1901–1992) and 20% validation set (1993–2014).

2.1.2 | Observation data

The CRU TS (Climate Research Unit gridded Time Series) is an extensive monthly land climate dataset which includes temperature and precipitation (http://doi.org/10/gbr3nj) over global land, except the Antarctica, with a high spatial resolution at 0.5° × 0.5° from 1901 to 2019. It was made from the analysis of more than 4,000 independent weather station and has high credibility (Harris et al., 2020). From the birth of the first version of CRU TS data in 2000, it has been widely used in many applications such as data evaluation (Fan et al., 2020), climate variability analysis (Wang et al., 2013) and independent climate mode (Renard and Tilman, 2019) due to its high quality. Due to the resolution mismatch of the observation data in the ocean such as ERSST (Huang et al., 2017a) and the CRU dataset time-series, our study concentrates only on the land. We use the 2 m air temperature from the latest version of the CRU TS as the reference data to train our model.

2.2 | Methods

In this study, we aim to establish a relationship between the historical CMIP6 ESMs temperature data and the observation data based on the DL models from 1901 to 2014. In essence, deep learning methods is a data merge and downsampling technology based on a nonlinear relationship. We then use the optimal model to predict the future data under four scenarios (SSP1-2.6, SSP2-4.5, SSP3-7.0, and SSP5-8.5) and extend the historical data from 1850 to 1900. The DL super resolution models mainly consist of a feature extraction part and up-sample part. First, the features of the input data are extracted which are up-sampled to the target resolution as output. The flowchart of this study is presented in Figure 1.

2.2.1 | Five different DL models

To find a suitable model, we select different methods with different classic DL structures such as the Residual Network (ResNet), Attention mechanism, Dense Convolutional Network (DenseNet), Generative Adversarial Networks (GAN) and Deformable Convolution Networks (DCN). Five SR methods are designed in this study. Four networks are single image SR methods, including the Enhanced Deep Residual Network (EDSR), Residual Dense Network (RDN), Residual Channel Attention Networks (RCAN) and Enhanced Super-Resolution Generative Adversarial Networks (ESRGAN); the Enhanced Deformable Convolutional Networks for Video Restoration (EDVR) is the last SR method, originally designed for video super-resolution. Each method is briefly described below:

ResNet solves network depth limitation due to the vanishing gradient problem (He et al., 2016). It achieved great success and is widely applied in different fields of computer vision. An example is the image super resolution method SRResNet (Ledig et al., 2017) with structural improvements later by proposed Wang et al. (2019a) and named EDSR. The most obvious feature is the removal of the batch normalization layers in the network as seen in Figure S1a. Besides improved performance, this new baseline model requires few parameters, saving about 40% computational time.

DenseNet is a novel network that also deals with the vanishing gradient problem. The main structure is that the input of each next layer is the output of all previous layers; this means every layer has direct access to the final output. It has a narrow internal structure and few parameters (Huang et al., 2017b). Because most deep CNN-based SR networks do not fully utilize hierarchical features from the initial low-resolution (LR) image, an RDN with a creative residual dense network based on ResNet and DenseNet (Figure S1b) is proposed to deal with this problem (Xu et al., 2018).

An attention mechanism is an indispensable part of natural language processing (NLP) algorithms in applications such as speech recognition and question answering.
Attention mechanism solve a problem that long distance information is very difficult to remember. After the creation of the transformer network structure, mainly based on attention blocks (Vaswani et al., 2017), attention mechanism became popular in the other fields of computer vision, such as image SR. Researchers invented a residual in residual (RIR) structure to form a very deep network, adding channel attention to a residual block as shown in Figure S1c (Zhang et al., 2018). RCAN can learn low and high frequency information and adaptive rescaling of channel characteristics by considering the interdependence between channels.

The GAN comprises two models: a generative model G to capture the feature and a discriminative model D to estimate the result from the G. The task of D is to minimize the probability of G making an error (Goodfellow et al., 2014). To retain finer texture detail at large upscaling factors, an image SR method—SRGAN—is proposed based on GAN (Ledig et al., 2017). However, the result of the SRGAN is often accompanied by unpleasant artefacts; the ESRGAN is designed to improve the performance of this network (Wang et al., 2019c).

A CNN has a fixed model geometric structure that restricts geometric transformations. To enhance model transformation, the DCN is proposed for the enlargement of the spatial sampling regions (Dai et al., 2017). The EDVR applies the DCN and Pyramid structure to design a Pyramid, Cascading and Deformable (PCD) network as illustrated in Figure S2 (Wang et al., 2019b). Although this method is proposed for video restoration, it also achieves a state-of-art performance in video SR. Unlike the single image SR algorithm, the EDVR needs a time series sequence input to learn spatial and temporal information and output the central frame image. Climate data have a strong relationship in the temporal dimension; so we also use the method to have an attempt to merge the ESM data.

### 2.2.2 Evaluation metrics

**Training and validation metrics**

To evaluate the different SR models, we adopt the same loss function for all networks. In the training process, the mean square error (MSE) is used as the loss function. The equation is described as follows:

\[
MSE = \frac{1}{N} \sum_{n=1}^{N} (Y_n - P_n)^2
\]

where \(N\) is the number of samples, \(Y_n\) is the observation temperature and \(P_n\) is the output of model data. mean absolute error (MAE), Pearson correlation coefficient (R) and MSE are used for a more comprehensive results evaluation in the validation phase. MAE and R are calculated as follows:

\[
MAE = \frac{1}{N} \sum_{n=1}^{N} |Y_n - P_n|
\]
The parameter meaning is the same as formula (1).

Besides, Taylor diagrams (Taylor, 2001) are used to compare the observation, merged data and model data. The $R$, standard deviations of error and unbiased root-mean-square deviation (ubRMSD) are calculated as follows:

$$R = \frac{\sum_{n=1}^{N} (Y_n - \bar{Y})(P_n - \bar{P})}{\sqrt{\sum_{n=1}^{N} (Y_n - \bar{Y})^2 \sum_{n=1}^{N} (P_n - \bar{P})^2}}$$  \hspace{1cm} (3)$$

$$RMSD = \sqrt{\frac{1}{N} \sum_{n=1}^{N} (Y_n - P_n)^2}$$  \hspace{1cm} (4)$$

$$BIAS = \frac{1}{N} \sum_{n=1}^{N} (Y_n - P_n)$$  \hspace{1cm} (5)$$

$$ubRMSD = \sqrt{RMSD^2 - BIAS^2}$$  \hspace{1cm} (6)$$

where $\beta$ is the ratio of the coefficient of variation between the observation data and the model simulated data. $\sigma_{\text{sim}}$ and $\mu_{\text{sim}}$ are standard deviation and mean value of model output respectively; $\sigma_{\text{obs}}$ and $\mu_{\text{obs}}$ are standard deviation and mean value of model output respectively. $\gamma$ is the histogram intersection for the histogram $K$ of the observation data and the histogram $L$ of the output of model, each containing $n$ bins. To compare different units variables and to guarantee bias insensitivity, the z score method is used to calculate $\gamma$. The closer this value is to 1, the better the data quality.

### 2.2.3 Model training settings

Due to the lack of suitable ocean data, we use zero padding to cover the ocean regions to reduce the effect during the training process. We divide global land into six regions (AUS, SA, NA, EUR, AFR and ASIA) as shown in Figure 2. Considering the boundary effects which means the output of the DL models will have more errors in the boundary regions because of the convolution struc-

![Figure 2](wileyonlinelibrary.com)
ture, we reserve duplicate areas between different zones. In the training process, we introduce the land-ocean mask to eliminate ocean interference in the results.

For a fair comparison and the selection of a suitable model, we use the same parameter settings such as the loss function (MSE), learning rate and training epochs for all models. The learning rate is 0.0001 at the beginning of the training and we use the learning rate attenuation strategy (learning rate divide by 3 per 20 epochs) to reach a better result. We set 100 epochs for each model.
and each epoch iterates over the training dataset about 1,000 times. Besides, we use the Adam optimizer with a gradient clip in the network to reduce over-fitting in the training process.

3 | RESULTS AND DISCUSSION

Figure 3 depicts the loss variations in training and validation phases. A sharp loss decline is observed for all models at the start with subsequent moderate fluctuations till model convergence. Almost all model achieve convergence at about 50 epochs. According to the loss performance in the validation phase, we save the best model of individual model according to the result of validation dataset in different areas.

The evaluation metrics for the validation phase of all models in various regions are shown in Table 2. It is noteworthy that three metrics (MSE, MAE and R) have little differences between various models. The EDSR achieve the best performance in all regions except AUS. The EDVR has the best scores for the MAE and MSE in AUS, and is slightly better than the EDSR; however, the R of EDSR is better than that of EDVR. All regions have very high R values with relatively large differences in MSE and MAE values. The MSE and MAE values for EUR and ASIA are larger compared with other regions. The SA and AFR regions have the least error. NA and ASIA have relatively higher error results compared with the rest of regions. In summary, the EDSR model outperforms the other DL approaches; therefore, we choose EDSR for the data merging. To understand the spatial error distribution, we calculate the mean MAE by pixel of all model outputs in the validation dataset. We also compute the mean spatial deviation by the EM data and compare with the DL results. For an intuitive comparison, we use bilinear interpolation to resample all the ESM model data to a 0.5° × 0.5° resolution, same as the observation and merged data.

Figure 4 shows the spatial distribution of the MAE for various methods in the validation phase. Figure 4a–e shows the results of DL models and Figure 4f reveals the mean error of EM methodology. The results of the DL models show significant improvement in almost every pixel compared with the EM results. Overall, the performance of a DL model is to a great extent determined by the input ESM data. From the EM results, the whole Northern Hemisphere, including Asia, Europe and North American show high errors (MAE is more than 2 in most regions), indicating that the raw model data is not accurate. In the Southern Hemisphere, the situation is much better; only some sub-regions show high errors (MAE exceeds 2) with most areas having relatively low errors (MAE less than 1.5). Remarkably, the DL model results have similar spatial patterns of MAE. Majority of the regions have considerably small errors except the upper

| Region | ESRGAN | RCAN | RDN | EDSR | EDVR |
|--------|--------|------|-----|------|------|
| AUS    | MSE: 1.09 | MSE: 1.10 | MSE: 1.07 | MSE: 1.05 | MSE: 1.03 |
|        | MAE: 0.80 | MAE: 0.80 | MAE: 0.79 | MAE: 0.79 | MAE: 0.77 |
|        | R: 0.9853 | R: 0.9855 | R: 0.9859 | R: 0.9860 | R: 0.9859 |
| SA     | MSE: 0.77 | MSE: 0.72 | MSE: 0.74 | MSE: 0.67 | MSE: 0.72 |
|        | MAE: 0.66 | MAE: 0.63 | MAE: 0.64 | MAE: 0.60 | MAE: 0.64 |
|        | R: 0.9937 | R: 0.9940 | R: 0.9945 | R: 0.9947 | R: 0.9938 |
| NA     | MSE: 3.90 | MSE: 3.63 | MSE: 3.86 | MSE: 3.36 | MSE: 3.84 |
|        | MAE: 1.45 | MAE: 1.39 | MAE: 1.42 | MAE: 1.33 | MAE: 1.42 |
|        | R: 0.9936 | R: 0.9939 | R: 0.9937 | R: 0.9943 | R: 0.9935 |
| EUR    | MSE: 4.45 | MSE: 4.33 | MSE: 4.59 | MSE: 4.24 | MSE: 4.88 |
|        | MAE: 1.54 | MAE: 1.51 | MAE: 1.55 | MAE: 1.50 | MAE: 1.66 |
|        | R: 0.9783 | R: 0.9792 | R: 0.9778 | R: 0.9798 | R: 0.9752 |
| AFR    | MSE: 0.86 | MSE: 0.73 | MSE: 0.75 | MSE: 0.71 | MSE: 0.75 |
|        | MAE: 0.71 | MAE: 0.65 | MAE: 0.66 | MAE: 0.64 | MAE: 0.65 |
|        | R: 0.9873 | R: 0.9896 | R: 0.9892 | R: 0.9898 | R: 0.9895 |
| ASIA   | MSE: 4.73 | MSE: 4.41 | MSE: 4.43 | MSE: 4.16 | MSE: 4.53 |
|        | MAE: 1.55 | MAE: 1.47 | MAE: 1.47 | MAE: 1.42 | MAE: 1.47 |
|        | R: 0.9917 | R: 0.9926 | R: 0.9924 | R: 0.9930 | R: 0.9920 |

Note: The boldened values indicate the best performance.
northern latitudes. An evident regularity can be observed; the error becomes increasingly larger along the latitudes in North Hemisphere. The Tibetan plateau region is one of the polar regions in the world, very hard to simulate. From our results, a significant improvement of the error in the Tibetan plateau area is observed. The same conclusion can be obtained that the EDSR is the optimal model in our task. Although DL methods have similar spatial distributions of MAE, a few high MAE regions are observed in a detailed inspection of the EDSR model result in Figure 5a.

We randomly select six pixels in individual continents to evaluate performance in the temporal dimension. The time series is shown in Figure 5. From Figure 5, the merged data has a remarkably better performance than the EM data compared to the observation data. Generally, the peak high and low temperatures tend to have higher bias over the time series. Six pixels can be classified into 3 group based on the results. Figure 5a,d is in one group; the observation data have a lot of small fluctuations besides the cyclic variation; the merged data, notwithstanding, reduce the bias but not reconstruct the wave motion of the whole series. The second group consists of Figure 5b,c; the observation data exhibits a relatively smooth period and the merged data has a good fit with the reference data. Figure 5e,f comprises the last group; the observed data also have a smooth period; however, there are fluctuations in place of the high and low in the time series. The merged data learns this wave despite the considerable errors in these places. The merged data have excellent results where the observation data have few fluctuations. On the contrary, where the observation data have a lot of waves, the performance of the merged data decreases.

We use the Taylor diagrams to further evaluate the DL merged data, EM data and the observation. Figure 6 depicts the Taylor diagrams for results of the monthly mean temperature from 1993 to 2014 in different
datasets. It is not surprising to find that the merged data achieve stable and superior performance (low bias and high R) in all regions compared to other data. Overall, all ESMs and EM data have very high R, exceeding 0.95. The EM data have a relatively better performance than the individual models globally. On the continental scale, there are clear differences between individual ESMs. In Asia and North America, as shown in Figure 6c,f, the performances of individual ESMs have no obvious variations and get results similar to the global. In other continents, the entire ESMs results show a dispersion trend and a relatively low performance. For R, the EM data also achieve excellent results, second to the merged data. However, for standard deviation and ubRMSD, the performance of the EM data is not as good as the merged data, even not better than single ESM data in some areas. The worst result is observed in Africa as shown in Figure 6g. All ESMs with R less than 0.95 and ubRMSD larger than 0.5 have more scattered distributions.

The Taylor diagrams for mean temperature show that the DL merged data has the best performance. Besides, we also use the SPAEF to compare DL data with individual ESM and EM data. The SPAEF of the merged data, EM data and different ESMs data depicted in Figure 7. It is obvious that the merged data has the highest performance on global and continental scales. On the contrary, the SPAEF of EM and ESMs data have different outcomes in global and regional areas. From Figure 7a, the EM data also have excellent SPAEF scores, second to the merged data globally. On the continental scale, the performances of EM data are not stable, similar to the performances of the ESMs data. In Figure 7c,g, the SPAEF of EM data in Asia and Africa are obviously lower than that of some single ESM results. It is noteworthy that the range of SPAEF is not restricted between 0 and 1 like some other correlation metrics. Generally, global scale scores of all data exceed 0.6; nonetheless, several negative values appear in some regional results of Asia, Australia, Europe and North America as seen in Figure 7c–f. Almost all data SPAEF performances decrease, moving from global to regional scale. Three ESMs (KACE-1-0-G, MCM-UA-1-0 and MIROC-ES2L) data achieve almost negative SPAEF in Australia, Europe and North America; this is mainly due to the raw low resolution (less than 2° × 2°) of these models as shown in Table 1.

Based on the ESMs data, we use the best DL model not only to merge the future temperature dataset but also the historical data from 1850 to 1900. From the Figure 8a, we can see anomalies of the merged data and all CMIP6 models from 1850 to 2100, using the observation data (1995–2014) as the reference period. Obviously, the highest temperature increase occurred in SSP-5.85. The uptrends are also observed in SSP3-7.0, SSP2-4.5 and SSP1-2.6 scenarios. Overall, the EM data tend to be lower in temperature than the observation data, while the DL merged data have results closer to the global observation data. The merged data is slightly higher than the EM data in both historical period and future scenarios. The consistency of
the results prove the stability of the DL model in this work. We therefore conclude that the merged data is credible. In Figure 8b, we use the EM and DL data to calculate the temperature variations respectively in different periods. In order, we select the mean historical temperature (1851–1870) and mean future temperature (2081–2100) to match the reference period of 1995 to 2014. The results of the DL data show slight temperature increase in historical and different future scenarios than the EM data. This phenomenon reveals the EM data have high estimation of temperature increase; it is observed in South America from the regional results in Figure S4 and Table S1. Moreover, the decadal trend of the four future scenarios in Figure S5 showing an upward trend.

From the evaluation metrics, the DL merged data outperforms the individual ESM and EM data in all aspects. Overall, the performance of EM data is second to the DL merged data although poorer than single models on a regional scale. Deep-learning methods are data-driving algorithms (Reichstein et al., 2019) and their performances are
FIGURE 7  SPAEF comparison of CMIP6 models, observations, DL merged data and EM data from 1901 to 2014 on global and continental land [Colour figure can be viewed at wileyonlinelibrary.com]
to a large extent determined by the input data. From our results and other studies (Fan et al., 2020), the errors of raw ESM data in the Northern Hemisphere are very high (MAE of most regions exceeds 2). The DL merged data have great improvements in Northern Hemisphere except for some regions far north where high errors exist (MAE exceeds than 1.5). It is worth noting that the bias in the Tibetan plateau area is relatively low in the DL merged data (MAE less than 1) under some raw model have a 5°C underestimate in this area (Fan et al., 2020). The outcome proves the effectiveness of the DL model bias correction during spatial downscaling and date merging.

For the pixel time series, the DL merged data cannot achieve a good outcome when the temperature has a lot of small waves in a big cycle and in the other condition it shows a good performance. All ESMs do not perform well in Africa compared to other continents from the time series evaluation (Figure 6g); this is mainly due to the limited availability of ground measurements for CRU in Africa (Collins, 2011). The merged data have a relatively low progress in the temporal dimension compared to the spatial dimension. This is because the selected DL model is a single-image DL approach which does not consider information in a temporal dimension. The temperature is showing an obviously increase trend under historical and four future scenarios, this result is similar with other result (Fan et al., 2020; Wu et al., 2022). The DL merged data have showed a relatively small rise than the EM data globally. From the regional scale, researchers found that EM mean temperature data underestimates temperature in three main river basins (Yellow River basin, Yangtze River basin and Southwest River basin) of China (Lu et al., 2022).

In previous studies, the DL methods based on simple architectures were reported to have good performance in the spatial downscaling (Huang, 2020). The DL downscaling method has been applied successfully in the CMIP6 soil moisture datasets of China, the DL method also performed a large improvement than the raw CMIP6 soil moisture datasets (Feng et al., 2022). In our work, we apply methods with relatively complex architectures. The best model according to the validation metrics is not the newest algorithm and does not outperform the other methods in other computer vision tasks. Besides, the video SR algorithm considers information in both spatial and temporal dimensions (Chan et al., 2021); this method should have been ideal to analyse the spatio-temporal relationship of climate data and achieve better results compared with a method that considers spatial features only. However, the EDVR did not achieve the best performance in our work. This outcome calls for more comparative experiments to be done for specific applications of computer vision in other disciplines.

Our work used DL methods to generate high-resolution future temperature datasets under four scenarios (SSP1-2.6, SSP2-4.5, SSP3-7.0, SSP5-8.5). However, there are some limitations of this work. First, the observation data we use is a reanalysis gridded data collected from meteorological stations; some errors that cannot be avoided affect the accuracy of this work. In the future work, weather station data need to be combined with the reanalysis data to achieve better results. Second, no suitable ocean observation data are available to match the spatial resolution of the CRU temperature datasets and time series simultaneously; we only merge high-resolution data in the land area. Additionally, the lack of SST also makes us unable to consider SST in the model via the training process. Taking the ocean into account in future studies will most likely improve the model outputs because the DL
models can learn extra information between the ocean and land. Third, our work does not explicitly take advantage of the temporal information of climate data. In other applications such as temperature reconstruction (Kadow et al., 2020), making full use of temporal information is a challenge which is worth looking into in future studies.

4 | CONCLUSIONS

In this study, we use deep learning (DL) methods to generate global land temperature datasets (except Antarctica) at a higher spatial resolution (0.5°) based on 31 different CMIP6 ESMs. Our methods can perform bias correction, spatial downscaling and data merging simultaneously. Five different DL methods are evaluated and the optimal model is selected to generate the historical data from 1851 to 1900, and future scenarios simulation data (SSP1-2.6, SSP2-4.5, SSP3-7.0 and SSP5-8.5) from 2015 to 2100. The merged data have a remarkably better quality compared with the individual ESMs and the ensemble of all ESMs in both spatial and temporal dimensions. The error metrics (e.g., R and ubRMSE) also show that the merged data have better accuracy in both global and regional areas.

The merged data have great improvement in spatial error. Besides the high latitudes of Asia and North America, most areas have a low MAE (less than 1°C). It means the merged data take full advantage of the raw CMIP6 model data. From the pixel time series, the DL merged data can achieve excellent performance when the observation is smooth with few fluctuations. Overall, the merged data progresses in the temporal dimension. The merged data have better results in spatial distribution than in temporal variation since the DL model mainly focuses on spatial features.

Our work is one of the first to perform spatial downscaling, bias correction and data merging simultaneously. Our results demonstrate the successful application of DL SR models for data merging and downscaling tasks when enough training data are available. The generated merged data with improved accuracy have great potential for many applications such as spatial–temporal change of future climate, and the influence of climate change on water resources and agriculture.

AUTHOR CONTRIBUTIONS

Xikun Wei: Conceptualization; methodology; software; data curation; investigation; validation; formal analysis; writing – original draft. Guojie Wang: Conceptualization; project administration; resources; supervision; funding acquisition. Donghan Feng: Data curation; software. Zheng Duan: Data curation. Daniel Fiifi Tawia Hagan: Software. Liangliang Tao: Visualization; validation. Lijuan Miao: Investigation; validation. Buda Su: Validation; visualization. Tong Jiang: Visualization; formal analysis.

ACKNOWLEDGEMENTS

This work was supported by National Key Research and Development Program of China (2019YFC1510203), General project of National Natural Science Foundation of China (41875094), the Sino-German Cooperation Group Project of China (GZ1447) and the Postgraduate Research and Practice Innovation Program of Jiangsu Province, grant number of KYCX22_1155.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

ORCID

Lijuan Miao https://orcid.org/0000-0003-2050-9157

REFERENCES

Adachi, S.A. and Tomita, H. (2020) Methodology of the constraint condition in dynamical downscaling for regional climate evaluation: a review. Journal of Geophysical Research-Atmospheres, 125(11), e2019JD032166. https://doi.org/10.1029/2019JD032166.

Weigel, A.P., Liniger, M. and Appenzeller, C. (2008) Can multimodel combination really enhance the prediction skill of probabilistic ensemble forecasts? Quarterly Journal of the Royal Meteorological Society, 1227, 496. https://doi.org/10.1002/qj.

Asong, Z.E., Khaliq, M.N. and Wheater, H.S. (2016) Projected changes in precipitation and temperature over the Canadian Prairie Provinces using the generalized linear model statistical downscaling approach. Journal of Hydrology, 539, 429–446.

Avaria-Llautureo, J., Venditti, C., Rivadeneira, M.M., Inostroza-Michael, O., Rivera, R.J., Hernández, C.E. and Canales-Aguirre, C.B. (2021) Historical warming consistently decreased size, dispersal and speciation rate of fish. Nature Climate Change, 11(9), 787–793. https://doi.org/10.1038/s41558-021-01123-5.

Baghanam, A.H., Nourani, V., Sheikhbabaei, A. and Seifi, A.J. (2020) Statistical downscaling and projection of future temperature change for Tabriz city, Iran. In: IOP Conference Series Earth and Environmental Science, Vol. 491. Kerala: IOP Publishing, pp. 012009.

Barth, A., Alvera-Azcárate, A., Licer, M. and Beckers, J.M. (2020) DINCAE 1.0: a convolutional neural network with error estimates to reconstruct sea surface temperature satellite observations. Geoscientific Model Development, 13(3), 1609–1622. https://doi.org/10.5194/gmd-13-1609-2020.

Blau, Y., Mechrez, R., Timofte, R., Michaeli, T. and Zelnik-Manor, L. (2018) The 2018 FIRM challenge on perceptual image super-resolution. In: The European Conference on Computer Vision (ECCV) Workshops. Munich.

Bruyère, C.L., Done, J.M., Holland, G.J. and Fredrick, S. (2014) Bias corrections of global models for regional climate simulations of high-impact weather. Climate Dynamics, 43, 1847–1856.
Calvin, K., Bond-Lamberty, B., Clarke, L., Edmonds, J., Eom, J., Hartin, C., Kim, S., Kyle, P., Link, R., Moss, R., McJeon, H., Patel, P., Smith, S., Waldhoff, S. and Wise, M. (2017) The SSP4: a world of deepening inequality. Global Environmental Change, 42, 284–296. https://doi.org/10.1016/j.gloenvcha.2016.06.010.

Chan, K.C.K., Wang, X., Yu, K., Dong, C. and Loy, C.C. (2021) BasicVSR: the search for essential components in video super-resolution and beyond. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), online, June. Available at: http://arxiv.org/abs/2012.02181

Chen, Y., Liu, S. and Wang, X. (2020) Learning continuous image representation with local implicit image function. arXiv.

Choi, D., Gao, Z. and Jiang, W. (2020) Attention to global warming. Review of Financial Studies, 33(3), 1112–1145. https://doi.org/10.1093/rfs/hhz086.

Clem, K.R., Fogt, R.L., Turner, J., Lintner, B.R., Marshall, G.J., Miller, J.R. and Renwick, J.A. (2020) Record warming at the south pole during the past three decades. Nature Climate Change, 10(8), 762–770. https://doi.org/10.1038/s41558-020-0815-z.

Colette, A., Vautard, R. and Vrac, M. (2012) Regional climate downscaling with prior statistical correction of the global climate forcing. Geophysical Research Letters, 39, L13707.

Collins, J.M. (2011) Temperature variability over Africa. Journal of Climate, 24(14), 3649–3666. https://doi.org/10.1175/2011JCLI3753.1.

Dai, J., Qi, H., Xiong, Y., et al. (2017) Deformable convolutional networks. In: Proceedings of the IEEE International Conference on Computer Vision. Venice: IEEE, pp. 764–773.

Das, L. and Akhter, J. (2019) How well are the downscaled CMIP5 models able to reproduce the monsoon precipitation over seven homogeneous zones of India? International Journal of Climatology, 39, 3323–3333.

Eyring, V., Bony, S., Meehl, G.A., Senior, C.A., Stevens, B., Stouffer, R.J. and Taylor, K.E. (2016) Overview of the coupled model Intercomparison project phase 6 (CMIP6) experimental design and organization. Geoscientific Model Development, 9(5), 1937–1958. https://doi.org/10.5194/gmd-9-1937-2016.

Eyring, V., Cox, P.M., Flato, G.M., Gleckler, P.J., Abramowitz, G., Caldwell, P., Collins, W.D., Gier, B.K., Hall, A.D., Hoffman, F.M., Hurtt, G.C., Jahn, A., Jones, C.D., Klein, S.A., Krasting, J.P., Kwiatkowski, L., Lorenz, R., Maloney, E., Meehl, G.A., Pendergrass, A.G., Pincus, R., Ruan, A.C., Russell, J.L., Sanderson, B.M., Santer, B.D., Sherwood, S.C., Simpson, I.R., Stouffer, R.J. and Williamson, M.S. (2019) Taking climate model evaluation to the next level. Nature Climate Change, 9(2), 102–110. https://doi.org/10.1038/s41558-018-0355-y.

Fan, X., Liao, L. and Gou, J. (2021) Statistical downscaling and projection of future temperatures across the loess plateau, China. Weather and Climate Extremes, 32, 100328. https://doi.org/10.1016/j.wace.2021.100328.

Feng, D.H., Wang, G.J., Wei, X.K., Amankwah, S.O.Y., Hu, Y.F., Lou, Z.C., et al. (2022) Merging and downsampling soil moisture data from CMIP6 projections using deep learning method. Frontiers in Environmental Science. https://doi.org/10.3389/fenvs.2022.847475.

Fricko, O., Havlík, P., Rogelj, J., Klimont, Z., Gusti, M., Johnson, N., Kolp, P., Strubegger, M., Valin, H., Amann, M., Ermolieva, T., Forssell, N., Herrero, M., Heyes, C., Kindermann, G., Krey, V., McCollum, D.L., Obersteiner, M., Pachauri, S., Rao, S., Schmid, E., Schoepf, W. and Riahi, K. (2017) The marker quantification of the shared socioeconomic pathway 2: a middle-of-the-road scenario for the 21st century. Global Environmental Change, 42(July), 251–267. https://doi.org/10.1016/j.gloenvcha.2016.06.004.

Fujisori, S., Hasegawa, T., Masui, T., Takahashi, K., Herran, D.S., Dai, H., Hijikoa, Y. and Kainuma, M. (2017) SSP3: AIM implementation of shared socioeconomic pathways. Global Environmental Change, 42, 268–283. https://doi.org/10.1016/j.gloenvcha.2016.06.009.

Gillett, N.P., Kirchmeier-Young, M., Ribes, A., Shiohama, H., Hegerl, G.C., Knutti, R., Gastineau, G., John, J.G., Li, L., Nazarenko, L., Rosenbloom, N., Seland, Ø., Wu, T., Yukimoto, S. and Ziehn, T. (2021) Constraining human contributions to observed warming since the pre-industrial period. Nature Climate Change, 11, 207–212. https://doi.org/10.1038/s41558-020-00965-9.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y. (2014) Generative adversarial nets. Communications of the ACM, 63(11), 139–144. https://doi.org/10.1145/3422622.

Gudmundsson, L., Brennnes, J.B., Haugen, J.E. and Engen-Skaugen, T. (2012) Technical note: downscaling RCM precipitation to the station scale using statistical transformations—a comparison of methods. Hydrology and Earth System Sciences, 16(9), 3383–3390. https://doi.org/10.5194/hess-16-3383-2012.

Gutiérrez, J.M., San-Martín, D., Brands, S., Manzanas, R. and Herrera, S. (2013) Reassessing statistical downscaling techniques for their robust application under climate change conditions. Journal of Climate, 26(1), 171–188. https://doi.org/10.1175/JCLI-D-11-00687.1.

Ham, Y.G., Kim, J.H. and Luo, J.J. (2019) Deep learning for multi-year ENSO forecasts. Nature, 573(7775), 568–572. https://doi.org/10.1038/s41586-019-1559-7.

Harris, I., Osborn, T.J., Jones, P. and Lister, D. (2020) Version 4 of the CRU TS monthly high-resolution gridded multimarine climate dataset. Scientific Data, 7(1), 1–19. https://doi.org/10.1038/s41597-020-0453-3.

Hasegawa, T., Sakurai, G., Fujisori, S., Takahashi, K., Hijiyoka, Y. and Masui, T. (2021) Extreme climate events increase risk of global food insecurity and adaptation needs. Nature Food, 2(8), 587–595. https://doi.org/10.1038/s43016-021-00335-4.

He, K., Zhang, X., Ren, S. and Sun, J. (2016) Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Vol. 2016. New York: IEEE, pp. 770–778. https://doi.org/10.1109/CVPR.2016.90.

Ho, C.K., Stephenson, D.B., Collins, M., Ferro, C.A.T. and Brown, S.J. (2012) Calibration strategies a source of additional uncertainty in climate change projections. Bulletin of the American Meteorological Society, 93(1), 21–26. https://doi.org/10.1175/2011BAMS3110.1.
Huang, B., Thorne, P.W., Banzon, V.F., Boyer, T., Chepurin, G., Lawrimore, J.H., Menne, M.J., Smith, T.M., Vose, R.S. and Zhang, H.M. (2017a) Extended reconstructed sea surface temperature, version 5 (ERSSTv5): upgrades, validations, and intercomparisons. *Journal of Climate*, 30(20), 8179–8205. https://doi.org/10.1175/JCLI-D-16-0386.1.

Huang, G., Liu, Z., Van Der Maaten, L. and Weinberger, K.Q. (2017b) Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. New York: IEEE, pp. 4700–4708. https://doi.org/10.1109/CVPR.2017.243.

Huang, X. (2020) Deep-learning based climate downsampling using the super-resolution method: a case study over the western US. *Geoscientific Model Development Discussion*. https://doi.org/10.5194/gmd-2020-214.

Jakob Theimefl, M., Gobiet, A. and Leuprecht, A. (2011) Empirical-statistical downsampling and error correction of daily precipitation from regional climate models. *International Journal of Climatology*, 31(10), 1530–1544. https://doi.org/10.1002/joc.2168.

Jevrejeva, S., Jackson, L.P., Riva, R.E.M., Grinsted, A. and Moore, J. (2013) A nonparametric kernel regression approach to reconstructing missing climate information. *Nature Geoscience*, 6(10), 770–773. https://doi.org/10.1038/ngeo1983.

Kannan, S. and Ghosh, S. (2013) A nonparametric kernel regression model for downsampling multisite daily precipitation in the Mahanadi basin. *Water Resources Research*, 49(3), 1360–1385.

Koch, J., Demirel, M.C. and Sisen, S. (2018) The SPAtial EFFiciency metric (SPAEF): multiple-component evaluation of spatial patterns for optimization of hydrological models. *Geoscientific Model Development*, 11(5), 1873–1886. https://doi.org/10.5194/gmd-11-1873-2018.

Kriegler, E., Bauer, N., Popp, A., Humpenöder, F., Leimbach, M., Streffer, J., Baumstark, L., Bodirsky, B.L., Hilaire, J., Klein, D., Moraflowdou, I., Weindl, I., Bertram, C., Dietrich, J.P., Luderer, G., Pehl, M., Pietzcker, R., Piontek, F., Lotze-Campen, H., Biewald, A., Bonsch, M., Giannousakis, A., Kreidenweis, U., Müller, C., Rolinski, S., Schultes, A., Schwantz, J., Stevanovic, M., Calvin, K., Emmerling, J., Fujimori, S. and Edenhofer, O. (2017) Fossil-fuel development (SSP5): an energy and resource intensive scenario for the 21st century. *Global Environmental Change*, 42(September 2018), 297–315. https://doi.org/10.1016/j.gloenvcha.2016.05.015.

Lafferty, D.C., Sriver, R.L., Haqui, I., Hertel, T.W., Keller, K. and Nichols, R.E. (2021) Statistically bias-corrected and downscaled climate models underestimate the adverse effects of extreme heat on U.S. maize yields. *Communications Earth & Environment*, 2(1), 1–10. https://doi.org/10.1038/s43247-021-00266-9.

Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z. and Shi, W. (2017) Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. New York: IEEE, pp. 4681–4690. https://doi.org/10.1109/CVPR.2017.19.

Liu, Y., Ganguly, A.R. and Dy, J. (2020) Climate downsampling using YNet: a deep convolutional network with skip connections and fusion. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York: Association for Computing Machinery, pp. 3145–3153. https://doi.org/10.1145/3394486.3403566.

Lovino, M.A., Pierrestegui, M.J., Müller, O.V., Berbery, E.H., Müller, G.V. and Pasten, M. (2021) Evaluation of historical cmip6 model simulations and future projections of temperature and precipitation in Paraguay. *Climatic Change*, 164(3), 1–24.

Lu, K., Arshad, M., Ma, X., Ullah, I., Wang, J. and Shao, W. (2022) Evaluating observed and future spatiotemporal changes in precipitation and temperature across China based on CMIP6-GCMs. *International Journal of Climatology*, 1–18. https://doi.org/10.1002/joc.7673.

Lu, J., Wang, G., Li, S., et al. (2021) Projected land evaporation and its response to vegetation greening over China under multiple scenarios in the CMIP6 models. Journal of geophysical research. *Biogeosciences*, 126(9), e2021JG006327.

Maraun, D. and Widmann, M. (2018) Structure of statistical downscaling methods. In: *Statistical Downscaling and Bias Correction for Climate Research*. Cambridge: Cambridge University Press, pp. 135–140.

Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S.L., Péan, C., Berger, S., et al. (2021) IPCC, 2021: summary for policymakers. In: *Climate Change 2021: The Physical Science Basis*. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge, USA: Cambridge University Press. In Press.

Miao, C., Ashouri, H., Hsu, K.L., Sorooshian, S. and Duan, Q. (2015) Evaluation of the PERSIANN-CDR daily rainfall estimates in capturing the behavior of extreme precipitation events over China. *Journal of Hydro meteorology*, 16(3), 1387–1396. https://doi.org/10.1175/JHM-D-14-0174.1.

Miao, C., Su, L., Sun, Q. and Duan, Q. (2016) A nonstationary bias-correction technique to remove bias in GCM simulations. *Journal of Geophysical Research-Atmospheres*, 175(4449), 238. https://doi.org/10.1002/2015JD025380.

Moss, R.H., Edmonds, J.A., Hibbard, K.A., Manning, M.R., Rose, S.K., Van Vuuren, D.P., Carter, T.R., Emori, S., Kainuma, M., Kram, T., Meehl, G.A., Mitchell, J.F.B., Nakicenovic, N., Riahi, K., Smith, S.J., Stouffer, R.J., Thomson, A.M., Weyant, J.P. and Wilbanks, T.J. (2010) The next generation of scenarios for climate change research and assessment. *Nature*, 463(7282), 747–756. https://doi.org/10.1038/nature08823.

Naumann, G., Cammalleri, C., Mentaschi, L. and Feyen, L. (2021) Increased economic drought impacts in Europe with anthropogenic warming. *Nature Climate Change*, 11(6), 485–491. https://doi.org/10.1038/s41558-021-01044-3.

Norris, J., Hall, A., Chen, D., Thackeray, C.W. and Madakumbura, G.D. (2021) Assessing the representation of synoptic variability associated with California extreme precipitation in CMIP6 models. *Journal of Geophysical Research-Atmospheres*, 126, e2020JD033938.

O’Neill, B.C., Tebaldi, C., Van Vuuren, D.P., Eyring, V., Friedlingstein, P., Hurtt, G., Knutti, R., Kriegler, E., Lamarque, J.F., Lowe, J., Meehl, G.A., Moss, R., Riahi, K. and Sanderson, B.M. (2016) The scenario model Intercomparison
project (ScenarioMIP) for CMIP6. *Geoscientific Model Development*, 9(9), 3461–3482. https://doi.org/10.5194/gmd-9-3461-2016.

Papalexiou, S.M., Rajulapati, C.R., Clark, M.P. and Lehner, F. (2020) Robustness of CMIP6 historical global mean temperature simulations: trends, long-term persistence, autocorrelation, and distributional shape. *Earth’s Futures*, 8(10), e2020EF001667. https://doi.org/10.1029/2020EF001667.

Parsons, L.A., Brennan, M.K., Wills, R.C.J. and Proistosescu, C. (2020) Magnitudes and spatial patterns of Interdecadal temperature variability in CMIP6. *Geophysical Research Letters*, 47(7), e2019GL086588. https://doi.org/10.1029/2019GL086588.

Phillips, T.J. and Gleckler, P.J. (2006) Evaluation of continental precipitation in 20th century climate simulations: the utility of multimodel statistics. *Water Resources Research*, 42(3), 1–10. https://doi.org/10.1029/2005WR004313.

Quante, L., Willner, S.N., Middelans, R. and Levermann, A. (2021) Regions of intensification of extreme snowfall under future warming. *Scientific Reports*, 11(1), 1–9. https://doi.org/10.1038/s41598-021-95979-4.

Riahi, K., van Vuuren, D.P., Kriegler, E., Edmonds, J., Bauer, N., Calvin, K., DELLink, R., Fricco, O., Lutz, W., Popp, A., Crespo Cuaresma, J., Leimbach, M., Jiang, I., Kram, T., Rao, S., Emmerling, J., Ebi, K., Hasegawa, T., Havlik, P., Humphenöder, F., Da Silva, A., Smith, S., Stehfest, E., Bosetti, V., Eom, J., Gernaat, D., Masui, T., Rogelj, J., Strefler, J., Drouet, L., Krey, V., Luderer, G., Harmsen, M., Takahashi, K., Baumstark, L., Doelman, J., Kainuma, M., KLimont, Z., Marangoni, G., Lotze-Campen, H. and Osterbein, M. (2017) The shared socioeconomic pathways and their energy, land use, and greenhouse gas implications: an overview. *Global Environmental Change*, 42, 153–168.

Seif, G. and Androutos, D. (2018) Large receptive field networks for high-scale image super-resolution. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Work*. Salt lake City: IEEE, pp. 876–885. https://doi.org/10.1109/CVPRW.2018.00120.

Shi, W., Caballero, J., Huszar, F., Totz, J., Aitken, A.P., Bishop, R., Rueckert, D. and Wang, Z. (2016) 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*. New York: IEEE, pp. 1874–1883. https://doi.org/10.1109/CVPR.2016.207.

Shrestha, R.R., Schnorbus, M.A., Werner, A.T. and Zwiers, F.W. (2014) Evaluating hydroclimatic change signals from statistically and dynamically downscaled GCMs and hydrologic models. *Journal of Hydrometeorology*, 15(2), 844–860. https://doi.org/10.1175/JHM-D-13-030.1.

Stockhause, M. and Lautenschlager, M. (2017) CMIP6 data citation of evolving data. *Data Science Journal*, 16, 1–13. https://doi.org/10.5334/dsj-2017-030.

Tang, G., Clark, M.P. and Papalexiou, S.M. (2021a) Sc-earth: a station-based serially complete earth dataset from 1950 to 2019. *Journal of Climate*, 34(16), 6493–6511.

Tang, G., Clark, M.P., Papalexiou, S.M., Newman, A.J., Wood, A. W., Brunet, D. and Whitfield, P.H. (2021b) EMDNA: an Ensemble Meteorological Dataset for North America. *Earth System Science Data*, 13, 3337–3362. https://doi.org/10.5194/essd-13-3337-2021.

Taylor, K.E. (2001) Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research: Atmospheres*, 106(D7), 7183–7192.

Tian, L., Sun, J., Chang, J., Xia, J., Zhang, Z., Kolomenskii, A.A., Schuessler, H.A. and Zhang, S. (2021) Retrieval of gas concentrations in optical spectroscopy with deep learning. *Measurement: Journal of the International Measurement Confederation*, 182(July), 109739. https://doi.org/10.1016/j.measurement.2021.109739.

Toms, B.A., Barnes, E.A. and Ebert-Uphoff, I. (2020) Physically interpreted neural networks for the geosciences: applications to earth system variability. *Journal of Advances in Modeling Earth Systems*, 12(9), 1–20. https://doi.org/10.1029/2019MS002002.

Tong, X.Y., Xia, G.S., Lu, Q., Shen, H., Li, S., You, S. and Zhang, L. (2020) Land-cover classification with high-resolution remote sensing images using transferable deep models. *Remote Sensing of Environment*, 237, 111322. https://doi.org/10.1016/j.rse.2019.111322.

Touma, D., Stevenson, S., Lehner, F. and Coats, S. (2021) Human-driven greenhouse gas and aerosol emissions cause distinct regional impacts on extreme fire weather. *Nature Communications*, 12(1), 1–8. https://doi.org/10.1038/s41467-020-20570-w.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L. and Polosukhin, I. (2017) Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 5999–6009.

van Vuuren, D.P., Kriegler, E., O’Neill, B.C., Ebi, K.L., Riahi, K., Carter, T.R., Edmonds, J., Hallegatte, S., Kram, T., Mathur, R. and Winkler, H. (2014) A new scenario framework for climate change research: scenario matrix architecture. *Climatic Change*, 122(3), 373–386. https://doi.org/10.1007/s10584-013-0906-1.

Wang, H., Liao, K., Yan, B. and Ye, R. (2019a) Deep residual network for single image super-resolution. In *Proceedings of the 2nd International Conference on Control and Computer Vision*. New York: Association for Computing Machinery. pp. 66–70. https://doi.org/10.1145/3341016.3341030.

Wang, J., Yang, B., Ljungqvist, F.C. and Zhao, Y. (2013) The relationship between the Atlantic multidecadal oscillation and temperature variability in China during the last millennium. *Journal of Quaternary Science*, 28(7), 653–658. https://doi.org/10.1002/jqs.2658.

Wang, X., Chan, K.C.K., Yu, K., Dong, C. and Loy, C.C. (2019b) EDVR: video restoration with enhanced deformable convolutional networks. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Work, 2019, 1954–1963. https://doi.org/10.1109/CVPRW.2019.00247.

Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Qiao, Y. and Loy, C.C. (2019c) ESRGAN: enhanced super-resolution generative adversarial networks. In *Proceedings of the European
Wei, X., Hamdi, R., Ochege, F.U., Kurban, A. and De Maeyer, P. (2021) The sensitivity of global surface air temperature to vegetation greenness. International Journal of Climatology, 41(1), 483–496. https://doi.org/10.1002/joc.6633.

Zang, Z., Li, D., Guo, Y., Shi, W. and Yan, X. (2021) Superior PM2.5 estimation by integrating aerosol fine mode data from the Himawari-8 satellite in deep and classical machine learning models. Remote Sensing, 13(14), 2779.

Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B. and Fu, Y. (2018) Image super-resolution using very deep residual channel attention networks. In Proceedings of the IEEE International Conference on Computer Vision Workshops, Vol. 2018. Cambridge: IEEE, pp. 286–301. https://doi.org/10.1109/ICCVW.2018.9008.

Zheng, G., Li, X., Zhang, R.H. and Liu, B. (2020) Purely satellite data-driven deep learning forecast of complicated tropical instability waves. Science Advances, 6(29), eaba1482. https://doi.org/10.1126/sciadv.aba1482.

Zhukov, A. and Gushcha, S. (2020) Stop global warming. Journal of Education, Health and Sport, 10(3), 18–33. https://doi.org/10.12775/jehs.2020.10.03.002.

**SUPPORTING INFORMATION**

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Wei, X., Wang, G., Feng, D., Duan, Z., Hagan, D. F. T., Tao, L., Miao, L., Su, B., & Jiang, T. (2023). Deep-learning-based harmonization and super-resolution of near-surface air temperature from CMIP6 models (1850–2100). International Journal of Climatology, 43(3), 1461–1479. https://doi.org/10.1002/joc.7926