A 148-year precipitation oxygen isoscape for China generated based on data fusion and bias correction of iGCMs simulations

Jiacheng Chen [1], Jie Chen [1,2], Xunchang J. Zhang [3], Peiyi Peng [4], Camille Risi [5]

[1]State Key Laboratory of Water Resources & Hydropower Engineering Science, Wuhan University, Wuhan, 430072, China
5 [2]Hubei Key Laboratory of Water System Science for Sponge City Construction, Wuhan University, Wuhan, 430072, China
[3]USDA-ARS, Grazinglands Research Laboratory, 7207W. Cheyenne St., El Reno, OK 73036, USA
[4]Chongqing Southwest Research Institute for Water Transport Engineering, Chongqing Jiaotong University, Chongqing, 400016, China
[5]Laboratoire de Meteorologie Dynamique, IPSL, CNRS, Ecole Normale Superieure, Sorbonne Universite, PSL Research
10 University, Paris, France

Correspondence to: Jie Chen (jiechen@whu.edu.cn)

Abstract. The precipitation oxygen isotopic composition is a useful environmental tracer for climatic and hydrological studies. However, the observed precipitation oxygen is limited at both temporal and spatial scales. Isotope-equipped general circulation models (iGCMs) can compensate for the temporal and spatial discontinuity of observation networks, but they
15 suffer from coarse spatial resolutions and systematic biases. The objective of this study is to build a high-resolution precipitation oxygen isoscape in China for a period of 148 years by integrating observed and iGCMs-simulated precipitation oxygen isotope composition ($\delta^{18}O_p$) using data fusion and bias correction methods. The temporal and spatial resolutions are month and 50-60 km for the isoscape, respectively. Prior to building the oxygen isoscape, the performance of two bias correction methods (BCMs) and three data fusion methods (DFMs) is compared after post-processing of eight iGCM
20 simulations. Results show that the outputs of the Convolutional Neural Networks (CNN) fusion method exhibit the strongest correlation with observations with correlation coefficient mostly larger than 0.8, and the smallest bias with root mean square error mostly smaller than 2‰. The other two DFMs also perform slightly better than the two BCMs, which show similar performance. Thus, precipitation oxygen isoscape is generated by using the CNN fusion method for the 1969-2007 period in which all iGCMs have output and by using the bias correction methods for the remaining years. Based on the precipitation
25 oxygen isoscape, the spatiotemporal patterns of $\delta^{18}O_p$ across China are investigated. The generated isoscape shows similar spatial and temporal distribution characteristics to observations. In general, the distribution pattern of $\delta^{18}O_p$ is consistent with the temperature effect in northern China, and with the precipitation amount effect in southern China, and be more specific in the Qinghai-Tibet Plateau of China. The $\delta^{18}O_p$ time series mirrors a fluctuating upward trend of the temperature or precipitation in most regions of China. The temporal and spatial distribution characteristics of the generated isoscape are
30 consistent with the characteristics of atmospheric circulation and climate change, indicating successful assimilation and extension of the observed precipitation oxygen isotopes in time and space. Overall, the built isoscape is reliable and useful for providing strong support for tracking atmospheric and hydrological processes. The dataset is available in Zenodo at https://doi.org/10.5281/zenodo.5703811 (Chen et al., 2021).

1 Introduction

35 Stable oxygen and hydrogen isotopes are the components of water molecules in natural water bodies. The fractionation of stable isotopes, namely, the distribution of stable isotopes in different ratios between two phases, occurs during each phase transition in the water cycle (Dansgaard, 1964). Therefore, stable isotopes are very sensitive to environmental changes and can record the internal process of water cycles, providing an effective tracking method in the study of complex hydrological and climatic processes (Gibson et al., 2005; Gat, 1996; Galewsky et al., 2016; Ansari et al., 2020).

40 The stable isotopes in precipitation are generally obtained from station measurements. In 1961, the International Atomic Energy Agency (IAEA) and the World Meteorological Organization (WMO) launched the Global Network of Isotopes in Precipitation (GNIP). Due to the scarcity of stations on the Tibetan Plateau, the Chinese Academy of Sciences (CAS) launched the Tibetan Plateau Network of Isotopes in Precipitation (TNIP) in 1991 (Yu et al., 2016a). However, most Chinese stations in GNIP stopped monitoring in the early 2000s (Zhang and Wang, 2016). In order to continue the systematic study,

45 the CAS established the Chinese Network of Isotopes in Precipitation (CHNIP) based on the Chinese Ecosystem Research Network (CERN) in 2004. Due to the difficulty and high cost of measuring precipitation isotope ratios (Allen et al., 2018), most of the observed data are short in length. The spatial distribution of observation stations is uneven, with few stations in inaccessible areas (Wang et al., 2015).

The stable isotopes in precipitation can also be simulated by isotope-equipped general circulation models (iGCMs). In

50 contrast to observations, iGCMs can provide time-continuous and space-regular isotope data. The general circulation model (GCM) mainly describes the temporal and spatial changes of physical variables through physical equations, and the essence of iGCMs is to introduce the cycle of water stable isotopes into each stage of the water cycle in GCMs (Xi, 2014). Driven by the external climate boundary conditions, iGCMs can provide the composition of precipitation stable isotopes by calculating the basic mass, momentum, energy and water balance equations of air columns and reasonably parameterizing the basic

55 meteorological processes (Sturm et al., 2010). Because the initial conditions, boundary conditions and process parameters of each model are different, the outputs of each iGCM are various. The Stable Water Isotope Intercomparison Group (SWING) and the second phase of SWING have been established to have a comprehensive comparison among different iGCMs and the related isotope measurements (Zhang et al., 2012). On this basis, the comparison and evaluation of iGCMs have been conducted in many studies (Zhang et al., 2012; Wang et al., 2015; Risi et al., 2012; Che et al., 2016). Even though iGCMs

60 have reliable physical bases and can reflect the dynamic characteristics of water stable isotopes, they are computationally expensive and require thorough validation (Sturm et al., 2010; Galewsky et al., 2016). Moreover, the spatial resolution of iGCMs is on the order of hundreds kilometres, which is too coarse to investigate the water cycle for a specific region or watershed. Furthermore, outputs of iGCMs are usually subjected to systematic bias when compared with gauged observations, and there is no single iGCM that consistently performs better than the others (Zhang et al., 2012; Wang et al.,

65 2015; Conroy et al., 2013). The use of ensemble mean of multiple iGCMs is a desirable approach for better use of information available from all iGCMs.

Overall, both observations and iGCM simulations have advantages and disadvantages. The effort of constructing a database by taking advantages and circumventing disadvantages of both becomes a challenge. Accordingly, in order to make full use of the advantages of high quality observations and continuous iGCM simulations, this study aims at constructing a
70   high-resolution precipitation oxygen isoscape in mainland China by data fusion and bias correction of iGCM simulations. In order to determine the best scheme to build the dataset, two bias correction methods (BCMs) and three neural network data fusion methods (DFMs) are first compared in terms of bias correcting and fusing iGCM simulations. The new isoscape in monthly temporal and approximately 0.5° spatial resolutions is produced by combining the Convolutional Neural Networks (CNN) fusion method and bias correction method for the 1870-2017 period. The spatial and temporal distribution
75   characteristics of oxygen isotopes in precipitation are then analysed for China.
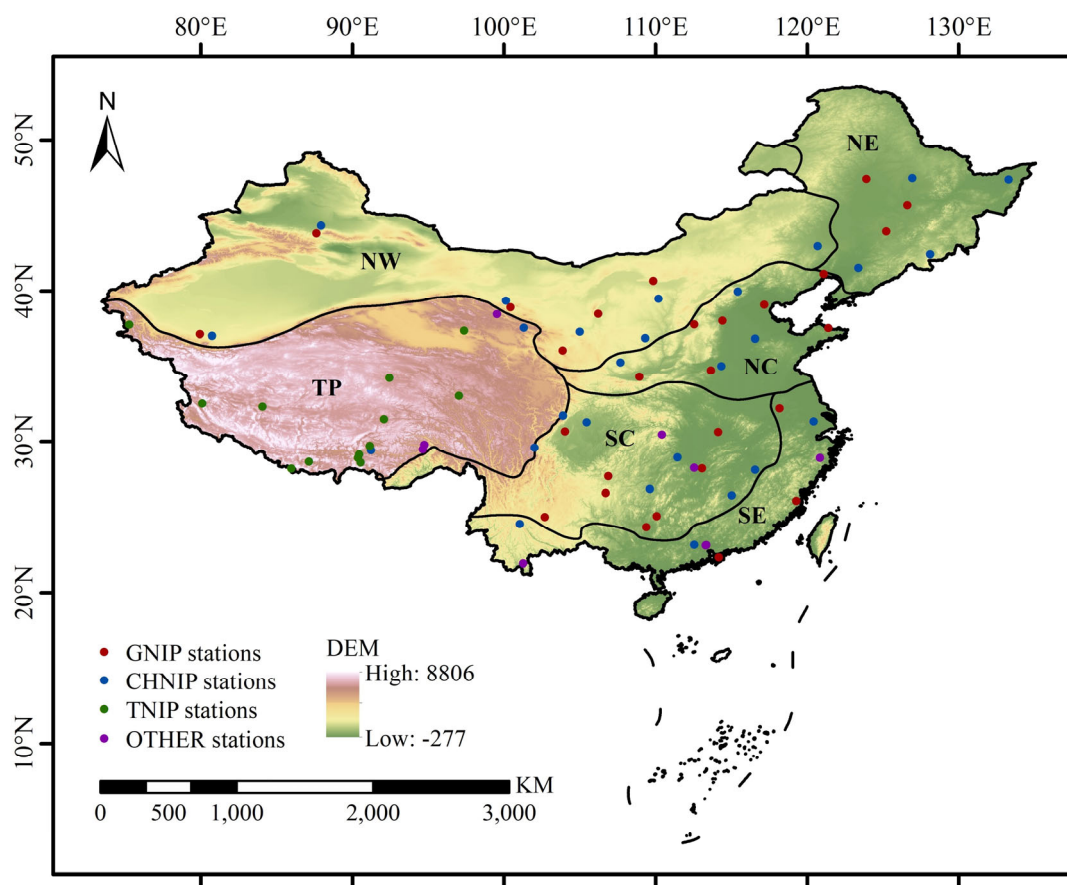

Figure 1. Map of the station locations and topography in mainland China. The dots indicate the distribution of isotope observation stations, with different colours representing different sources. The six sub-regions are plotted (NE – Northeast China, NC – North China, SC – South China, SE – Southeast China, TP – Tibetan Plateau, NW – Northwest China).

2 Study area and data

2.1 Study area

China is located in the east of Eurasia and on the west coast of the Pacific Ocean. The topography of China generally presents three steps descending to the east. The climate is complex and diverse in China. Heavily influenced by the continents and oceans, the monsoon climate is significant, especially for the east of China. The spatiotemporal variation of precipitation stable isotopes are very complex due to the significant changes in winter and summer circulation (Liu et al., 2014). Mainland China can be geographically classified into three sub-regions, the eastern monsoon region, the arid northwest region (NW) and the Qinghai-Tibet Plateau region (TP), according to topography, climate, soil and vegetation. According to temperature and precipitation, the eastern monsoon region is further divided into two sub-regions: north (NE and NC) and south (SC and SE). In addition, the southern part of the eastern monsoon region is further divided into southeast coastal region and inland region, considering the characteristics of observed isotope data. To sum up, the study area is divided into six sub-regions for our analysis: Northeast China (NE), North China (NC), South China (SC), Southeast China (SE), Qinghai-Tibet Plateau (TP) and Northwest China (NW), as shown in Fig. 1.

2.2 Datasets

There are 78 oxygen isotope observation stations in the study area (Fig. 1), including 29 GNIP stations (available at https://nucleus.iaea.org/wiser), 27 CHNIP stations (Liu et al., 2014), 13 TNIP stations (Yao et al., 2013) and 9 stations from other sources (mainly from references). Monthly oxygen isotope composition of precipitation ($\delta^{18}O_p$) is used for analysis. The time span of GNIP data mostly ranges from 1980 to 2000, and the length of the time period is basically 5-15 years. For CHNIP, most time periods are about 2-5 years ranging from 2005 to 2010. Most TNIP data were between 1995 and 2005, with varying lengths.

Eight $\delta^{18}O_p$ spatio-temporal fields simulated by five iGCMs (CAM2, GISS E, HadAM3, LMDZ4 and MIROC32) are used, which are selected from the SWING2 archive (available at https://data.giss.nasa.gov/swing2/). Six of eight simulations are free-running, forced only by observed sea surface conditions. The remaining two (GISS E and LMDZ4) were nudged to constrains of large-scale atmosphere circulation, so that the dynamical fields in simulations are close to the observations (Yoshimura et al., 2008). In addition to SWING2 simulations, a zoomed simulation by LMDZ4, with the horizontal resolution of 50-60km (Risi et al., 2010; Gao et al., 2011) and nudged by reanalyses, is also used. Detailed information about these iGCMs can be found in Table 1.

Table 1. Time periods and basic outputs information of selected iGCMs.

GCM	Simulation method	Horizontal resolution (longitude × latitude)	Time period	Key references
CAM2	Free-running	2.8° × 2.8°	1958-2003	Lee et al. (2007)
GISS E	Free-running and	2.5° × 2°	1969-2009	Schmidt et al. (2007)

GCM	Simulation method	Horizontal resolution (longitude × latitude)	Time period	Key references
HadAM3	nudged by NCEP Free-running	3.75° × 2.5°	1870-2001	Tindall et al. (2009)
LMDZ4	Free-running and nudged by ECMWF	3.75°×2.5°	1979-2007	Risi et al. (2010)
	Zoomed (nudged by ECMWF)	50-60 km	1979-2017	Gao et al. (2011)
MIROC32	Free-running	2.8°×2.8°	1979-2007	Kurita et al. (2011)

The time span of the isoscape built in this study covers the union set of all simulations ranging from 1870 to 2017. Since the temporal lengths of eight iGCM are not identical, the number of iGCM simulations used to build the isoscape varies. Specifically, for 1979-2001, a total of eight simulations from all five iGCMs were used; for 2002-2007, six simulations from three iGCMs (GISS E, LMDZ4 and MIROC32) were used; and for 1969-1978, four simulations from three iGCMs (CAM2, GISS E and HadAM3) were used. For the remaining periods, there is only one simulation or two: for 1958-1968, CAM2 and HadAM3 were used; for 1870-1957, HadAM3 was used; and for 2008-2017, LMDZ4 zoomed was used.

The stable isotope composition of precipitation is expressed in the relative permillage (‰) derived from the standard sample (Clark and Fritz, 1997) as:

$$\delta = \left(\frac{R_{\text{sample}}}{R_{\text{V-SMOW}}} - 1 \right) \times 1000‰ \tag{1}$$

where R is the ratio of heavier isotope to common isotope ($^{18}O/^{16}O$), and the subscripts sample and V-SMOW represent standard sample and Vienna Standard Mean Ocean Water, respectively.

3 Methods

3.1 Bias correction methods

BCMs aim to correct the mean, variance and/or quantile of the climate model time series, so that the corrected model time series can better match those of the observations (Maraun, 2013). In this study, two typical methods (i.e. linear scaling (LS) and distribution translation (DT)) were used to correct the bias of iGCM at the monthly timescale. These two BCMs can be classified into mean-based scaling (i.e. LS) and distribution-based correction (i.e. DT) approaches (Chen et al., 2013). The mean-based scaling uses a constant correction factor for the entire time series, while the distribution-based approach uses correction factors that vary with the quantiles of the distribution (Chen et al., 2016).

The LS method is the simplest bias correction method. The differences between observations and raw iGCM simulations were applied to simulations to obtain the bias-corrected isotope time series for each season and sub-region. Specifically, for a particular sub-region, the differences (defined as correction factors) in mean values between observed and simulated isotopes

Earth System
Science
Data

were first calculated at the seasonal basis using Eq. (2). The calculated correction factors were then applied to simulated isotopes for the entire period using Eq. (3).

$$R_{\text{LS}} = \overline{\delta O}_{\text{obs,s,sr}}^{\text{ref}} - \overline{\delta O}_{\text{raw,s,sr}}^{\text{ref}} \tag{2}$$

$$\delta O_{\text{cor,s,sr}} = \delta O_{\text{raw,s,sr}} + R_{\text{LS}} \tag{3}$$

135 where R_{LS} is the correction factor; the superscript ref represents the reference period; the subscripts obs, raw and cor represent observations, raw simulations and corrected simulations; and s and sr represent a specific season and a sub-region.

The implementation of the DT method is similar to the LS method. However, the differences (i.e. correction factors) between observed and simulated isotopes were calculated for each of 100 integral percentiles as shown in Eq. (4) – (6), to represent the distribution for each season in each sub-region. The correction factors of grid points were obtained by

140 interpolating or extrapolating the factors of observation stations using Eq. (5).

$$R_{\text{DT}}^{\text{ref}} = \delta O_{\text{obs,q}}^{\text{ref}} - \delta O_{\text{raw,q}}^{\text{ref}} \tag{4}$$

$$R_{\text{DT}}^{\text{ref}} \xrightarrow{\text{Interpolation/Extrapolation}} R_{\text{DT}} \tag{5}$$

$$\delta O_{\text{cor,s,sr}} = \delta O_{\text{raw,s,sr}} + R_{\text{DT}} \tag{6}$$

where the subscript q is a percentile for a specific season in a sub-region. Other superscripts and subscripts are the same as

145 Eq. (2) and (3).

3.2 Neural network data fusion

The neural network is a kind of mathematical model, which imitate the behaviour characteristics of animal neural network and carry out distributed parallel computing (Rumelhart et al., 1994; Krenker et al., 2011). Performing calculations and spreading information through large numbers of interconnected neurons, neural networks are often used to describe complex

150 relationships between inputs and outputs, or to explore the internal structure and patterns of data (Hsu et al., 1995; French et al., 1992). In this study, Back Propagation Neural Network (BP), Long Short-Term Memory (LSTM) Neural Network and Convolutional Neural Network (CNN) are adopted for data fusion, considering BP's simplicity and practicality, LSTM's advantages in time series prediction and CNN's outstanding performance in various fields. The structures of these three neural network DFMs are presented in Fig. 2.

155 BP, first proposed by Rumelhart et al. (1986), is a multilayer feed-forward network trained by the error back-propagation algorithm. BP is one of the most widely used neural network models with high simulation accuracy for nonlinear functions. The main characteristic of BP is that the input signal is processed layer by layer from the input layer to the hidden layer and then to the output layer, and each neuron carries out the weighted sum of the input signal through the activation function. If the error between the actual output and the expected output is larger than the set value, the weight and bias of the network

160 are continuously corrected by the backpropagation to minimize the loss function. The isotope simulations of iGCMs were selected as the input of BP and the observations as the expected output to calculate the loss function. The input layer was the

corresponding input parameter, and the output layer was the fusion isotope value. The hidden layer had two layers with 16 and 4 neurons respectively. The activation function was the Rectified Linear Unit (ReLU). The Stochastic Gradient Descent (SGD) was used to optimize the loss function iteratively. Mean Square Error (MSE) was chosen as the loss function. The

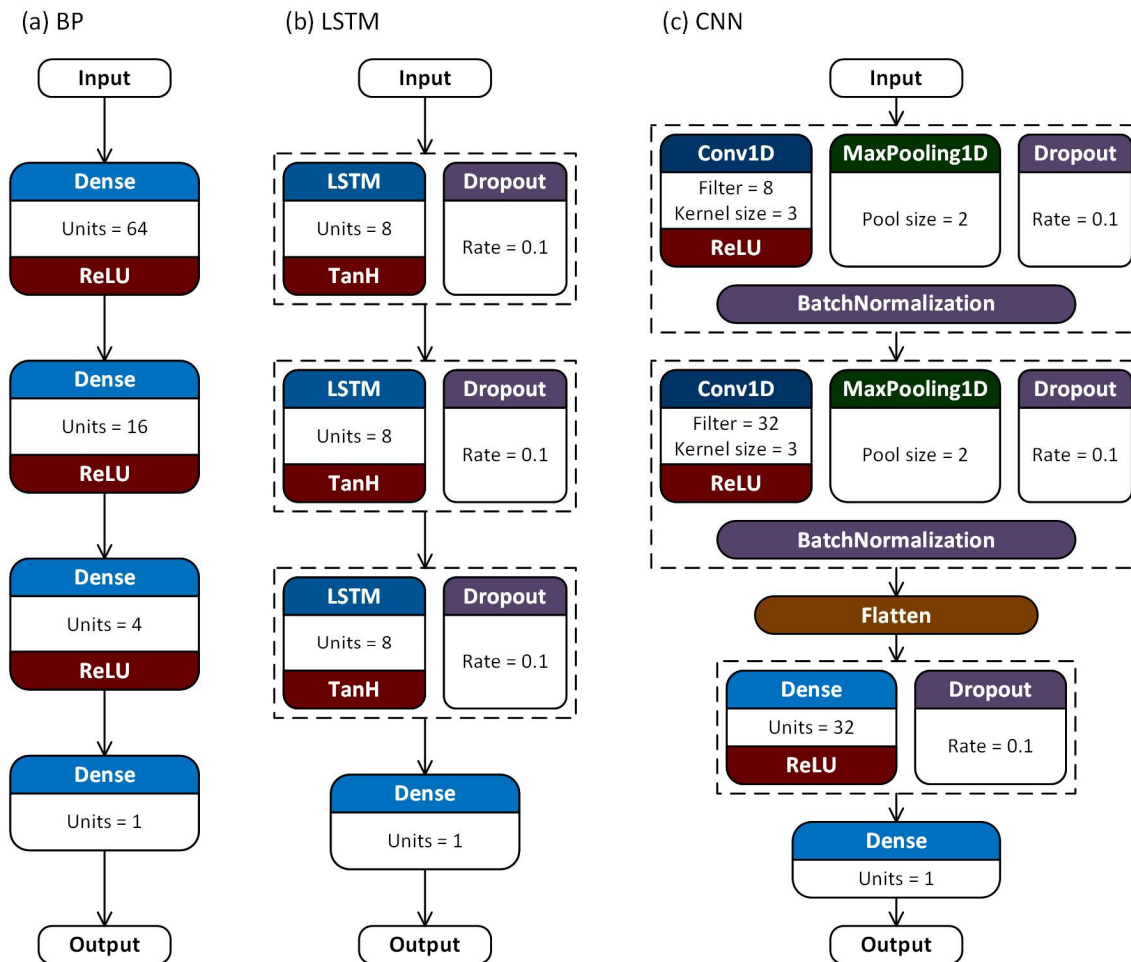165   maximum number of iterations was 1000, and the training stopped if the loss function did not decrease within 150 epochs.

Figure 2. Network structure of BP, LSTM and CNN fusion method.

LSTM is very efficient for sequential data and is derived from Recurrent Neural Network (RNN) with memory function. RNN has a sequential feed-forward connection, so that the information of the past moment can affect the output of the

170   present moment (Zhang et al., 2018). The traditional RNN has the problems of vanishing gradient and exploding gradient (Hochreiter, 1998). To solve these problems, the LSTM Neural Network was proposed (Hochreiter and Schmidhuber, 1997). A basic LSTM neuron usually consists of a memory cell and three gates (i.e. input gate, forget gate and output gate). Memory cells are used to store past information, realizing long-distance dependent learning of sequence features. The input gate determines which inputs are saved to the cell; forget gate determines what information is retained from the previous

175　moment; output gate determines what information needs to be output. In this study, a fully connected layer was added after three LSTM layers to generate fusion results. The number of neurons in the LSTM layer was set to 8. Dropout layers were applied to the three LSTM layers of the network with the probability of 0.1 to make the model more robust. The activation function used the most common setting in LSTM, the Hyperbolic Tangent (TanH) function. The optimizer of the model was set as the Adaptive Moment Estimation (Adam) algorithm, which performs optimally. Similar to BP, MSE was chosen as the

180　loss function. The maximum number of iterations was 1000 with the patience of 100 epochs.

　　CNN was first proposed by Lecun (1989), for the problem of handwritten digit recognition. CNN combines three advantages of local connectivity, weight sharing and pooling. On one hand, it reduces the number of weights, making the network easy to optimize. On the other hand, it reduces the complexity of the model and alleviates the overfitting problem. It is one of the most widely used neural networks with the best performance. The convolutional layer and pooling layer of the

185　hidden layer are the core modules of CNN. The function of the convolutional layer is to extract features of the input data by convolutional kernels. The pooling layer performs feature selection and information filtering on the feature map output by the convolution layer. In this study, CNN was mainly composed of two convolutional layers and pooling layers. The two convolutional layers had 8 and 32 convolutional kernels, respectively, and the size of the convolution kernels was 3. The pooling layers used the max pooling method, and the pooling size was 2. Two fully connected layers were added at the end

190　with 32 and 1 neurons, respectively, to remove the spatial topology and output the results. The convolutional layer and the fully connected layer were connected by flattening the output of the convolutional layer through the flatten layer. Batch normalization layers were inserted into the model to improve the speed, performance and stability of the neural network. Dropout layers were also applied with the probability of 0.1. The activation function was the ReLU. The optimizer of the model was the Adam, with the loss function of MSE. The maximum number of iterations was 1000 with the patience of 100

195　epochs.

3.3 Cross-validation experiments

In order to make full use of the data and reduce the variation of model accuracy caused by the difference between the training set and the test set, K-fold cross-validation was adopted. In the K-fold (K=5 in this study) experiment, the data set was randomly divided into K groups, and one of them was used as the test set each time, leaving K-1 groups as the training

200　set. To fully consider the variations of random division, K-fold cross-validation was repeated 100 times.

3.4 Generation of isoscape

The performance of BCMs and DFMs were evaluated by using correlation coefficient (CC) and root mean square error (RMSE) as metrics for the common period of 1969-2007. On the basis of the performance comparison, the optimal method was selected for each period to generate the oxygen isoscape. For the common period (1969-2007), the isoscape was

205　generated using the best-performing methods among BCMs or DFMs; while for the rest periods, only the BCM was used since there is only one or two simulations. When using the BCM, the ensemble mean of all iGCM simulations and two

BCMs were used. To sum up, the generated isoscape presents monthly $\delta^{18}O_p$ on a 50-60km spatial resolution over mainland China for the 1870-2017 period.

4 Results

4.1 Evaluation of bias correction and data fusion methods

Prior to applying BCMs and DFMs to build the isoscape, the performance of iGCM simulations was evaluated by comparing gauged observations for the 1969-2007 period. Fig. 3 shows the cumulative distribution functions (CDFs) of $\delta^{18}O_p$ for observations and iGCMs simulations in each sub-region. Generally, the CDFs of observed $\delta^{18}O_p$ can be well represented by iGCM simulations for each sub-region, as the observed CDFs distribute in the centre of simulated ones. For specific regions, the envelope of CDFs is the narrowest for NE and SE, indicating that iGCMs perform consistently better for these two regions. For NC, SC and NW, the CDFs of LMDZ4 (free and nudged) simulated $\delta^{18}O_p$ are relatively close to the observations, while other iGCM simulations generally overestimate the $\delta^{18}O_p$. For TP, the differences between CDFs of observed and simulated $\delta^{18}O_p$ are the largest, indicating the large variability of $\delta^{18}O_p$ simulations. This is expected, as climate models generally perform worse for TP than other regions (Zhu and Yang, 2020; Su et al., 2013).
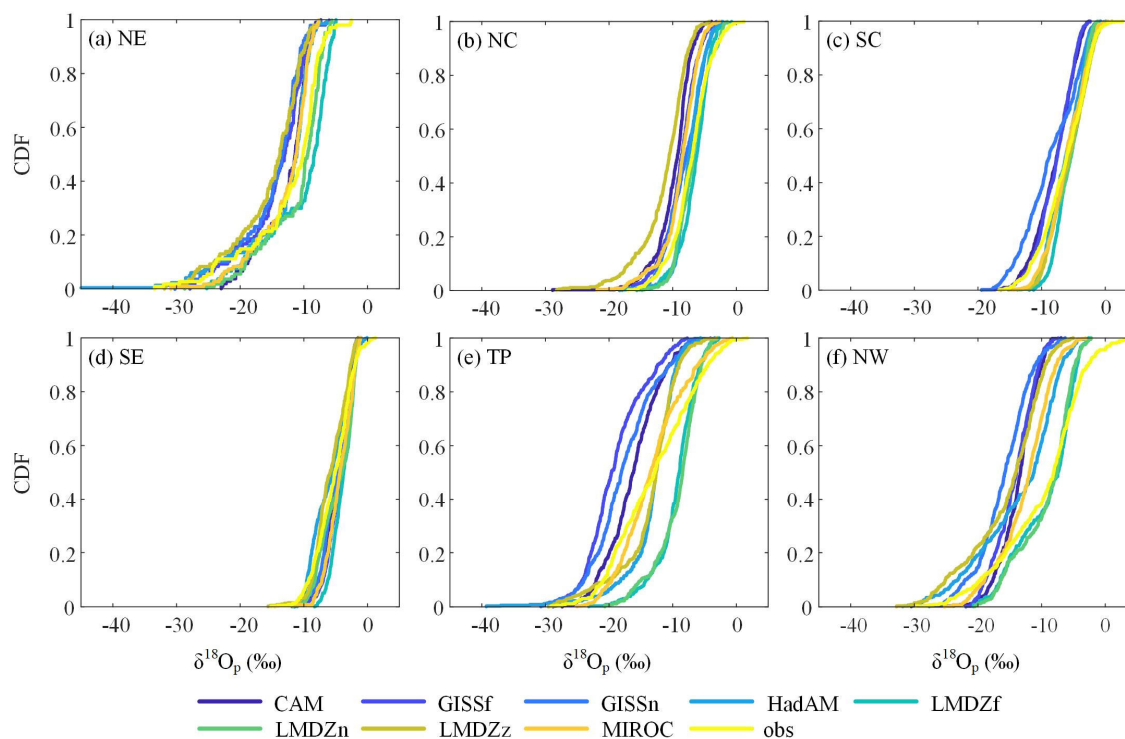
Figure 3. Cumulative distribution functions of $\delta^{18}O_p$ for eight iGCM simulations in six sub-regions.

Root mean square error (RMSE) and correlation coefficient (CC) were also calculated to evaluate the accuracy of iGCM simulated $\delta^{18}O_p$. Fig. 4 presents the RMSE and CC for raw iGCM simulations, bias corrected and fused simulations for six sub-regions over the validation periods (1979-2001). Generally, DFMs perform better than BCMs, and both perform better

225 than raw iGCM simulations. In addition, all the simulations are correlated with the observations with CC ranging between 0.12 and 0.95.
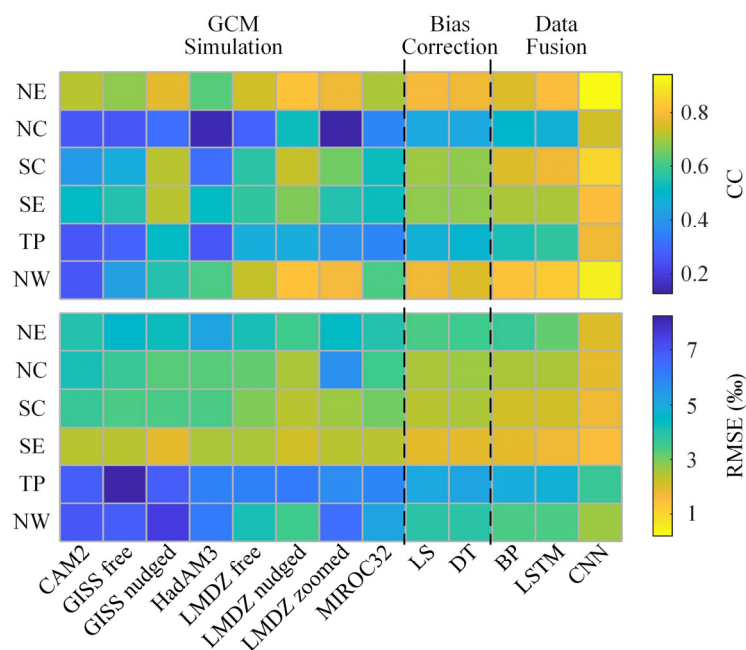
Figure 4. Average correlation coefficient (CC) and root mean square error (RMSE) metrics of raw, bias-corrected and fused $\delta^{18}O_p$ in six sub-regions.

230 The CC and RMSE vary considerably for raw iGCM simulations, with CC ranging from 0.12 to 0.83 and RMSE ranging from 2.0‰ to 9.1‰. The simulations of nudged GISS E and LMDZ4 have the strongest correlation with the observations, and their CC is basically above 0.5, ranging from 0.30-0.83. The error of nudged LMDZ4 is the smallest, and its RMSE ranges between 2.2‰ and 6.8‰.

Generally, the performance of LS and DT is similar, even though the LS method performs slightly better than the DT

235 method. The CCs range from 0.40 to 0.88 for the LS and DT corrected simulations, with an average increase from 0.52 to 0.64 relative to raw simulations; the RMSEs are between 2.0‰ and 5.7‰, with an average decrease of 23.9%.

For DFMs, BP and LSTM show similar performance, while CNN performs the best. The CC of CNN-generated simulations is all greater than 0.74, increasing from 0.52 to 0.84 on average compared with the raw simulations. The RMSE of CNN-generated simulations is all smaller than 4.0‰, showing 48% reduction relative to the raw simulation on average.

Earth System
Science
Data

240   Fig. 4 shows that CC and RMSE of simulations in different sub-regions are quite different. For all raw, fused, and bias corrected simulations, CC is smaller for NC and TP while RMSE is larger for TP and NW than other sub-regions. For DFMs, especially CNN, the differences of CC and RMSE between different regions are smaller.

Generally, all simulations perform worse in NC, TP and NW than in other sub-regions. The poor simulation performance in NC may be due to complex air mass movements (Yang et al., 2016; Peng et al., 2020a), which are difficult to be
245   accurately simulated by iGCMs. The poor simulation performance in TP and NW may be due to the fact that GCMs cannot accurately describe the atmospheric physical process and simulate precipitation and other meteorological factors in these regions (Miao et al., 2012; Jiang et al., 2016; Chen and Frauenfeld, 2014).

The performance of BCMs and DFMs is also evaluated for three periods (1969-1978, 1979-2001 and 2002-2007) and six sub-regions at a seasonal basis. The seasonal average CC and RMSE for two BCMs and three neural network DFMs are
250   presented in Figs. 5-6. Generally, all BCMs and DFMs perform very similar for all three periods.

As for CC, the simulations in the northern region (NE, NC, and NW) show strong correlation with observations in spring, while those in the southern region (SC, SE) show strong correlation in summer and autumn. The correlation of CNN fusion simulations is significantly higher than that of the other methods, with CC being mostly above 0.8. The correlation of LS and DT corrected simulations is similar to that of BP and LSTM fusion simulations, with CC being mostly between 0.3 and 0.6,
255   varying with sub-regions and seasons. The BP and LSTM fusion methods perform slightly worse in NE, but better in TP and NW, compared with BCMs.

As for RMSE, the northern region has a small error of about 2‰ in summer, except for NW with a larger error of 3‰; while the southern region shows little seasonal difference in error, only RMSE in SE being slightly larger in autumn and winter, mostly ranging between 1‰ and 3‰. On the whole, the DFMs perform better than the two BCMs. The errors of
260   CNN simulations are the smallest in all regions and seasons, which are smaller than 4‰ in TP and NW and mostly smaller than 2‰ in other sub-regions. The errors of BP and LSTM simulations are slightly smaller than those of LS and DT simulations.

From the error bars, the LS and DT methods show smaller uncertainties in CC and RMSE than the BP, LSTM and CNN fusion methods. In addition to the uncertainties brought by cross-validation, the simulations of DFMs also show uncertainties
265   brought by neural network itself. Generally, CNN fusion simulations show smaller uncertainties than the other two fusion methods.

The above results show that the CNN fusion method consistently performs better than the other methods. To further confirm this conclusion, scatter plots of fused and corrected against observed $\delta^{18}O_p$ are presented in Fig. 7 for the 1979-2001 period. The overall mean CC and RMSE corresponding to the figure are shown in Table 2. It can be seen that the CNN
270   fusion method shows a stronger correlation with the observations than the other fusion methods and BCMs. The CNN fused $\delta^{18}O_p$ consistently shows the largest CC and smallest RMSE, showing a strong positive linear correlation with the observations with CC being almost all larger than 0.85.
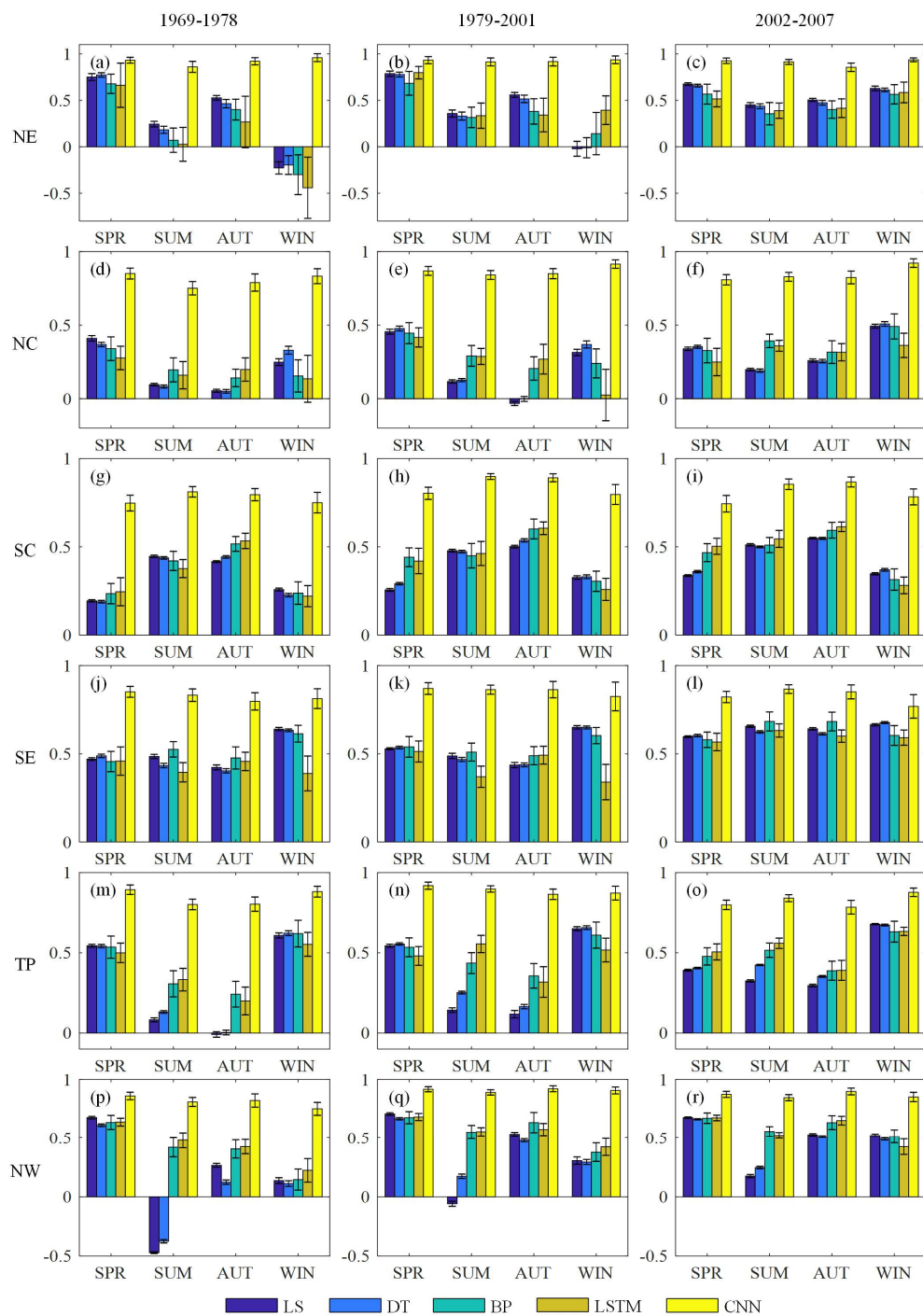
275 **Figure 5. Seasonal average results of correlation coefficient (CC) metrics for BCMs and DFMs in six sub-regions. The whiskers denote +/- one standard deviation.**
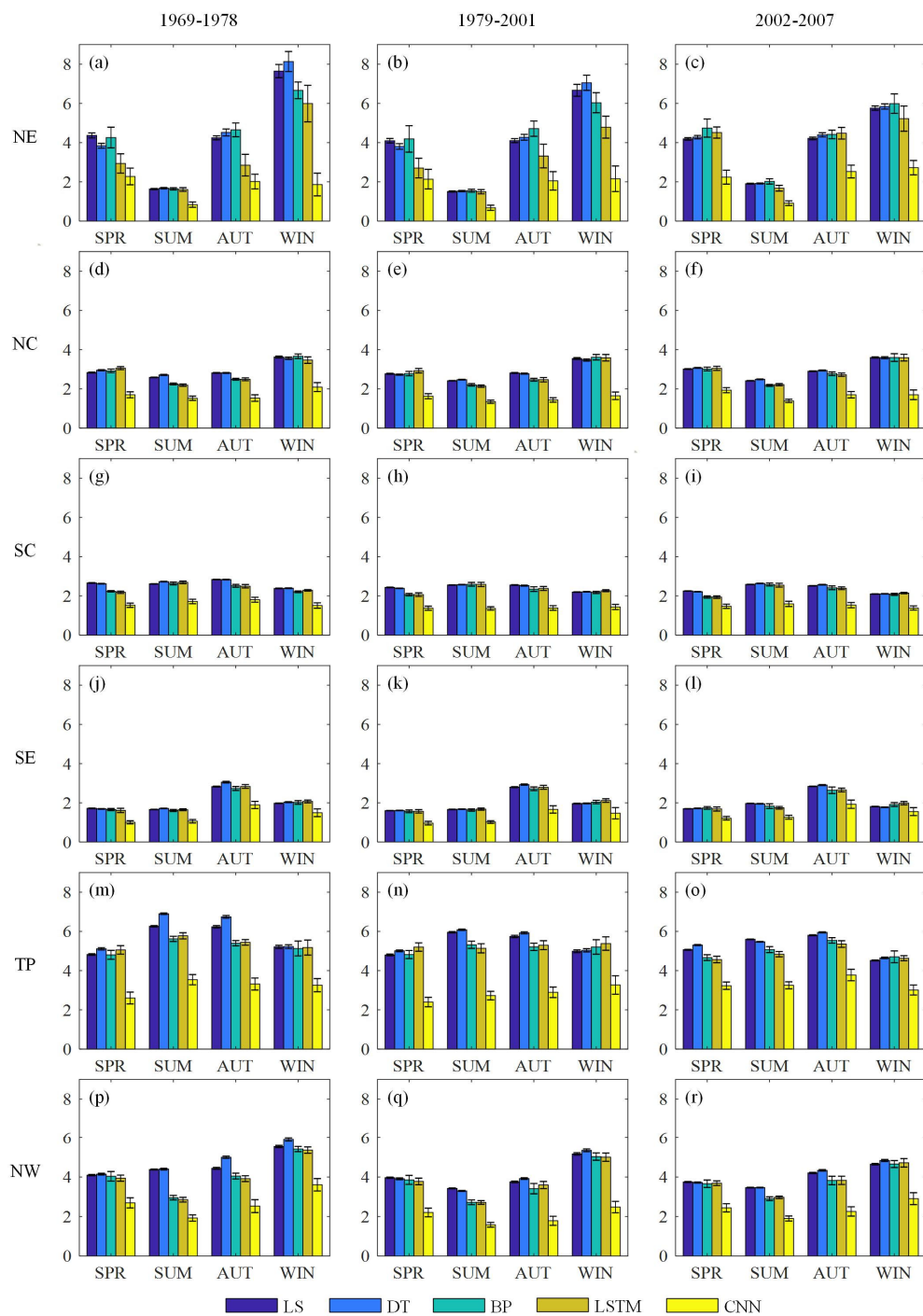
Figure 6. Seasonal average results of root mean square error (RMSE) metrics (‰) for BCMs and DFMs in six sub-regions. The whiskers denote +/- one standard deviation.
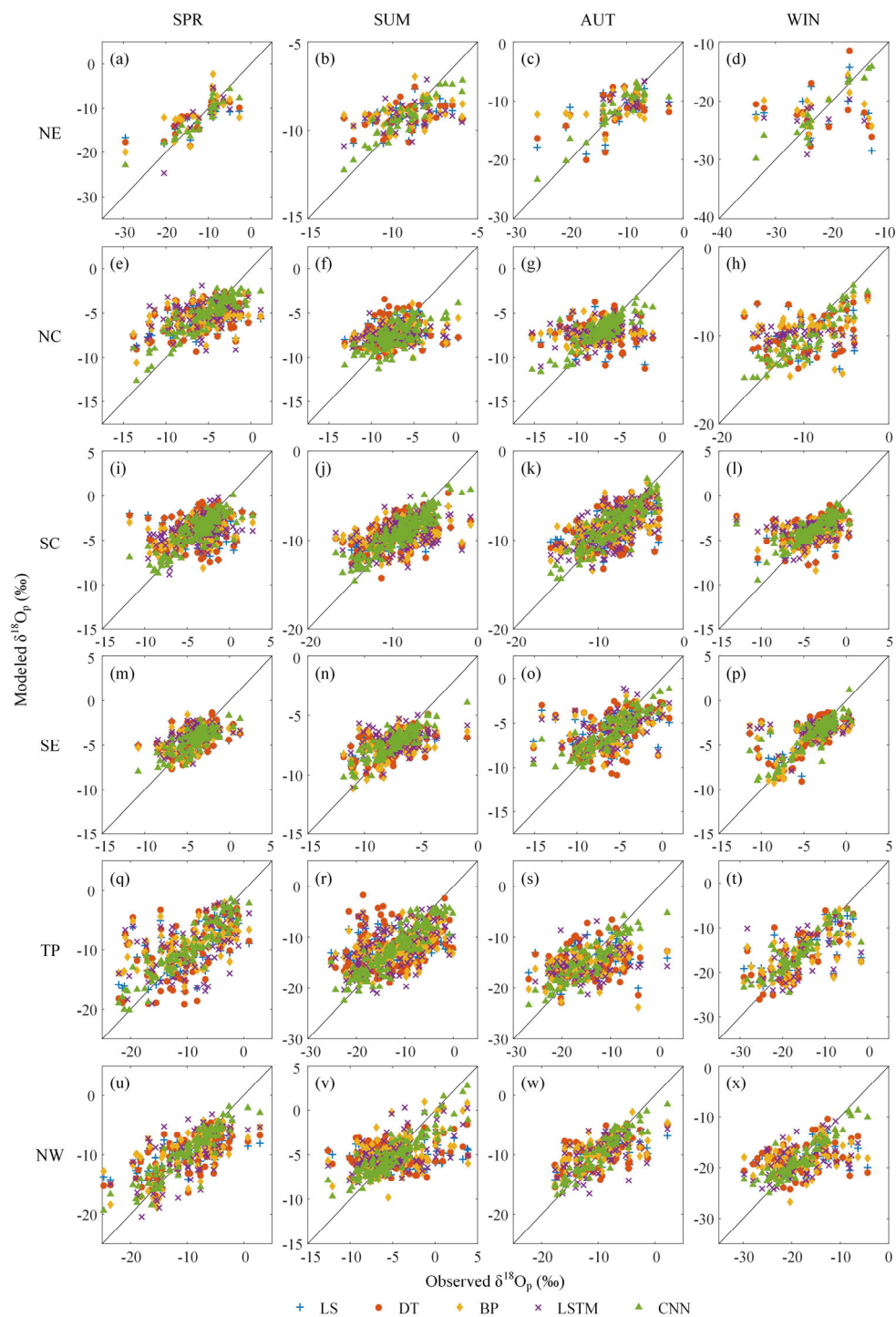
Figure 7. Scatter plots of seasonal δ¹⁸Oₚ from bias-corrected and fused output against observations in six sub-regions.

285   **Table 2. Correlation coefficient (CC) and root mean square error (RMSE, ‰) metrics corresponding to the Fig. 7.**

Sub-region	Season	LS		DT		BP		LSTM		CNN	
		CC	RMSE	CC	RMSE	CC	RMSE	CC	RMSE	CC	RMSE
NE	SPR	0.788	4.097	0.774	3.837	0.675	4.315	0.802	2.777	0.934	2.349
	SUM	0.362	1.508	0.324	1.537	0.310	1.552	0.329	1.498	0.915	0.730
	AUT	0.559	4.099	0.513	4.258	0.382	4.562	0.333	2.889	0.913	2.091
	WIN	-0.018	6.713	-0.018	6.985	0.145	6.073	0.373	4.658	0.936	2.203
NC	SPR	0.460	2.766	0.479	2.723	0.447	2.775	0.414	2.943	0.869	1.631
	SUM	0.111	2.420	0.123	2.481	0.300	2.204	0.284	2.150	0.843	1.396
	AUT	-0.029	2.797	-0.011	2.795	0.200	2.457	0.268	2.527	0.849	1.422
	WIN	0.311	3.582	0.368	3.477	0.240	3.728	0.086	3.406	0.908	1.679
SC	SPR	0.258	2.424	0.297	2.382	0.441	2.061	0.430	2.047	0.808	1.369
	SUM	0.478	2.556	0.474	2.571	0.457	2.583	0.467	2.568	0.895	1.409
	AUT	0.499	2.557	0.539	2.516	0.599	2.351	0.613	2.350	0.895	1.340
	WIN	0.326	2.192	0.330	2.215	0.304	2.191	0.250	2.292	0.795	1.457
SE	SPR	0.529	1.607	0.533	1.620	0.540	1.574	0.513	1.522	0.873	0.972
	SUM	0.490	1.663	0.463	1.685	0.526	1.618	0.376	1.677	0.869	1.030
	AUT	0.434	2.806	0.438	2.926	0.488	2.728	0.491	2.810	0.869	1.624
	WIN	0.654	1.956	0.649	1.964	0.606	2.039	0.339	2.114	0.824	1.460
TP	SPR	0.549	4.781	0.556	4.986	0.537	4.814	0.481	5.299	0.916	2.409
	SUM	0.145	5.951	0.252	6.099	0.434	5.341	0.551	5.205	0.900	2.691
	AUT	0.121	5.731	0.164	5.921	0.355	5.209	0.315	5.380	0.864	2.852
	WIN	0.649	4.992	0.659	5.024	0.616	5.180	0.526	5.294	0.877	3.305
NW	SPR	0.706	3.975	0.661	3.932	0.674	3.872	0.687	3.776	0.920	2.207
	SUM	-0.060	3.424	0.171	3.294	0.555	2.718	0.561	2.671	0.896	1.582
	AUT	0.536	3.746	0.481	3.908	0.643	3.440	0.574	3.553	0.929	1.735
	WIN	0.307	5.162	0.291	5.337	0.377	5.070	0.429	4.979	0.913	2.460

4.2 Spatial variability of precipitation oxygen isotope

Since the above section shows that the CNN method consistently performs better than other methods, it is used to generate precipitation oxygen isotopes for the 1969-2007 period with a spatial resolution of 50-60km on latitude and longitude. Fig. 8

290   presents the spatial variability of mean $\delta^{18}O_p$ for observed and fused data for the 1969-2007 period. Generally, the CNN fused $\delta^{18}O_p$ shows similar spatial distribution to observations for all four seasons.

In NE, $\delta^{18}O_p$ decreases with increasing latitude for all seasons, and its spatial variation is basically parallel to the latitude, which reflects the latitude effect (Rozanski et al., 1993). Indeed, most of the water vapour in the atmosphere is formed at low latitudes, and Rayleigh distillation continuously depletes the residual water vapour as air masses move toward higher

295   latitudes, thus depleting the $\delta^{18}O_p$ of the residual water vapour and thus of the rain forming in clouds.
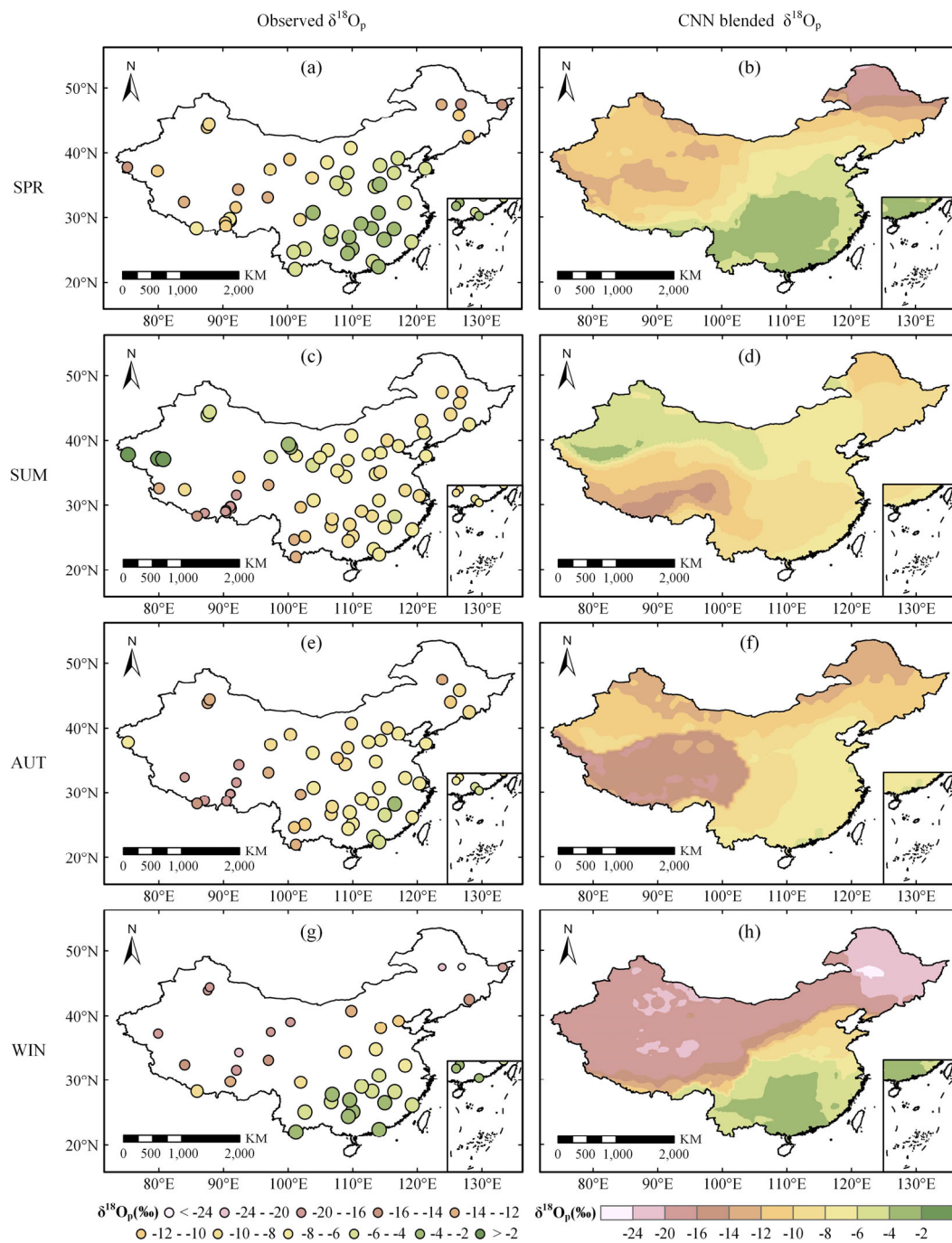
Figure 8. Seasonal averaged observations (a, c, e, and g) and CNN fused simulations (b, d, f, and h) of δ¹⁸Oₚ in the mainland China.

In SC and SE, $\delta^{18}O_p$ decreases from the southeast coast to inland. This phenomenon is consistent with the continental effect. As water vapour transfers from the ocean to the interior of the continent, precipitation is formed along the way. The

300   separation process of heavy isotopes takes place preferentially than that of light isotopes, which leads to the gradual dilution of heavy isotopes in the cloud, and thus makes the proportion of heavy isotopes in the subsequent precipitation lower.

The $\delta^{18}O_p$ in TP is low in general except for the southeast corner, which is mainly due to the special effect of large landforms (Zhao et al., 2012). The low $\delta^{18}O_p$ in TP is mainly due to its high altitude, with an average altitude being above 4,000m. The moisture in the air mass is gradually removed during the orographic uplift, with heavy isotopes preferred to be

305   removed during the condensation process, which leads to the dilution of heavy isotopes in water vapour (Rozanski et al., 1993). Higher $\delta^{18}O_p$ in the southeast corner of TP indicates closer vapour sources such as the Bay of Bengal and the Arabian Sea in the Indian Ocean. Due to the terrain barrier of the Himalayas, most of the water vapour can only pass through its southeast corner, along the valley of the major rivers (Nujiang River, Jinsha River, etc.) into the plateau, or through the Yarlung Zangbo River valley into the plateau (Araguás-Araguás et al., 1998).

310   The $\delta^{18}O_p$ in NW is lower than that in the southern region, but higher than that in NE and TP. This is because NW is far away from the ocean and has a dry climate, so the amount of heavy isotope in water vapour from the ocean is limited. However, a large part of water vapour to generate precipitation in NW comes from terrestrial evaporation (Li et al., 2016). The $\delta^{18}O_p$ in surface water in the arid area is high, resulting in high $\delta^{18}O_p$ in evaporation water vapour and heavy isotope enrichment in precipitation. Another process is re-evaporation of raindrop in arid climate, enriching heavy isotope in

315   precipitation water. Under the joint control of both, the $\delta^{18}O_p$ in this sub-region varies greatly. In the southern part of the Taklimakan Desert, Xinjiang, $\delta^{18}O_p$ is obviously higher. This is because, the Taklimakan Desert is located in the heart of Eurasia, and it is surrounded by high mountains and has extremely low rainfall. In the southern part of the desert, there is more precipitation in the Kashi-Hotan line, and the water vapour mainly comes from the evaporation of local lakes and rivers, so the ratio of isotopes in precipitation is high.

320   The seasonality of $\delta^{18}O_p$ varies in sub-regions and is influenced by various factors. For NE and NC, $\delta^{18}O_p$ is lower in winter than in other seasons. NE and NC are influenced by westerly wind and polar continental air mass constantly, with no convergence or strong convection with isotope-depleted air mass. Compared with Pacific air mass, westerly wind and polar air mass are drier and have higher $\delta^{18}O_p$ (Tang et al., 2017). The seasonal distribution pattern of $\delta^{18}O_p$ in NE and NC is consistent with the temperature effect (Dansgaard, 1964). Although the amount effect is not significant at the annual scale in

325   these regions, it cannot be ignored in the wet season (Yamanaka et al., 2007). In particular, the maximum value of $\delta^{18}O_p$ is observed in spring for NC, when a large part of the precipitation vapour comes from local re-evaporation. The temperature effect is also reflected in NW. Due to the long-term influence of continental air mass, the temperature difference between winter and summer is large, and the $\delta^{18}O_p$ changes synchronously with temperature in these two seasons.

For SC and SE, $\delta^{18}O_p$ is lower in summer than in other seasons. The climate features of SC and SE are related to the deep

330   convection driven by the East Asian monsoon (Vuille et al., 2005), which brings water vapour from the Pacific Ocean to eastern China and dominates the sub-regions in summer. Air masses from the Pacific Ocean are more isotopically depleted than those from SC and SE (Zhao et al., 2012; Peng et al., 2020a), so convergence with the Pacific depleted air masses will dilute the isotopic content of precipitation in SC and SE. Therefore, although temperature in summer is generally higher,

depletion of $\delta^{18}O_p$ is usually larger during monsoon season than winter season. The effect of surface temperature on isotopic

335   fractionation during precipitation is masked by the effect of precipitation amount (Araguás-Araguás et al., 1998). The temporal distribution pattern of $\delta^{18}O_p$ in SC and SE is influenced by heavy monsoon precipitation and follows the amount effect (Rozanski et al., 1993; Dansgaard, 1964).

     In TP, $\delta^{18}O_p$ is positively correlated with temperature in the non-monsoon region (northern part of the plateau), with high $\delta^{18}O_p$ in summer and low in winter, reflecting the temperature effect. For the monsoon region (southern part of the plateau),

340    $\delta^{18}O_p$ is high in winter and spring and low in summer and autumn, which is obviously influenced by marine air mass and shows obvious amount effect. These results are similar to previous studies (Yu et al., 2021; Yao et al., 2013).

4.3 Temporal variability of precipitation oxygen isotope

As mentioned earlier, the $\delta^{18}O_p$ were generated by a combination of bias correction and data fusion method for the 1870-2017 period. The CNN fusion was used for 1969-2007, and BCMs (mean of LS and DT) was used for the rest of the period.

345   Fig. 9 shows the monthly time series of generated $\delta^{18}O_p$ and their 12-month moving averages for eight sub-regions over the 1870-2017 period. TP is divided into monsoon and non-monsoon regions, according to the research of Yu et al. (2021). In our study, the region with significant correlation between $\delta^{18}O_p$ and temperature is the non-monsoon region, while the rest is the monsoon region. The Mann-Kendall tests show that the $\delta^{18}O_p$ significantly increased in NE and NC for recent 40 years at the P=0.01 level. These two regions are consistent with the temperature effect, and it can be inferred that the temperature of

350   these sub-regions has a rising trend during this period. This has been proved in many studies. For example, the studies of Ren et al. (2012) and Ding et al. (2007) have shown that the temperature in China has a rising trend in recent years, especially in northeast, north and northwest China. A slight upward trend is also observed in NW from the 1920s to the 1970s and 1980s to 2000s. The temperature effect is more significant in inland areas at middle and high latitudes. In winter, NW is mainly controlled by the westerlies, and the amount effect can be ignored (Yang et al., 2011; Yang et al., 2017).

355   Therefore, the temperature effect in NW is more significant in winter. From 2000 to 2012, the $\delta^{18}O_p$ values in NW showed a decreasing trend, mainly in winter. Some studies have shown that the temperature in NW during this period is consistent with the global land warming hiatus phenomenon, and even shows obvious cooling, especially in winter (Wen et al., 2017; Ma et al., 2019).

     The $\delta^{18}O_p$ in SC presents a gentle upward trend in recent 70 years, indicating that precipitation has a downward trend in

360   this period, since the $\delta^{18}O_p$ conforms to the rule of amount effect in this region. While the $\delta^{18}O_p$ in SE shows a fluctuating trend of decline for the past 80 years (significant at 0.05 level), which indicates an upward trend in precipitation. These trends are consistent with the existing researches that precipitation increased in the east coast and northwest of China, but decreased in southwest, north and northeast (Zhai et al., 2005; Liang et al., 2016). There is no significant trend for $\delta^{18}O_p$ in TP. However, $\delta^{18}O_p$ shows a significant decreasing trend for the monsoon region of TP over the past 90 years, while it shows

365   a significant increasing trend for the non-monsoon region at the significant level of P=0.01. This is because the non-monsoon region shows temperature effect, while the monsoon region shows amount effect, which is consistent with the increasing

trend of temperature and precipitation in the Qinghai-Tibet region for recent years (Yu et al., 2021; Yao et al., 2013). Overall, the change trend of temperature and precipitation derived from the isotope effect analysis is consistent with that analysed directly using temperature and precipitation data in China. This further proved the reasonable performance of isoscape built
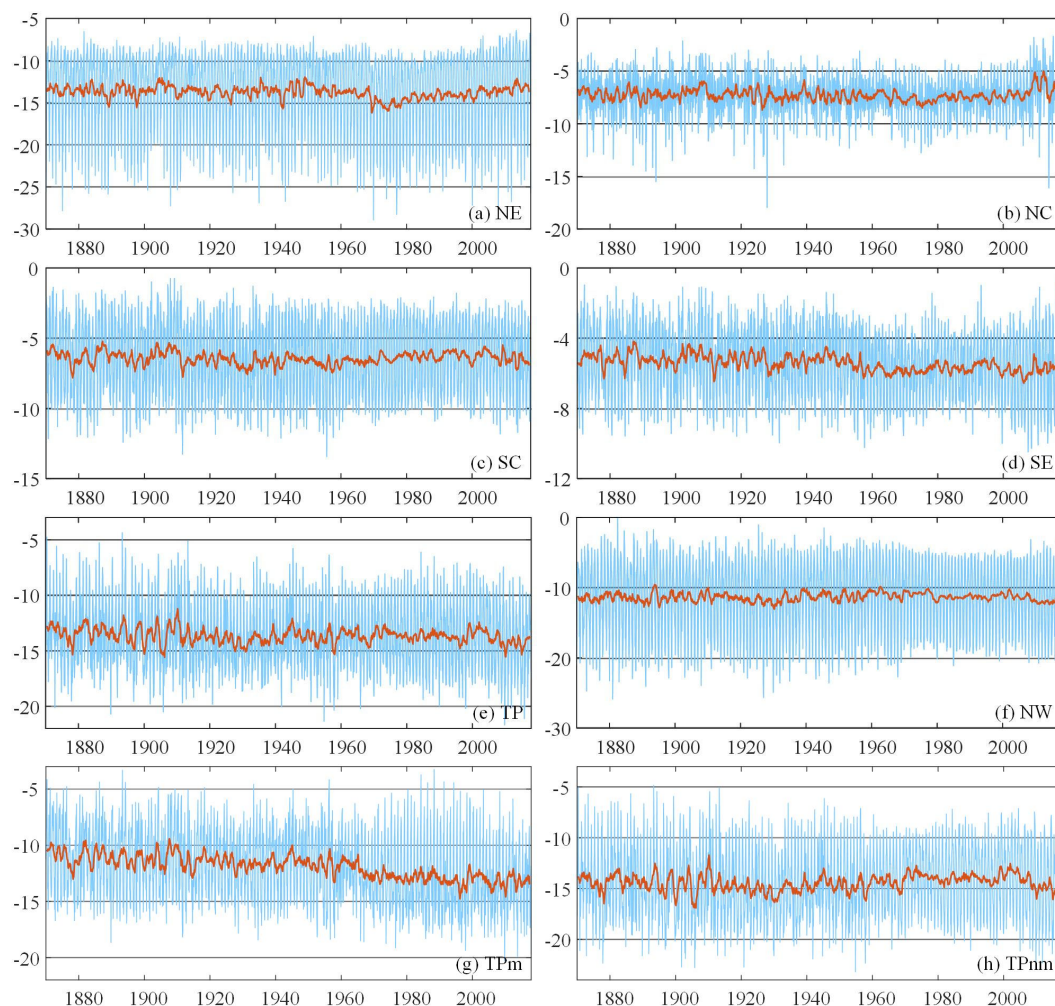
370 in this study.

Figure 9. Monthly time series of the generated $\delta^{18}O_p$ (‰) and their 12-month moving average in eight sub-regions from 1870 to 2017. (g) shows the monsoon region. (h) shows the non-monsoon region.

5 Data application

375 The dataset includes the stable oxygen isotope of precipitation for mainland China over the 1870-2017 period. Since the observation is limited and climate model simulations have different time durations, the entire duration was constructed by using the various methods and datasets for different periods as shown in Fig. 10.

Open Access — Earth System Science Data Discussions

HadAM3 LS & DT ensemble average	CAM2 & HadAM3 LS & DT ensemble average	CAM2, GISS E & HadAM3 CNN fusion	CAM2, GISS E, HadAM3, LMDZ4 & MIROC32 CNN fusion	GISS E, LMDZ4 & MIROC32 CNN fusion	LMDZ4 zoomed LS & DT ensemble average
1870-1957	1958-1968	1969-1978	1979-2001	2002-2007	2008-2017

Figure 10. The generation mode of dataset in each period.

380   Based on this new isoscape, the physical mechanisms driving the spatio-temporal variation of $\delta^{18}O_p$ can be deeply explored. This dataset is useful for tracing atmospheric and hydrological processes. For example, it can be used to study the effect of meteorological variables and air mass trajectory on stable isotope distribution, and quantify the source and fate of moisture (Tian et al., 2007; Peng et al., 2020b; Gao et al., 2011). The data can also help track the water cycle, to assess the source of water resources using complementary isotopic data from lakes, rivers and groundwater (Yao et al., 2013; Gao et al.,

385   2009). The generated dataset is also useful for paleoclimate studies, by studying the relationship between modern isotopes and meteorological variables (Yu et al., 2016b). When using this dataset, it should be noted that $\delta^{18}O_p$ for the 1979-2007 period generated using the CNN fusion method should be more reliable, because the performance comparison for BCMs and DFMs show that CNN fusion method performed better than BCMs, and more iGCM simulations were used for data fusion in this period that other periods.

390   **6 Data availability**

The generated dataset is available in Zenodo at https://doi.org/10.5281/zenodo.5703811 (Chen et al., 2021). The GNIP data can be obtained from the GNIP Database of IAEA/WMO (https://nucleus.iaea.org/wiser). The TNIP data was derived from the National Tibetan Plateau Data Centre (DOI: 10.11888/Geogra.tpdc.270940). The detailed information and sources of observations from other sources are shown in Table 3. The iGCM data can be downloaded from SWING2 (accessible at

395   https://data.giss.nasa.gov/swing2/), where LMDZ4 zoomed data was provided by Dr. Camille Risi at Laboratoire de Météorologie Dynamique in France.

Table 3. Detailed information and sources of observations from previous publications in literatures.

Sub-region	Site	Latitude	Longitude	Period	Data source
SC	Heshang	30.45	110.42	2011-2018	Wang et al. (2020)
SC	Changsha	28.25	112.55	2010-2017	Zhou et al. (2019) (DOI: 10.17632/975k2wzw3p.1)
SE	Guangzhou	23.13	113.32	2007-2014	Yang et al. (2017) (DOI: 10.11888/AtmosPhys.tpe.249477.db.)
SE	Yongan	28.89	120.85	2014-2018	Hu et al. (2020) (DOI: 10.17632/vndfs3dpyn.1)
SE	Xishuangbanna	21.93	101.27	2002-2004	Liu et al. (2007)

Earth System
Science
Data

Open Access

Discussions

Sub-region	Site	Latitude	Longitude	Period	Data source
SE	Guangzhou	23.15	113.35	2007-2009	Xie et al. (2011)
TP	Lulang	29.77	94.73	2007-2014	Yang et al. (2017) (DOI: 10.11888/AtmosPhys.tpe.249477.db.)
TP	Nuxia	29.47	94.65	2009-2014	Yang et al. (2017) (DOI: 10.11888/AtmosPhys.tpe.249477.db.)
TP	Yeniugou	38.46	99.54	2008-2009	Zhao et al. (2011)

7 Conclusions

400   Long-time sequence of $\delta^{18}O_p$ are of great significance for hydrological and meteorological studies. In view of the lack of long and reliable $\delta^{18}O_p$ datasets in China, this study generates a new dataset by integrating multi-iGCM data to overcome the limitations of short duration and uneven distribution of observed data. This dataset contains monthly $\delta^{18}O_p$ over mainland China for the 1870-2017 period with a spatial resolution of 50-60 km. The dataset from 1969 to 2007 was generated by using the CNN fusion method, when the observed time series and multiple iGCM simulation are available. For other periods, it

405   was generated by bias correcting iGCMs simulations. Two BCMs (i.e. LS and DT) with similar performances were used to produce ensemble mean. Prior to building the isoscape, the performance of two BCMs (LS and DT) and three DFMs (BP, LSTM and CNN) were evaluated using RMSE and CC as criteria. The results showed that CNN fusion method consistently performed the best for all sub-regions in China, and BP and LSTM fusion methods performed slightly better than LS and DT (BCMs). The performance of LS and DT method was similar. In terms of spatial distribution and temporal variability of

410    $\delta^{18}O_p$, the generated data showed very similar spatial distributions to observations, and the temporal trend of $\delta^{18}O_p$ was consistent with the observed changes in precipitation and temperature for different regions in China. All these showed that the built isoscape is reliable and useful to extending the time and space of observations in China.

Author contributions

JieC and XJZ designed the framework of the research. JiaC collected data, performed the analyses, and wrote the manuscript.

415   PP and CR provided data and methodological support. JieC supervised the progress of research and provided resources and financial support. All authors contributed to the discussion and revised the submitted manuscript.

Competing interests

The authors declare that they have no conflict of interest.

References

430   Allen, S. T., Kirchner, J. W., and Goldsmith, G. R.: Predicting Spatial Patterns in Precipitation Isotope (δ2H and δ18O) Seasonality Using Sinusoidal Isoscapes, Geophysical Research Letters, 45, 4859-4868, https://doi.org/10.1029/2018gl077458, 2018.

Ansari, M. A., Noble, J., Deodhar, A., and Saravana Kumar, U.: Atmospheric factors controlling the stable isotopes (δ18O and δ2H) of the Indian summer monsoon precipitation in a drying region of Eastern India, Journal of Hydrology, 584,
435   124636, https://doi.org/10.1016/j.jhydrol.2020.124636, 2020.

Araguás-Araguás, L., Froehlich, K., and Rozanski, K.: Stable isotope composition of precipitation over southeast Asia, Journal of Geophysical Research: Atmospheres, 103, 28721-28742, https://doi.org/10.1029/98jd02582, 1998.

Che, Y., Zhang, M., Wang, S., Wang, J., Liu, Y., and Zhang, F.: Stable water isotopes of precipitation in China simulated by SWING2 models, Arabian Journal of Geosciences, 9, 732, https://doi.org/10.1007/s12517-016-2755-5, 2016.

440   Chen, J., Brissette, F. P., Chaumont, D., and Braun, M.: Finding appropriate bias correction methods in downscaling precipitation for hydrologic impact studies over North America, Water Resources Research, 49, 4187-4205, https://doi.org/10.1002/wrcr.20331, 2013.

Chen, J., St-Denis, B. G., Brissette, F. P., and Lucas-Picher, P.: Using Natural Variability as a Baseline to Evaluate the Performance of Bias Correction Methods in Hydrological Climate Change Impact Studies, Journal of Hydrometeorology,
445   17, 2155-2174, https://doi.org/10.1175/jhm-d-15-0099.1, 2016.

Chen, J., Chen, J., Zhang, X. J., Peng, P., and Risi, C.: Precipitation oxygen isoscape for mainland China from 1870 to 2017 generated based on data fusion and bias correction of iGCMs simulations [dataset], https://doi.org/10.5281/zenodo.5703811, 2021.

Chen, L. and Frauenfeld, O. W.: A comprehensive evaluation of precipitation simulations over China based on CMIP5
450   multimodel ensemble projections, Journal of Geophysical Research: Atmospheres, 119, 5767-5786,
        https://doi.org/10.1002/2013jd021190, 2014.

Clark, I. D. and Fritz, P.: Environmental Isotopes in Hydrogeology, CRC Press, https://doi.org/10.1201/9781482242911,
        1997.

Conroy, J. L., Cobb, K. M., and Noone, D.: Comparison of precipitation isotope variability across the tropical Pacific in
455   observations and SWING2 model simulations, Journal of Geophysical Research: Atmospheres, 118, 5867-5892,
        https://doi.org/10.1002/jgrd.50412, 2013.

Dansgaard, W.: Stable isotopes in precipitation, Tellus, 16, 436-468, https://doi.org/10.3402/tellusa.v16i4.8993, 1964.

Ding, Y., Ren, G., Zhao, Z., Xu, Y., Luo, Y., Li, Q., and Zhang, J.: Detection, causes and projection of climate change over
        China: An overview of recent progress, Advances in Atmospheric Sciences, 24, 954-971, https://doi.org/10.1007/s00376-
460   007-0954-4, 2007.

French, M. N., Krajewski, W. F., and Cuykendall, R. R.: Rainfall forecasting in space and time using a neural network,
        Journal of hydrology, 137, 1-31, https://doi.org/10.1016/0022-1694(92)90046-X, 1992.

Galewsky, J., Steen-Larsen, H. C., Field, R. D., Worden, J., Risi, C., and Schneider, M.: Stable isotopes in atmospheric water
        vapor and applications to the hydrologic cycle, Rev Geophys, 54, 809-865, https://doi.org/10.1002/2015RG000512, 2016.

465   Gao, J., Tian, L., Liu, Y., and Gong, T.: Oxygen isotope variation in the water cycle of the Yamzho lake Basin in southern
        Tibetan Plateau, Science Bulletin, 54, 2758-2765, https://doi.org/10.1007/s11434-009-0487-6, 2009.

Gao, J., Masson-Delmotte, V., Yao, T., Tian, L., Risi, C., and Hoffmann, G.: Precipitation Water Stable Isotopes in the
        South Tibetan Plateau: Observations and Modeling*, Journal of Climate, 24, 3161-3178,
        https://doi.org/10.1175/2010jcli3736.1, 2011.

470   Gat, J. R.: Oxygen and hydrogen isotopes in the hydrologic cycle, Annual Review of Earth and Planetary Sciences, 24, 225-
        262, https://doi.org/10.1146/annurev.earth.24.1.225, 1996.

Gibson, J. J., Edwards, T. W. D., Birks, S. J., St Amour, N. A., Buhay, W. M., McEachern, P., Wolfe, B. B., and Peters, D.
        L.: Progress in isotope tracer hydrology in Canada, Hydrological Processes, 19, 303-327, https://doi.org/10.1002/hyp.5766,
        2005.

475   Hochreiter, S.: The vanishing gradient problem during learning recurrent neural nets and problem solutions, Int J Uncertain
        Fuzz, 6, 107-116, https://doi.org/10.1142/S0218488598000094, 1998.

Hochreiter, S. and Schmidhuber, J.: Long short-term memory, Neural Comput, 9, 1735-1780,
        https://doi.org/10.1162/neco.1997.9.8.1735, 1997.

Hsu, K.-l., Gupta, H. V., and Sorooshian, S.: Artificial Neural Network Modeling of the Rainfall-Runoff Process, 31, 2517-
480   2530, https://doi.org/10.1029/95WR01955, 1995.

Hu, M., Zhang, Y., Wu, K., Shen, H., Yao, M., Dahlgren, R. A., and Chen, D.: Assessment of streamflow components and hydrologic transit times using stable isotopes of oxygen and hydrogen in waters of a subtropical watershed in eastern China, Journal of Hydrology, 589, 125363, https://doi.org/10.1016/j.jhydrol.2020.125363, 2020.

Jiang, D., Tian, Z., and Lang, X.: Reliability of climate models for China through the IPCC Third to Fifth Assessment
485   Reports, International Journal of Climatology, 36, 1114-1133, https://doi.org/10.1002/joc.4406, 2016.

Krenker, A., Bešter, J., and Kos, A.: Introduction to the artificial neural networks, Artificial Neural Networks: Methodological Advances and Biomedical Applications. InTech, 1-18, https://doi.org/10.5772/15751, 2011.

Kurita, N., Noone, D., Risi, C., Schmidt, G. A., Yamada, H., and Yoneyama, K.: Intraseasonal isotopic variation associated with the Madden-Julian Oscillation, Journal of Geophysical Research: Atmospheres, 116, D24101,
490   https://doi.org/10.1029/2010jd015209, 2011.

LeCun, Y.: Generalization and network design strategies, Connectionism in perspective, 19, 143-155, 1989.

Lee, J.-E., Fung, I., DePaolo, D. J., and Henning, C. C.: Analysis of the global distribution of water isotopes using the NCAR atmospheric general circulation model, Journal of Geophysical Research, 112, D16306, https://doi.org/10.1029/2006jd007657, 2007.

495   Li, Z., Feng, Q., Wang, Q. J., Kong, Y., Cheng, A., Yong, S., Li, Y., Li, J., and Guo, X.: Contributions of local terrestrial evaporation and transpiration to precipitation using δ 18 O and D-excess as a proxy in Shiyang inland river basin in China, Global and Planetary Change, 146, 140-151, https://doi.org/10.1016/j.gloplacha.2016.10.003, 2016.

Liang, Y., Qian, H., and Zhang, L.: Regionalization of the annual precipitation change in the last 50 years in China (1961-2010), Acta Meteorologica Sinica, 74, 31-45, https://doi.org/10.11676/qxxb2016.001, 2016.

500   Liu, J., Song, X., Yuan, G., Sun, X., and Yang, L.: Stable isotopic compositions of precipitation in China, Tellus B: Chemical and Physical Meteorology, 66, 22567, https://doi.org/10.3402/tellusb.v66.22567, 2014.

Liu, W. J., Liu, W. Y., Li, P. J., Gao, L., Shen, Y. X., Wang, P. Y., Zhang, Y. P., and Li, H. M.: Using stable isotopes to determine sources of fog drip in a tropical seasonal rain forest of Xishuangbanna, SW China, Agricultural Forest Meteorology, 143, 80-91, https://doi.org/10.1016/j.agrformet.2006.11.009, 2007.

505   Ma, L., Li, H., Liu, T., and Liang, L.: Abrupt temperature change and a warming hiatus from 1951 to 2014 in Inner Mongolia, China, Journal of Arid Land, 11, 192-207, https://doi.org/10.1007/s40333-019-0100-z, 2019.

Maraun, D.: Bias Correction, Quantile Mapping, and Downscaling: Revisiting the Inflation Issue, Journal of Climate, 26, 2137-2143, https://doi.org/10.1175/jcli-d-12-00821.1, 2013.

Miao, C., Duan, Q., Yang, L., and Borthwick, A. G.: On the applicability of temperature and precipitation data from CMIP3
510   for China, PLoS One, 7, e44659, https://doi.org/10.1371/journal.pone.0044659, 2012.

Peng, P., John Zhang, X., and Chen, J.: Bias correcting isotope-equipped GCMs outputs to build precipitation oxygen isoscape for eastern China, Journal of Hydrology, 589, 125153, https://doi.org/10.1016/j.jhydrol.2020.125153, 2020a.

Peng, P., Zhang, X. J., and Chen, J.: Modeling the contributions of oceanic moisture to summer precipitation in eastern China using 18O, Journal of Hydrology, 581, 124304, https://doi.org/10.1016/j.jhydrol.2019.124304, 2020b.

515   Ren, G., Ding, Y., Zhao, Z., Zheng, J., Wu, T., Tang, G., and Xu, Y.: Recent progress in studies of climate change in China,
        Advances in Atmospheric Sciences, 29, 958-977, https://doi.org/10.1007/s00376-012-1200-2, 2012.

Risi, C., Bony, S., Vimeux, F., and Jouzel, J.: Water-stable isotopes in the LMDZ4 general circulation model: Model
        evaluation for present-day and past climates and applications to climatic interpretations of tropical isotopic records,
        Journal of Geophysical Research, 115, D12118, https://doi.org/10.1029/2009jd013255, 2010.

520   Risi, C., Noone, D., Worden, J., Frankenberg, C., Stiller, G., Kiefer, M., Funke, B., Walker, K., Bernath, P., Schneider, M.,
        Bony, S., Lee, J., Brown, D., and Sturm, C.: Process-evaluation of tropospheric humidity simulated by general circulation
        models using water vapor isotopic observations: 2. Using isotopic diagnostics to understand the mid and upper
        tropospheric moist bias in the tropics and subtropics, Journal of Geophysical Research: Atmospheres, 117, D05304,
        https://doi.org/10.1029/2011jd016623, 2012.

525   Rozanski, K., Araguás-Araguás, L., and Gonfiantini, R.: Isotopic Patterns in Modern Global Precipitation, Geophysical
        Monograph-American Geophysical Union, 78, 1-36, https://doi.org/10.1029/GM078p0001, 1993.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J.: Learning Representations by Back-Propagating Errors, Nature, 323,
        533-536, https://doi.org/10.1038/323533a0, 1986.

Rumelhart, D. E., Widrow, B., and Lehr, M. A.: The basic ideas in neural networks, Communications of the ACM, 37, 87-92,
530   https://doi.org/10.1145/175247.175256, 1994.

Schmidt, G. A., LeGrande, A. N., and Hoffmann, G.: Water isotope expressions of intrinsic and forced variability in a
        coupled   ocean-atmosphere   model,   Journal   of   Geophysical   Research: Atmospheres,   112,   D10103,
        https://doi.org/10.1029/2006jd007781, 2007.

Sturm, C., Zhang, Q., and Noone, D.: An introduction to stable water isotopes in climate models: benefits of forward proxy
535   modelling for paleoclimatology, Clim Past, 6, 115-129, https://doi.org/10.5194/cp-6-115-2010, 2010.

Su, F., Duan, X., Chen, D., Hao, Z., and Cuo, L.: Evaluation of the Global Climate Models in the CMIP5 over the Tibetan
        Plateau, Journal of Climate, 26, 3187-3208, https://doi.org/10.1175/jcli-d-12-00321.1, 2013.

Tang, Y., Song, X., Zhang, Y., Han, D., Ai, L., Zhao, T., and Wang, Y.: Using stable isotopes to understand seasonal and
        interannual dynamics in moisture sources and atmospheric circulation in precipitation, Hydrological Processes, 31, 4682-
540   4692, http://doi.org/10.1002/hyp.11388, 2017.

Tian, L., Yao, T., MacClune, K., White, J. W. C., Schilla, A., Vaughn, B., Vachon, R., and Ichiyanagi, K.: Stable isotopic
        variations in west China: A consideration of moisture sources, Journal of Geophysical Research: Atmospheres, 112,
        D10112, https://doi.org/10.1029/2006jd007718, 2007.

Tindall, J. C., Valdes, P. J., and Sime, L. C.: Stable water isotopes in HadCM3: Isotopic signature of El Niño–Southern
545   Oscillation   and   the   tropical   amount   effect,   Journal   of   Geophysical   Research,   114,   D04111,
        https://doi.org/10.1029/2008jd010825, 2009.

Vuille, M., Werner, M., Bradley, R. S., and Keimig, F.: Stable isotopes in precipitation in the Asian monsoon region, Journal
        of Geophysical Research, 110, D23108, https://doi.org/10.1029/2005jd006022, 2005.

Wang, S., Zhang, M., Chen, F., Che, Y., Du, M., and Liu, Y.: Comparison of GCM-simulated isotopic compositions of
550   precipitation in arid central Asia, Journal of Geographical Sciences, 25, 771-783, https://doi.org/10.1007/s11442-015-1201-z, 2015.

Wang, Y., Hu, C., Ruan, J., and Johnson, K. R.: East Asian precipitation δ18O relationship with various monsoon indices, Journal of Geophysical Research: Atmospheres, 125, e2019JD032282, https://doi.org/10.1029/2019JD032282, 2020.

Wen, X., Wu, X., and Gao, M.: Spatiotemporal variability of temperature and precipitation in Gansu Province (Northwest
555   China) during 1951–2015, Atmospheric Research, 197, 132-149, https://doi.org/10.1016/j.atmosres.2017.07.001, 2017.

Xi, X.: A Review of Water Isotopes in Atmospheric General Circulation Models: Recent Advances and Future Prospects, International Journal of Atmospheric Sciences, 2014, 1-16, https://doi.org/10.1155/2014/250920, 2014.

Xie, L., Wei, G., Deng, W., and Zhao, X.: Daily δ18O and δD of precipitations from 2007 to 2009 in Guangzhou, South China: Implications for changes of moisture sources, Journal of Hydrology, 400, 477-489,
560   https://doi.org/10.1016/j.jhydrol.2011.02.002, 2011.

Yamanaka, T., Tsujimura, M., Oyunbaatar, D., and Davaa, G.: Isotopic variation of precipitation over eastern Mongolia and its implication for the atmospheric water cycle, Journal of Hydrology, 333, 21-34, https://doi.org/10.1016/j.jhydrol.2006.07.022, 2007.

Yang, Q., Ma, Z., and Xu, B.: Modulation of monthly precipitation patterns over East China by the Pacific Decadal
565   Oscillation, Climatic Change, 144, 405-417, https://doi.org/10.1007/s10584-016-1662-9, 2016.

Yang, X., Davis, M. E., Acharya, S., and Yao, T.: Asian monsoon variations revealed from stable isotopes in precipitation, Climate Dynamics, 51, 2267-2283, https://doi.org/10.1007/s00382-017-4011-4, 2017.

Yang, X., Yao, T., Yang, W., Yu, W., and Qu, D.: Co-existence of temperature and amount effects on precipitation δ18O in the Asian monsoon region, Geophysical Research Letters, 38, L21809, https://doi.org/10.1029/2011gl049353, 2011.

570   Yao, T., Masson-Delmotte, V., Gao, J., Yu, W., Yang, X., Risi, C., Sturm, C., Werner, M., Zhao, H., He, Y., Ren, W., Tian, L., Shi, C., and Hou, S.: A review of climatic controls on δ18O in precipitation over the Tibetan Plateau: Observations and simulations, Reviews of Geophysics, 51, 525-548, https://doi.org/10.1002/rog.20023, 2013.

Yoshimura, K., Kanamitsu, M., Noone, D., and Oki, T.: Historical isotope simulation using Reanalysis atmospheric data, Journal of Geophysical Research, 113, D19108, https://doi.org/10.1029/2008jd010074, 2008.

575   Yu, W., Wei, F., Ma, Y., Liu, W., Zhang, Y., Luo, L., Tian, L., Xu, B., and Qu, D.: Stable isotope variations in precipitation over Deqin on the southeastern margin of the Tibetan Plateau during different seasons related to various meteorological factors and moisture sources, Atmospheric Research, 170, 123-130, https://doi.org/10.1016/j.atmosres.2015.11.013, 2016a.

Yu, W., Tian, L., Risi, C., Yao, T., Ma, Y., Zhao, H., Zhu, H., He, Y., Xu, B., Zhang, H., and Qu, D.: δ18O records in water vapor and an ice core from the eastern Pamir Plateau: Implications for paleoclimate reconstructions, Earth and Planetary
580   Science Letters, 456, 146-156, https://doi.org/10.1016/j.epsl.2016.10.001, 2016b.

Yu, W., Yao, T., Thompson, L. G., Jouzel, J., Zhao, H., Xu, B., Jing, Z., Wang, N., Wu, G., Ma, Y., Gao, J., Yang, X., Zhang, J., and Qu, D.: Temperature signals of ice core and speleothem isotopic records from Asian monsoon region as

indicated by precipitation δ18O, Earth and Planetary Science Letters, 554, https://doi.org/10.1016/j.epsl.2020.116665, 2021.

585 Zhai, P., Zhang, X., Wan, H., and Pan, X.: Trends in total precipitation and frequency of daily precipitation extremes over China, Journal of climate, 18, 1096-1108, https://doi.org/10.1175/JCLI-3318.1, 2005.

Zhang, J., Zhu, Y., Zhang, X., Ye, M., and Yang, J.: Developing a Long Short-Term Memory (LSTM) based model for predicting water table depth in agricultural areas, Journal of Hydrology, 561, 918-929, https://doi.org/10.1016/j.jhydrol.2018.04.065, 2018.

590 Zhang, M. and Wang, S.: A review of precipitation isotope studies in China: Basic pattern and hydrological process, Journal of Geographical Sciences, 26, 921-938, https://doi.org/10.1007/s11442-016-1307-y, 2016.

Zhang, X., Sun, Z., Guan, H., Zhang, X., Wu, H., and Huang, Y.: GCM simulations of stable isotopes in the water cycle in comparison with GNIP observations over East Asia, Acta Meteorologica Sinica, 26, 420-437, https://doi.org/10.1007/s13351-012-0403-x, 2012.

595 Zhao, L., Xiao, H., Zhou, M., Cheng, G., Wang, L., Yin, L., and Ren, J.: Factors controlling spatial and seasonal distributions of precipitation δ18O in China, Hydrological Processes, 26, 143-152, https://doi.org/10.1002/hyp.8118, 2012.

Zhao, L., Yin, L., Xiao, H., Cheng, G., Zhou, M., Yang, Y., Li, C., and Zhou, J.: Isotopic evidence for the moisture origin and composition of surface runoff in the headwaters of the Heihe River basin, Chinese Science Bulletin, 56, 406-415, https://doi.org/10.1007/s11434-010-4278-x, 2011.

600 Zhou, H., Zhang, X., Yao, T., Hua, M., Wang, X., Rao, Z., and He, X.: Variation of δ18O in precipitation and its response to upstream atmospheric convection and rainout: A case study of Changsha station, south-central China, Science of the Total Environment, 659, 1199-1208, https://doi.org/10.1016/j.scitotenv.2018.12.396, 2019.

Zhu, Y.-Y. and Yang, S.: Evaluation of CMIP6 for historical temperature and precipitation over the Tibetan Plateau and its comparison with CMIP5, Advances in Climate Change Research, 11, 239-251, https://doi.org/10.1016/j.accre.2020.08.001, 605 2020.