



Adjusting spatial dependence of climate model outputs with cycle-consistent adversarial networks

Bastien François¹ · Soulivanh Thao¹ · Mathieu Vrac¹

Received: 4 March 2021 / Accepted: 25 June 2021 / Published online: 30 July 2021
© The Author(s) 2021

Abstract

Climate model outputs are commonly corrected using statistical univariate bias correction methods. Most of the time, those 1d-corrections do not modify the ranks of the time series to be corrected. This implies that biases in the spatial or inter-variable dependences of the simulated variables are not adjusted. Hence, over the last few years, some multivariate bias correction (MBC) methods have been developed to account for inter-variable structures, inter-site ones, or both. As proof-of-concept, we propose to adapt a computer vision technique used for Image-to-Image translation tasks (CycleGAN) for the adjustment of spatial dependence structures of climate model projections. The proposed algorithm, named MBC-CycleGAN, aims to transfer simulated maps (seen as images) with inappropriate spatial dependence structure from climate model outputs to more realistic images with spatial properties similar to the observed ones. For evaluation purposes, the method is applied to adjust maps of temperature and precipitation from climate simulations through two cross-validation approaches. The first one is designed to assess two different post-processing schemes (Perfect Prognosis and Model Output Statistics). The second one assesses the influence of nonstationary properties of climate simulations on the performance of MBC-CycleGAN to adjust spatial dependences. Results are compared against a popular univariate bias correction method, a “quantile-mapping” method, which ignores inter-site dependencies in the correction procedure, and two state-of-the-art multivariate bias correction algorithms aiming to adjust spatial correlation structure. In comparison with these alternatives, the MBC-CycleGAN algorithm reasonably corrects spatial correlations of climate simulations for both temperature and precipitation, encouraging further research on the improvement of this approach for multivariate bias correction of climate model projections.

Keywords Bias correction · Spatial dependence · Post-processing · Climate simulations · Generative adversarial networks · Model output statistics

1 Introduction

With ongoing climate change, mitigation and adaptation strategies have to be anticipated by decision makers in order to reduce potential future consequences of climate change on human societies and activities (IPCC 2014). Such consequences are commonly assessed through climate change impact studies, for instance in hydrology (e.g., Bates et al. 2008), agronomy (e.g., Wheeler and von Braun 2013) or epidemiology (e.g., Caminade et al. 2014). They rely on impact

model simulations, the quality of which highly depends on the reliability of the climate information used as inputs (e.g., Muerth et al. 2013; Ramirez-Villegas et al. 2013). Besides observations, global and regional climate models (GCM and RCM) are the major tools to understand the climate system and its evolutions in the future (Randall et al. 2007; Reichler and Kim 2008). However, despite considerable improvements in climate modelling, climate simulations often remain biased compared to observations: even for the current climate, key statistical features such as mean, variance or the dependence structures between physical variables or between sites can differ from those calculated for observational references (e.g., Eden et al. 2012; Cattiaux et al. 2013; Mueller and Seneviratne 2014). Consequently, biases are expected to be present in climate projections for future periods, making bias correction an often unavoidable

✉ Bastien François
bastien.francois@lsce.ipsl.fr

¹ Laboratoire des Sciences du Climat et l'Environnement (LSCE-IPSL) CNRS/CEA/UVSQ, UMR8212, Université Paris-Saclay, Gif-sur-Yvette, France

data pre-processing step for impact studies (e.g., Christensen et al. 2008; Maraun et al. 2010; Teutschbein and Seibert 2012).

In the recent years, many statistical bias correction (BC) methods have been developed that aim to correct (selected features of) the distribution of climate variables. The idea of statistical bias correction is to find a mathematical transformation that makes climate simulations have similar statistical properties as a reference dataset over the historical period, and then apply this transformation for the modeled projection. Such transformations may be determined with statistical models based on either perfect prognosis (PP) or model output statistics (MOS) approaches (Maraun et al. 2010). The PP approach consists in determining the statistical link between a variable of interest from references (predictand) and one or several observed variables (predictors) occurring at the same time. Simultaneous values of predictand and predictors are indeed required to implement the PP approach and learn the (synchronous) relationships between them. By applying these relationships to predictors from climate simulations, this approach implicitly makes the assumption that these predictors are realistically simulated (Wilks 2006). In the MOS approach, observed and simulated variables are not considered to be synchronized in time, and biases relate to differences in some statistics (such as means or variances) or in distributions between references and modeled climate variables. Adjustments can be made to the simulated mean (e.g., Delta method, Xu 1999), variance (e.g., simple scaling adjustment, Berg et al. 2012) and also all moments of higher order and percentiles (e.g., “quantile-mapping”, Haddad and Rosenfeld 1997; Déqué 2007; Gudmundsson et al. 2012). In particular, quantile-mapping technique has received a keen interest since it permits for adjusting not only the mean and variance but also the whole distribution of climate variables. It has been conducive to the development of many variants (e.g., Vrac et al. 2012, 2016; Trambly et al. 2013; Cannon et al. 2015), and applied for various studies (e.g., Vigaud et al. 2013; Defrance et al. 2017; Bartok et al. 2019; Tong et al. 2020). However, such BC methods are designed to only correct statistical aspects of univariate distributions. Simulated variables are indeed adjusted separately for each physical variable at each specific location. Thus, potential biases in the spatial dependence structure of modeled variables are not corrected (e.g., Wilcke et al. 2013), which can generate corrections with inappropriate multivariate situations and can affect subsequent analyses that depend on spatial characteristics of climate variables (e.g., Zscheischler et al. 2019). For instance, this can occur with flood risk assessment, that depends on spatial (and temporal) properties of precipitation, soil moisture and river flow (Vorogushyn et al. 2018) or with drought-related impacts, that depend on complex interaction of natural and anthropogenic processes (Van Loon et al. 2016). It is hence

crucial to provide end users with bias corrections of climate simulations that present not only relevant 1-dimensional information at each individual site but also appropriate spatial representation.

Over the last years, a few multivariate bias correction (MBC) methods have been developed to address the issues of biases in multivariate dependencies. Not only do these methods correct marginal properties of simulated variables, they are also designed to adjust statistical dependencies between variables. Although it has been found for specific cases that MBC methods do not particularly outperform univariate ones for the adjustment of dependencies between multiple variables (Räty et al. 2018), this finding cannot be generalized to all applications and methods. For instance, François et al. (2020) showed the added value of MBC to improve inter-variable dependence and spatial structures for temperature and precipitation over Europe. More generally, MBCs could be of great interest for compound events studies, where dependencies between drivers of extreme events with large impacts are crucial to evaluate their risks (Zscheischler et al. 2018).

A categorization of MBC methods in three main families of approaches has been proposed in the literature (e.g., Vrac 2018; François et al. 2020):

- the “marginal/dependence” correction approach, that consists of MBC methods adjusting in two distinct steps, i.e. separately, marginal distributions and multivariate dependencies of climate simulations (e.g., Bárdossy and Pegram 2012; Mehrotra and Sharma 2016; Hnilica et al. 2017; Nahar et al. 2018; Cannon 2018; Nguyen et al. 2019; Guo et al. 2019; Vrac and Thao 2020).
- the “successive conditional” category, made up of MBC methods performing successive univariate corrections of climate variables conditionally on the previously adjusted ones (e.g., Piani and Haerter 2012; Dekens et al. 2017).
- the “all-in-one” correction approach, that adjusts directly the whole statistical distribution (i.e. both univariate and multivariate properties) of climate simulations at the same time (e.g., Robin et al. 2019).

Based on this categorization, François et al. (2020) performed an intercomparison and critical review of MBC methods. It presents a global picture of the performances of MBCs in terms of multivariate adjustments of climate simulations, as well as the different assumptions and statistical techniques used.

In parallel, i.e., in contexts other than bias correction, over the last decades, machine learning techniques have emerged as a promising approach to model highly non-linear and complex relationships between statistical variables. Major improvements have been obtained with Deep Learning models (see the overview of Schmidhuber 2015),

which have proved to be efficient to extract high-level feature information from various datasets. In particular, convolutional neural networks (CNNs, see e.g., Lecun and Bengio 1995) showed that they can capture with great performances complex spatial structures. Initially developed for computer vision problems (e.g., Szegedy et al. 2015; He et al. 2016), they found numerous applications in climate sciences: for instance for weather forecast prediction uncertainty (Scher and Messori 2018), emulations of atmospheric dynamics (Shi et al. 2015; Scher and Messori 2019; Chapman et al. 2019), detection of extreme weather events (Liu et al. 2016; Racah et al. 2017) and statistical downscaling (Vandal et al. 2017; Rodrigues et al. 2018; Baño-Medina et al. 2020). A recent overview of Deep Learning applications for Earth system science is offered by Reichstein et al. (2019).

Recently, a new class of artificial neural networks, named Generative Adversarial Networks (GANs; Goodfellow et al. 2014), has led to tremendous interests due to their ability to infer high dimensional probability distributions. Initially, this machine-learning method has been developed for estimating the distribution of images from a target dataset, with the aim of sampling new (and unseen) images from this distribution. GANs, implemented with deep convolutional neural networks, have achieved impressive results in computer vision problems (e.g., Radford et al. 2016) and are a subject of active research to improve computing architectures (e.g., Salimans et al. 2016; Karras et al. 2018; Menick and Kalchbrenner 2018) and optimization techniques (e.g., Mao et al. 2017; Arjovsky et al. 2017; Roth et al. 2017). Conditional formulations of GANs have also been developed, for which additional information, such as class labels or images, can serve as inputs to condition the generation of the new images (e.g., Mirza and Osindero 2014; Gauthier 2014; Denton et al. 2015; Kim et al. 2017; Isola et al. 2017). In particular, image-conditional GANs permit to perform image-to-image translation tasks by learning how to map the statistical distribution of one set of images (source dataset) to the statistical distribution of another set (target dataset). Depending on the correspondence between images of the source and target datasets, different versions of image-conditional GANs have been developed. When all the images are paired (i.e., there is a known one-to-one correspondence between every images of the source and target datasets), conditional GANs are trained by supervised learning (Yoo et al. 2016; Isola et al. 2017). When only a few images are paired, semi-supervised is used (Gan et al. 2017) and when all points are unpaired, only unsupervised learning can be applied (Kim et al. 2017; Yi et al. 2017; Zhu et al. 2017). Due to the stochastic and high-dimensionality nature of many physical processes of the Earth system, GANs and conditional GANs are particularly appealing for atmospheric science problems. Recently, they have been used for various Earth-science related applications: for instance for statistical downscaling (Leinonen

et al. 2020; Wang et al. 2021), temporal disaggregation of spatial rainfall fields (Scher and Peßenteiner 2020), sampling of extreme values (Bhatia et al. 2020), modelling of chaotic dynamical systems (e.g., Xie et al. 2018; Wu et al. 2020), classification of snowflake images (Leinonen and Berne 2020), weather forecasting (Bihlo 2020) and stochastic parameterization in geophysical models (Gagne II et al. 2020).

In climate modelling context, no one-to-one correspondence exists between observations and model simulations as they have different internal variabilities and thus are not synchronized in time. Biases refer to differences in distributional properties between references and simulated climate variables. Hence, in this context, bias correction can be seen as an unsupervised image-to-image problem that aims to map daily images from model simulations to daily images from historical observational references in order to adjust the distributional properties of the climate model.

In this study, we adapt a specific formulation of conditional GANs, initially used for unsupervised image-to-image translation problems (CycleGAN, Zhu et al. 2017), for multi-site corrections of climate simulations. The new MBC method, referred to as MBC-CycleGAN in the following, is introduced and applied in a proof-of-concept context for the correction of daily temperature and precipitation fields with a simple neural network architecture. In order to investigate and evaluate the proposed methodology, applications and comparisons of MBC-CycleGAN based on PP (corresponding to a supervised context) and MOS (unsupervised context) approaches are performed through a cross-validation method. In addition, a second cross-validation method is used in this study to assess the performances of MBC-CycleGAN in a context of different degrees of nonstationarity of the climate model between present (i.e., calibration) and future (i.e., projection) periods. One univariate quantile-mapping-based BC method and two MBC algorithms are included in the study in order to gain a better understanding of the performances of MBC-CycleGAN concerning univariate, spatial and temporal properties.

The paper is organized as follows: Section 2 presents the model and reference data used, and Sect. 3 describes the MBC-CycleGAN algorithm. Then, Sect. 4 displays the experimental setup used in this study, and results are provided in Sect. 5. Conclusions, discussions and perspectives for future research are finally proposed in Sect. 6.

2 Reference and model data

In this study, the dataset employed as reference for the bias correction is the “Système d’Analyse Fournissant des Renseignements Atmosphériques à la Neige” (SAFRAN) reanalysis (Vidal et al. 2010) with an approximate $8 \text{ km} \times 8$

km spatial resolution. Daily temperature and precipitation time series from 1 January 1979 to 31 December 2016 are extracted over the region of Paris, France ($[47.878, 49.830^\circ \text{N}] \times [0.949, 3.947^\circ \text{E}]$), which corresponds to a domain with $28 \times 28 = 784$ continental grid cells.

For the climate simulations data to be corrected, daily temperature and precipitation time series are taken from runs of the IPSL-CM5A-MR Earth system model (Marti et al. 2010; Dufresne et al. 2013) with a $1.25^\circ \times 2.5^\circ$ spatial resolution over the same region of Paris. For the 1979–2005 period, a historical run is extracted and concatenated with a run under RCP 8.5 scenario (i.e., the scenario with highest CO_2 concentration) for the 2006–2016 period, to obtain the desired 1979–2016 period. To perform a bias correction, a one-to-one correspondence between model and reference grid cells is needed, i.e., spatial resolutions between reference and model data have to be the same. Hence, IPSL data are regridded to the SAFRAN spatial resolution with a bilinear interpolation for both temperature and precipitation.

More data are required for this study, in particular for the implementation of the PP approach and to assess the influence of nonstationary properties of climate simulations on the performance of the proposed MBC method. For sake of clarity and make reading easier, these data will be introduced thereafter in the appropriate sections.

For illustration purpose, Fig. 1a displays the topographic map of France with the region of Paris in a box, as well as the mean daily temperature (Fig. 1b, c) and precipitation (Fig. 1d, e) maps for SAFRAN and IPSL datasets during winter over the 1979–2016 period for Paris.

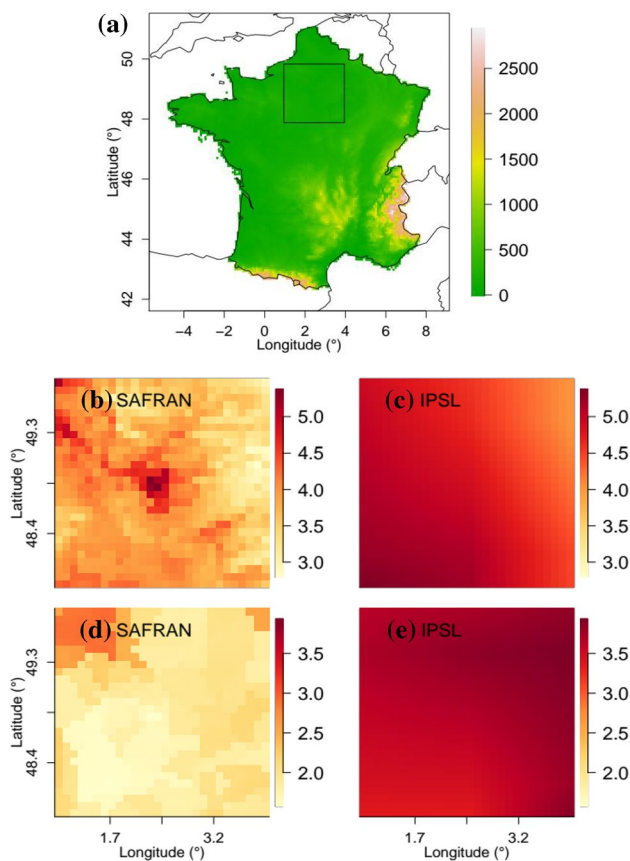


Fig. 1 a Topographic map of France with the selected region over Paris in a box, b, c temperature and d–e precipitation daily mean computed at each grid cell during winter over the 1979–2016 period for Paris. Results are shown for SAFRAN reference and plain IPSL outputs

3 Methodology

3.1 GAN

In its most basic formulation, a generative adversarial network consists of two neural networks that are trained jointly: a generator and a discriminator. We first consider one random variable \mathbf{Y} living in \mathbb{R}^d , with a probability distribution denoted $\mathbb{P}_{\mathbf{Y}}$. This random variable characterizes the available data, such as images of the target dataset (i.e., references), and hence takes its values in a high-dimensional space. We assume to have at hand samples $\mathbf{y}_1, \dots, \mathbf{y}_n$ drawn according to the density $\mathbb{P}_{\mathbf{Y}}$ on \mathbb{R}^d . The generator, denoted G , is a function from $\mathbb{R}^{d'}$ to \mathbb{R}^d and is intended to be applied to a d' -dimensional random variable \mathbf{W} , usually multivariate Gaussian random noise (with $d' \ll d$), such that the random variable $G(\mathbf{W})$ follows the law of \mathbf{Y} , i.e. $\mathbb{P}_{\mathbf{Y}} = \mathbb{P}_{G(\mathbf{W})}$. Let $\mathbf{w}_1, \dots, \mathbf{w}_n$ be a sample drawn from the distribution of \mathbf{W} . To train the generator G , the discriminator $D_{\mathbf{Y}}$, that is a function from \mathbb{R}^d to $[0, 1]$, is used as complex loss function (Goodfellow et al. 2014). This neural network is a binary

classifier that returns the probability that a given observation, or image, comes from $\mathbb{P}_{\mathbf{Y}}$. The discriminator is trained in a supervised way to return maximal probability values on the reference images \mathbf{y}_i and minimal values on the artificially generated images $G(\mathbf{w}_i)$. Conversely, the goal of the generator is to “fool” the discriminator by making the distribution of $G(\mathbf{w}_i)$ as indistinguishable as possible from that of \mathbf{y}_i , i.e., making difficult for the discriminator to determine that a sample $G(\mathbf{w}_i)$ comes from a distribution different from $\mathbb{P}_{\mathbf{Y}}$. Generator and discriminator are trained in turns and are in competition (i.e. “adversarial training”) to improve themselves until it reaches an optimal equilibrium state.

The original formulation of GANs explained above is unconditional: the generator G only takes as input noise vectors \mathbf{w}_i to produce new samples that are drawn from the target distribution $\mathbb{P}_{\mathbf{Y}}$. The idea of conditional GANs (e.g., Goodfellow et al. 2014; Mirza and Osindero 2014) is to add some information as inputs to direct the generation. By conditioning the generation on an input image, the generator is able to generate a corresponding output image, rendering the

conditional GANs appropriate for image-to-image translation tasks (e.g., Isola et al. 2017).

3.2 CycleGAN for unsupervised image-to-image translation

CycleGAN (Zhu et al. 2017) is a particular image-conditional GANs that is commonly used for unsupervised image-to-image translation. In the original application, CycleGAN has been applied with great success to transform photographs into the styles of master paintings by modifying colour information (i.e., RGB colour channels and/or spatial features of colours) of the photographs. Instead of the random noise \mathbf{W} , we introduce another random variable \mathbf{X} , with probability distribution $\mathbb{P}_{\mathbf{X}}$, living in the same dimensional space as \mathbf{Y} (i.e., \mathbb{R}^d). This random variable \mathbf{X} characterizes the images of the source dataset (i.e., biased simulations to correct). The CycleGAN approach consists in learning a mapping (i.e., a generator) $G_{\mathbf{X} \rightarrow \mathbf{Y}} : \mathbb{R}^d \rightarrow \mathbb{R}^d$ such that the random variable $G_{\mathbf{X} \rightarrow \mathbf{Y}}(\mathbf{X})$ follows the law of \mathbf{Y} (i.e., $\mathbb{P}_{\mathbf{Y}} = \mathbb{P}_{G_{\mathbf{X} \rightarrow \mathbf{Y}}(\mathbf{X})}$). In addition to samples $\mathbf{y}_1, \dots, \mathbf{y}_n$, we assume to have at hand image samples $\mathbf{x}_1, \dots, \mathbf{x}_n$ drawn according to density $\mathbb{P}_{\mathbf{X}}$ on \mathbb{R}^d . Similarly as unconditional GANs, the mapping $G_{\mathbf{X} \rightarrow \mathbf{Y}}$ is learned using an adversarial loss, i.e. with a discriminator $D_{\mathbf{Y}}$ which forces the generator $G_{\mathbf{X} \rightarrow \mathbf{Y}}$ to generate images from a distribution close to the target distribution $\mathbb{P}_{\mathbf{Y}}$. The adversarial loss is defined as:

$$L_{GAN}(G_{\mathbf{X} \rightarrow \mathbf{Y}}, D_{\mathbf{Y}}) = \frac{1}{n} \sum_{i=1}^n \ln D_{\mathbf{Y}}(\mathbf{y}_i) + \frac{1}{n} \sum_{i=1}^n \ln(1 - D_{\mathbf{Y}} \circ G_{\mathbf{X} \rightarrow \mathbf{Y}}(\mathbf{x}_i)). \tag{1}$$

$G_{\mathbf{X} \rightarrow \mathbf{Y}}$ aims to minimize this adversarial objective against $D_{\mathbf{Y}}$, that means, tries to fool the discriminator with its generated images (i.e., maximizing the probability $D_{\mathbf{Y}}(G_{\mathbf{X} \rightarrow \mathbf{Y}}(\mathbf{x}_i))$). On the contrary, the discriminator $D_{\mathbf{Y}}$ aims to maximize the adversarial loss by distinguishing between transferred samples $G_{\mathbf{X} \rightarrow \mathbf{Y}}(\mathbf{x}_i)$ and samples \mathbf{y}_i from the distribution $\mathbb{P}_{\mathbf{Y}}$. A perfect discriminator $D_{\mathbf{Y}}$ would return probability values equal to 1 for samples drawn from $\mathbb{P}_{\mathbf{Y}}$ and equal to 0 for samples generated by $G_{\mathbf{X} \rightarrow \mathbf{Y}}$. Hence, $G_{\mathbf{X} \rightarrow \mathbf{Y}}$ is designed to solve the optimization problem against $D_{\mathbf{Y}}$:

$$G_{\mathbf{X} \rightarrow \mathbf{Y}} = \arg \min_{G_{\mathbf{X} \rightarrow \mathbf{Y}}} \max_{D_{\mathbf{Y}}} L_{GAN}(G_{\mathbf{X} \rightarrow \mathbf{Y}}, D_{\mathbf{Y}}). \tag{2}$$

As highlighted by Zhu et al. (2017), this adversarial objective for unsupervised problems is under-constrained: there is no guarantee that “an individual input \mathbf{x}_i and output \mathbf{y}_i are paired up in a meaningful way” with such a mapping $G_{\mathbf{X} \rightarrow \mathbf{Y}}$. In fact, without further constraints, several different mappings can optimize similarly the adversarial loss by transferring the same set of images from $\mathbb{P}_{\mathbf{X}}$ to any random permutation of a same set of images from the distribution $\mathbb{P}_{\mathbf{Y}}$. Moreover, optimizing in practice this under-constrained

adversarial objective alone has been found to be difficult for unsupervised problems, often leading to a well-known problem called “mode collapse”. Mode collapse appears when a generator fails to model the complete range of input images. This results in a lack of diversity in the generated outputs. To address these issues, Zhu et al. (2017) propose to reduce the number of possible mapping functions by adding more constraints to the optimization problem. To do so, they introduce the inverse mapping $G_{\mathbf{Y} \rightarrow \mathbf{X}} : \mathbb{R}^d \rightarrow \mathbb{R}^d$, as well as a second discriminator $D_{\mathbf{X}}$ aimed to recognize images from the distribution $\mathbb{P}_{\mathbf{X}}$. Similarly to the mapping $G_{\mathbf{X} \rightarrow \mathbf{Y}}$, an equivalent adversarial loss can be used to learn the mapping $G_{\mathbf{Y} \rightarrow \mathbf{X}}$ by solving $\arg \min_{G_{\mathbf{Y} \rightarrow \mathbf{X}}} \max_{D_{\mathbf{X}}} L_{GAN}(G_{\mathbf{Y} \rightarrow \mathbf{X}}, D_{\mathbf{X}})$. Zhu et al. (2017) proposed to use $G_{\mathbf{Y} \rightarrow \mathbf{X}}$ to enforce the learned mappings to be cycle-consistent. That means that, for each input image \mathbf{x}_i , the mappings $G_{\mathbf{X} \rightarrow \mathbf{Y}}$ and $G_{\mathbf{Y} \rightarrow \mathbf{X}}$ can be constrained such that it learns to translate \mathbf{x}_i back to the initial image, i.e. $G_{\mathbf{Y} \rightarrow \mathbf{X}} \circ G_{\mathbf{X} \rightarrow \mathbf{Y}}(\mathbf{x}_i) \approx \mathbf{x}_i$ (and similarly for image \mathbf{y}_i , such that $G_{\mathbf{X} \rightarrow \mathbf{Y}} \circ G_{\mathbf{Y} \rightarrow \mathbf{X}}(\mathbf{y}_i) \approx \mathbf{y}_i$). This property can be enforced by using a “cycle-consistency” loss which is defined as:

$$L_{cyc}(G_{\mathbf{X} \rightarrow \mathbf{Y}}, G_{\mathbf{Y} \rightarrow \mathbf{X}}) = \frac{1}{n} \sum_{i=1}^n \|G_{\mathbf{Y} \rightarrow \mathbf{X}}(G_{\mathbf{X} \rightarrow \mathbf{Y}}(\mathbf{x}_i)) - \mathbf{x}_i\|_1 + \frac{1}{n} \sum_{i=1}^n \|G_{\mathbf{X} \rightarrow \mathbf{Y}}(G_{\mathbf{Y} \rightarrow \mathbf{X}}(\mathbf{y}_i)) - \mathbf{y}_i\|_1. \tag{3}$$

Finally, to ensure that images in $\mathbf{x}_1, \dots, \mathbf{x}_n$ that already seem to be draw from the distribution $\mathbb{P}_{\mathbf{Y}}$ (and vice-versa) are not mapped to another images, an identity mapping loss can also be defined as:

$$L_{id}(G_{\mathbf{X} \rightarrow \mathbf{Y}}, G_{\mathbf{Y} \rightarrow \mathbf{X}}) = \frac{1}{n} \sum_{i=1}^n \|G_{\mathbf{Y} \rightarrow \mathbf{X}}(\mathbf{x}_i) - \mathbf{x}_i\|_1 + \frac{1}{n} \sum_{i=1}^n \|G_{\mathbf{X} \rightarrow \mathbf{Y}}(\mathbf{y}_i) - \mathbf{y}_i\|_1, \tag{4}$$

which further reduces the solution space of mapping functions and prevents even more the optimization problem from being under-constrained. The full objective function of the CycleGAN architecture can be expressed as follows:

$$L(G_{\mathbf{X} \rightarrow \mathbf{Y}}, G_{\mathbf{Y} \rightarrow \mathbf{X}}, D_{\mathbf{X}}, D_{\mathbf{Y}}) = L_{GAN}(G_{\mathbf{X} \rightarrow \mathbf{Y}}, D_{\mathbf{Y}}) + L_{GAN}(G_{\mathbf{Y} \rightarrow \mathbf{X}}, D_{\mathbf{X}}) + \lambda_{cyc} L_{cyc}(G_{\mathbf{X} \rightarrow \mathbf{Y}}, G_{\mathbf{Y} \rightarrow \mathbf{X}}) + \lambda_{id} L_{id}(G_{\mathbf{X} \rightarrow \mathbf{Y}}, G_{\mathbf{Y} \rightarrow \mathbf{X}}), \tag{5}$$

where λ_{cyc} and λ_{id} control the relative importance of both cycle-consistency and identity losses. Finally, the CycleGAN aims to solve:

$$(G_{\mathbf{X} \rightarrow \mathbf{Y}}, G_{\mathbf{Y} \rightarrow \mathbf{X}}) = \arg \min_{G_{\mathbf{X} \rightarrow \mathbf{Y}}, G_{\mathbf{Y} \rightarrow \mathbf{X}}} \max_{D_{\mathbf{X}}, D_{\mathbf{Y}}} L(G_{\mathbf{X} \rightarrow \mathbf{Y}}, G_{\mathbf{Y} \rightarrow \mathbf{X}}, D_{\mathbf{X}}, D_{\mathbf{Y}}). \tag{6}$$

Although estimating the inverse mapping $G_{Y \rightarrow X}$ is not necessarily the initial goal of many image-to-image translation problems, its use to constrain the optimization problem has been found to be crucial in an unsupervised context for the convergence of the algorithm and the estimation of the desired mapping $G_{X \rightarrow Y}$. Illustrations of the adversarial, cycle-consistent and identity losses within the CycleGAN architecture are given in Fig. 2.

3.3 The MBC-CycleGAN approach

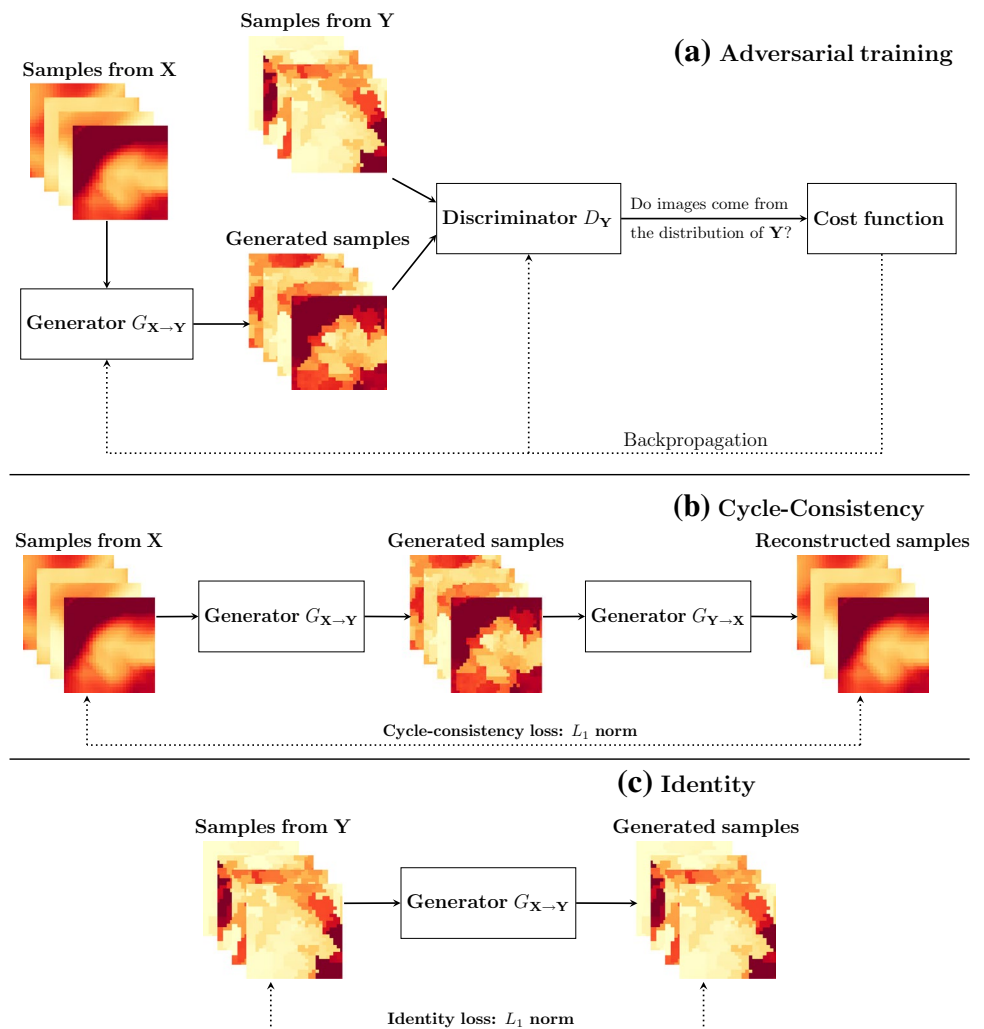
3.3.1 Adaptation of CycleGAN for MBC

The main idea of the proposed methodology, named MBC-CycleGAN, is to adapt the CycleGAN approach so that it turns daily maps of a simulated variable with spatial features inappropriate compared to a reference dataset, to more realistic maps. Here, MBC-CycleGAN is developed in the context of the “marginal/dependence” MBC category, i.e., correcting separately marginal distributions and dependence

relationships. In addition to marginal distributions, we consider the adjustment of spatial dependence structures. The algorithm is trained on a historical period (i.e., calibration) for which both climate simulations and reference datasets are available. Once the adversarial neural network has converged, adjustment of climate simulations over a projection period (e.g., a future time period) is performed using the pretrained algorithm. The MBC-CycleGAN proceeds as follows:

1. As MBC-CycleGAN belongs to the marginal/dependence category, univariate distributions of modeled climate variables are first corrected independently using a univariate BC method for both calibration and projection periods. In this study, the quantile-quantile (QQ) mapping method is used (Déqué 2007).
2. Then, quantile-quantile and reference data over the calibration time period are transformed to belong to $[0, 1]$ using a pointwise min-max normalization. For each grid cell, the minimum and maximum values from the refer-

Fig. 2 **a** Illustration of the adversarial training for the mapping function $G_{X \rightarrow Y}$, associated with the adversarial discriminator D_Y . D_Y encourages $G_{X \rightarrow Y}$ to generate outputs that are indistinguishable from the probability distribution of Y . A similar adversarial training is used for $G_{Y \rightarrow X}$ using D_X (not presented in this figure). In CycleGAN architectures, the mappings $G_{X \rightarrow Y}$ and $G_{Y \rightarrow X}$ are enforced to be cycle-consistent, i.e., **b** if an initial image from X is translated using $G_{X \rightarrow Y}$ and back again using $G_{Y \rightarrow X}$, the initial image should be obtained. **c** In addition, to ensure that images from X that already seem to be drawn from the distribution of Y are not modified too much, the identity property is used by enforcing $G_{X \rightarrow Y}$ applied to images from Y to resemble to initial inputs from Y (and vice versa for $G_{Y \rightarrow X}$). In our study, samples from X and Y are replaced by QQ outputs and references, respectively



ence during the calibration are taken to compute the normalization. The resulting daily maps are then given to a CycleGAN model to learn the transfer between the two distributions of images. Generators and discriminators are trained until the spatial distribution of the corrected maps stops improving. More details about the criteria used to evaluate spatial distributions are presented thereafter.

3. Once the CycleGAN model has been trained for the calibration period, the same pointwise normalization is performed for quantile-quantile data over the projection period, i.e., using the same minimum and maximum values from the reference during the calibration period. Normalized daily maps from quantile-quantile data in the projection period are translated in the normalized reference domain using the pretrained adversarial neural network. Then, the corrected outputs obtained are rescaled to physical values by applying the inverse of the pointwise min-max normalization used.
4. Finally, by taking advantage of the Schaake Shuffle technique (Clark et al. 2004), quantile-quantile data for the projection period obtained from Step 1 are reordered such that the rank structure of the data obtained from Step 3 is reproduced. This shuffling technique, already employed in a few multivariate bias correction methods (e.g., Vrac 2018; Cannon 2018; Mehrotra and Sharma 2019), permits here to obtain bias-corrected data with marginal properties from quantile-quantile outputs and rank dependence structure from CycleGAN outputs.

A summary of the successive steps in the form of a flowchart is provided in Fig. 3. More details about the different algorithmic steps are presented in Appendix 1.

3.3.2 Network architecture

To infer the weights for the cycle-consistency mapping loss λ_{cyc} and the identity mapping loss λ_{id} , preliminary tests have been conducted by checking a couple of combinations of weights and verifying that our optimization process improved the spatial structure of the climate simulations. With respect to these results (not shown), the weights have been chosen equal to $\lambda_{cyc} = 10$ and $\lambda_{id} = 1$.

Additionally, in this paper, we only present results obtained with a simple architecture for the CycleGAN neural networks. Our work being a proof of concept, we did not tune any further the architecture or the hyperparameters of the neural networks. However, the results presented later in Sect. 5 appear sufficient to illustrate the potential of CycleGANs for MBC. Schemes for the convolutional neural networks for both generators and discriminators are presented in Fig. 4. Architecture of generators for the mapping and inverse mapping are identical and are based on

deep convolutional layers (DCGAN, Radford et al. 2016). First, the daily maps, i.e. images of size 28×28 are given as inputs to the generators. Then, images flow through three 2D convolution layers with an increasing number of 3×3 filters (64–128–256). Two of them are performing convolutions that downsample input images to capture complex patterns at different scales. Then, two 2D transpose convolutional layers with a decreasing number of 4×4 filters (128–64) are used to perform inverse convolution operations and upsampling input data. Finally, one 2D convolution layer with one 1×1 filter is used to generate an output image of the same size as the initial one. Skip connections between convolution and transpose convolutional layers are used to ease the training of the CycleGAN network (He et al. 2016). All the other hyperparameters for the neural network architecture of the generators are detailed in Appendix 2.

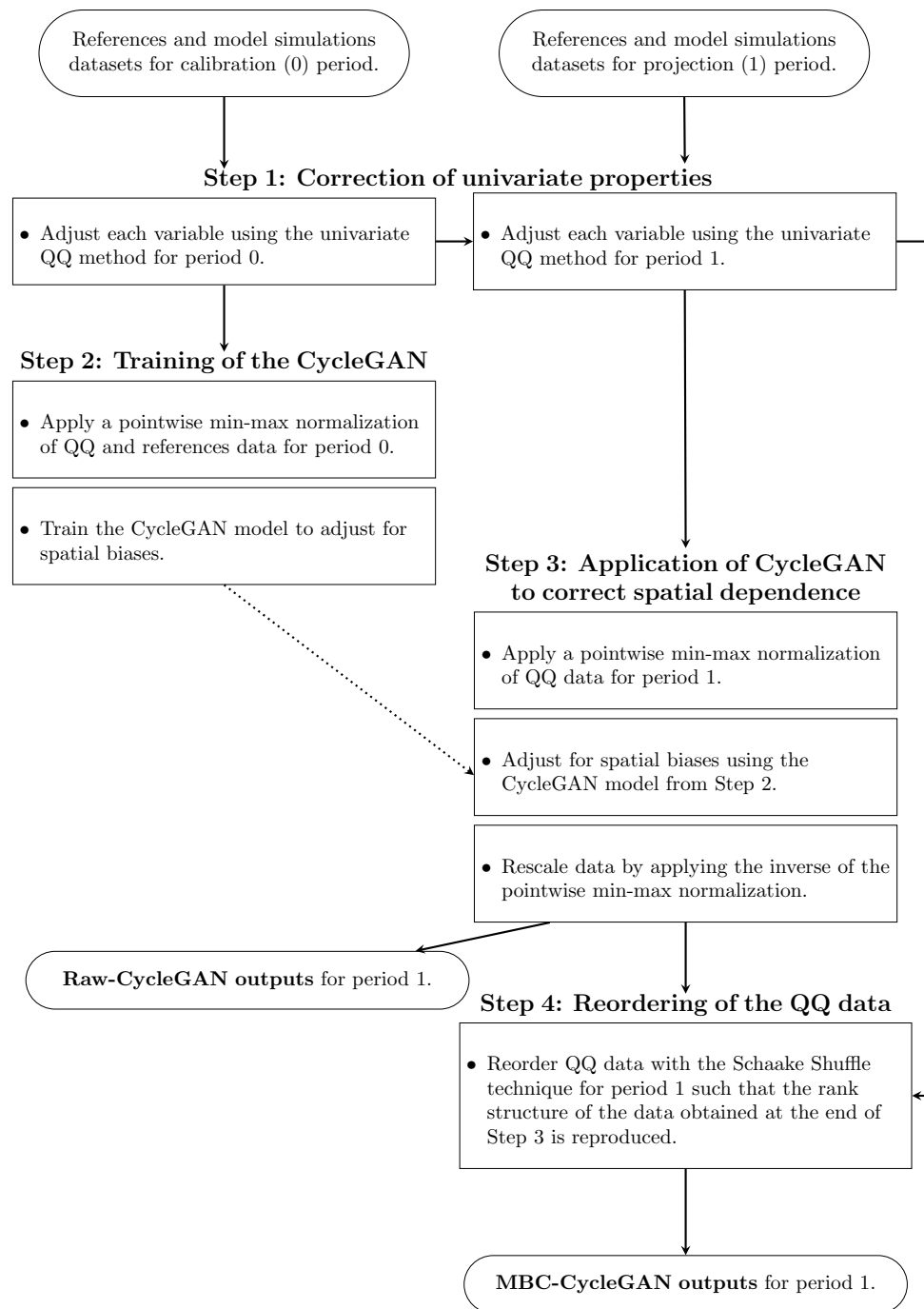
Concerning the discriminators, they take as well as inputs images of size 28×28 . Then, two 2D convolution layers with an increasing number of 3×3 filters (64–128) are used. Finally, outputs are flattened, i.e., are converted into a 1-dimensional array before being given to a fully connected layer (dense layer) that computes the sigmoid values (i.e., probabilities) for the classification of images.

The number of parameters is equal to 1,025,281 for each generator and 80,769 for each discriminator, bringing the total number of parameters to 2,212,100 for the whole CycleGAN architecture. Please note that each convolution and transpose convolutional layer used within the neural network architectures of both generators and discriminators includes a bias vector to fit. The number of parameters added by individual convolutional layers depends on its number of filters f_2 , the filter size (here 3×3) and also the number of filters f_1 from the previous convolutional layer. Adding an additional convolutional layer in a generator architecture with f_2 filters will add $(3 \times 3 \times f_1 + 1) \times f_2$ parameters. Hence, constructing a (deeper) neural network with more and more layers increases drastically the number of parameters to train. In order to keep an algorithm which is relatively fast to train while being stable, we decided not to add further layers to generators and discriminators architectures. For a concise summary of network architectures used, we refer to the Tables 3 and 4 in Appendix 2.

3.3.3 Training details

In this study, CycleGAN networks are trained using the Adam optimizer (Kingma and Ba 2017) with learning rates of $1e-4$ and $5e-5$ for the generators and discriminators, respectively. Please note that no grid search has been performed to determine optimal values of learning rates, and hence there is room for improvement. For the performance assessment of the CycleGAN model during training, the energy distance (Székely and Rizzo 2004; Székely and

Fig. 3 Flowchart for the MBC-CycleGAN method to adjust climate simulations for the projection period

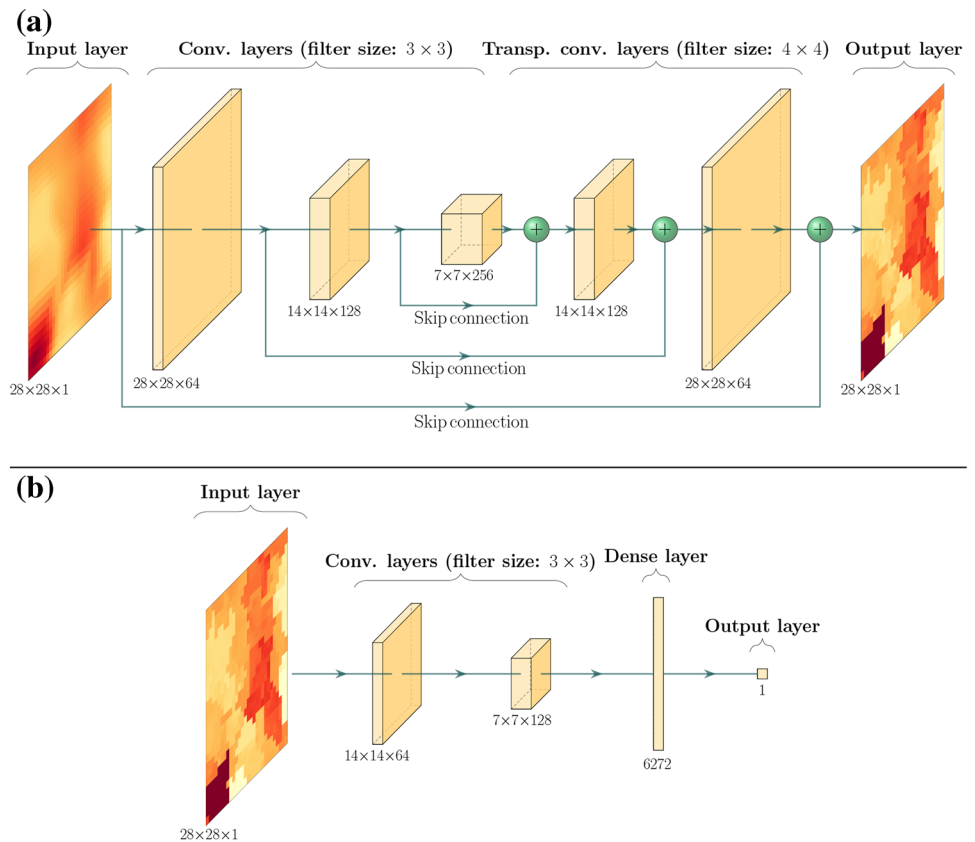


Rizzo 2013) is used. This metric, already used in the bias correction literature (e.g., Cannon 2018), permits to measure the statistical discrepancy between two multivariate distributions that are potentially in high dimension. Given two k -multivariate independent random vectors \mathbf{P} and \mathbf{Q} with multivariate probability distributions μ and ν respectively, the energy distance \mathcal{E} between the two distributions is:

$$\mathcal{E}(\mu, \nu) = \sqrt{2\mathbb{E}\|\mathbf{P} - \mathbf{Q}\| - \mathbb{E}\|\mathbf{P} - \mathbf{P}'\| - \mathbb{E}\|\mathbf{Q} - \mathbf{Q}'\|},$$

with \mathbb{E} denoting the expected value, \mathbf{P}' (resp. \mathbf{Q}') independent and identically distributed copy of \mathbf{P} (resp. \mathbf{Q}) and $\|\cdot\|$ the Euclidean distance. The corresponding energy statistic of \mathcal{E} between two k -dimensional statistical samples \mathbf{p} and \mathbf{q} can be computed as follows:

Fig. 4 Scheme of the convolutional neural networks for the **a** generators and **b** discriminators used in this study within the MBC-CycleGAN procedure. For each convolutional and transpose convolutional layers, the number of filters used is indicated by the third coordinate of their output size



$$\hat{\mathcal{E}}(\mathbf{p}, \mathbf{q}) = \left(\frac{2}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{m=1}^{n_2} \|\mathbf{p}_i - \mathbf{q}_m\| - \frac{1}{n_1^2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_1} \|\mathbf{p}_i - \mathbf{p}_j\| - \frac{1}{n_2^2} \sum_{l=1}^{n_2} \sum_{m=1}^{n_2} \|\mathbf{q}_l - \mathbf{q}_m\| \right)^{\frac{1}{2}},$$

where \mathbf{p}_i denotes the realizations of \mathbf{P} at the time step i across the k dimensions (and similarly for \mathbf{q}_m with \mathbf{Q}). The energy statistic goes to zero when the two multivariate samples \mathbf{p} and \mathbf{q} are drawn from the same distribution.

During training, computations of energy distances are performed every 10 epochs, i.e. each time that the CycleGAN has worked 10 times through the entire training dataset. Estimated energy distances $\hat{\mathcal{E}}$ are calculated on multivariate distributions of ranks between references and bias-corrected data. It permits to assess along the training the performance of the method to correct the whole spatial dependence structure of climate simulations. Computing energy distance using ranks instead of raw values allows the removal of the influence of univariate properties on the spatial relationships. The CycleGAN model that minimizes the energy distance on ranks during training is chosen for the correction of the projection period. Training 1000 epochs takes ~ 4 h on a single NVIDIA Tesla V100 GPU.

4 Design of experiments

For evaluation purposes, the proposed MBC-CycleGAN method is applied to adjust climate simulations outputs with SAFRAN data as references. Bias correction is performed on separate seasons in order to preserve seasonal properties. In the following, for sake of clarity, only the winter results are presented. Data are available for the 1979–2016 period (i.e. 3420 winter days), and need to be divided into a calibration period and a projection period to train and evaluate our algorithm. In accordance with common practices in machine learning, the 1979–2016 period is split as follows: 70% (2394 days) as training dataset and 30% (1026 days) as evaluation dataset. In this study, two different cross-validation methods—that differ in how calibration and projection periods are constructed—are used to evaluate our methodology.

4.1 Model output statistics (MOS) vs. Perfect prog (PP)

The first cross-validation method consists in drawing randomly the days that define the calibration and projection periods. As these periods are drawn randomly, the

potential climate change signal present in the data during the 1979–2016 period vanishes. Hence, for this cross-validation method, no changes in marginal and dependence properties are expected between the calibration and projection periods, allowing for the assessment of the method in a stationary context. We take advantage of this first stationary cross-validation technique to apply our method in both PP and MOS post-processing schemes for the adjustment of IPSL climate simulations. Implementing and evaluating both the PP and MOS approaches in such a validation context permits to determine which approach is better suited in our context of bias correction of climate simulations. For the MOS approach, MBC-CycleGAN is applied directly to IPSL data according to the 4 steps already described in Sect. 3.3. Concerning the implementation of the PP approach, the same procedure is applied but the CycleGAN model is trained in a slightly different way. Indeed, as already explained in Sect. 1, a PP approach consists in establishing the statistical relationships between large-scale predictors and local-scale predictands from observational or reanalysis data (including for the predictors) before applying them to climate model data. Hence, large-scale predictors temporally matching the SAFRAN dataset are needed to a PP approach. For this purpose, a new climate dataset is constructed for both temperature and precipitation as follows: initial local-scale SAFRAN data with 8 km × 8 km spatial resolution are upscaled using conservative interpolation on a large-scale grid of 32 km × 32 km spatial resolution. Then, the obtained large-scale data are regridded using bilinear interpolation to the initial grid of SAFRAN, allowing to train CycleGAN. It results in “biased” daily maps of temperature and precipitation (large-scale predictors) of the initial SAFRAN data (local-scale predictands), temporally matching the chronology of the SAFRAN time series. Using these new data—hereafter referred to as “low-resolution (LR) SAFRAN”—a CycleGAN model is trained for the implementation of the PP approach by learning the transfer of maps from 1d-BC large-scale predictors (QQ(LR SAFRAN)) to maps from local-scale predictands (SAFRAN). This trained model is then used to bias correct IPSL simulations over the projection period and, hence, evaluate the CycleGAN results in a PP context.

4.2 Nonstationarity investigation

To evaluate the nonstationary behavior of the proposed method, a second cross-validation method is defined, which consists in dividing the 1979–2016 period chronologically. By still defining the calibration and the projection periods based on the 70–30% split, it results in obtaining approximately the 1979–2005 and 2006–2016 portions as calibration and projection periods, respectively. Hence, the potential climate change signal between the calibration and

projection periods is not removed by the cross-validation technique. Within this second cross-validation method, IPSL simulations and SAFRAN references can potentially have different marginal and spatial dependence changes between calibration and projection periods. In this respect, depending on the level of agreement in changes between simulations and references, and how MBC methods account for these changes in their correction procedure, the quality of the correction for projection periods can possibly be different. Hence, to provide a global picture of the performances of the MBC-CycleGAN method in the nonstationary context, three bias correction exercises of climate data with different statistical changes are performed with respect to SAFRAN references:

- the correction of IPSL simulations that present different marginal and spatial properties from SAFRAN, and with potentially different changes than those from SAFRAN.
- the correction of LR SAFRAN dataset (presented above), whose marginal and spatial properties as well as their changes are in line with those from SAFRAN.
- the correction of a third dataset called IPSLbis (presented below) that presents different marginal and spatial properties from SAFRAN, but for which their changes are in line with those from SAFRAN.

For the sake of clarity, a summary of the different attributes of the three datasets to correct is presented in Table 1.

LR SAFRAN dataset already presented above has, by construction, little bias with SAFRAN references: its biases are only due to the interpolation technique used to obtain data with a lower resolution. Hence, statistical changes between the calibration and projection periods for LR SAFRAN are in line with those from the SAFRAN dataset. Adjusting LR SAFRAN data for the projection period permits to assess if the MBC-CycleGAN method is able to reproduce the changes from the reference in the correction. Also, the LR SAFRAN dataset presents the particularity of being synchronous in time with references. Hence, in addition to evaluate the proposed method in terms of distributional properties, which is not considered as sufficient to identify successful bias correction techniques (Maraun 2016), this pairwise correspondence between predictors and predictands offers the possibility to directly compare corrected daily maps with those from the references using classic forecast verification statistics.

As IPSL simulations compute a different combination of variability and warming than those from the SAFRAN reanalysis, IPSL model and SAFRAN references are likely to present disagreeing changes in their statistical (marginal and dependence) properties between calibration and projection periods. To evaluate the influence of these potential disagreeing changes on the performance of correction of the

proposed method, we constructed the third dataset, referred to as “IPSLbis”, for the projection period only. IPSLbis is specifically constructed so that its marginal and dependence changes between calibration and projection periods are in line with those from the reference. In order to ease the comparison of results with the first bias correction exercise, we forced IPSLbis to have the same changes as LR SAFRAN. This is reached by using a two-step procedure that takes advantage of a nonstationary quantile mapping technique for marginal changes (CDF-t, Vrac et al. 2012) and a matrix-recorrelation technique for dependence changes (Bárdossy and Pegram 2012). More details about the generation of the IPSLbis data can be found in Appendix 3 and a detailed evaluation of the evolution of statistical properties of the different dataset between the calibration and projection period is provided in Appendix 4. In particular, results presented in Appendix 4 indicate that, as expected, changes in spatial structures from SAFRAN references are (globally) in agreement with those from LR SAFRAN for both temperature and precipitation. However, concerning changes in spatial structures for IPSL simulations, conclusions are not the same depending on the physical variable. While, for temperature, simulated changes of spatial correlations are partially in line with those from LR SAFRAN, IPSL model presents discrepancy of changes for precipitation. Globally, the construction of IPSLbis with the two-step procedure described in Appendix 3 permits to impose to IPSL data spatial changes for both temperature and precipitation that are in line with those from LR SAFRAN.

4.3 Comparisons to existing MBCs: R^2D^2 and dOTC

Although evaluating the performance of correction for IPSL simulations is of primary interest, applying our method on these three datasets (IPSL, IPSLbis, LR SAFRAN) permits to assess gradually how well our method is performing depending on the biases present in the dataset to correct. Note that, as IPSL and IPSLbis data during calibration are identical, there is no need to train for a second time the CycleGAN model for IPSLbis data: the CycleGAN model trained with IPSL data can be used directly to adjust IPSLbis simulations for the projection period. In addition, two MBCs with different assumptions about nonstationarity are applied for comparison using the second cross-validation method: the “Rank Resampling For Distributions And Dependences” (R^2D^2 , Vrac and Thao 2020) and the “Dynamical Optimal Transport Correction” (dOTC, Robin et al. 2019) methods.

R^2D^2 , developed in the context of marginal/dependence category, relies on an analogue-based method that allows to resample ranks from a reference dataset according to some conditioning information and reconstructs dependence structure of the simulated time series. The information to condition the analogues can be multivariate by considering, for

example, a set of variables to be corrected at a given time t . Conditioning for the ranks resampling can also be extended to ranks sequences, i.e. conditioning by not only one but several lagged time steps. Please note that, for the different implementations of R^2D^2 in this study, the multivariate conditioning used includes 4 grid points that cover uniformly the region of interest. In addition, 5 lagged time steps are used for the conditioning, as it has been found to stabilize the R^2D^2 method (not shown). Also, the QQ method is used to correct the marginal properties for R^2D^2 outputs.

Concerning the dOTC method, it was developed in the all-in-one category, i.e., adjusting the univariate distributions and dependence structures at the same time. The dOTC method takes advantage of the optimal transport theory to construct a multivariate transfer function, named a transport plan, for the adjustment of climate simulations with respect to references while minimizing an associated cost function. This particular transfer function permits to link, through conditional laws, all the multivariate elements from the biased multivariate distribution to their corrections. Corrections are then derived by drawing directly from these conditional laws to obtain the bias corrected data.

Both R^2D^2 and dOTC methods are applied according to the spatial-dimensional configuration (hereinafter referred to as “Spatial-”), where all the 784 time series for a particular physical variable are corrected jointly. While R^2D^2 assumes spatial dependence structures (i.e., the rank correlations, or copulas) to be stable in time, the dOTC method makes the hypothesis of nonstationarity of the dependence structure between the calibration and the projection periods, which allows for taking into account the changes of the model (e.g., due to climate change) in the bias correction procedure. Intercomparing the results from both Spatial- R^2D^2 and Spatial-dOTC for adjusting spatial dependence structure of climate simulations with those from MBC-CycleGAN allows to better assess how the proposed method performs in a nonstationary context.

5 Results

In this section, analyses are presented for the winter season (December, January and February) only. CycleGAN models are trained during the calibration period and selected such that energy distances on ranks are minimized. All evaluations are performed on the projection period for the corrected outputs obtained from the two cross-validation methods and results are compared to those from the reference dataset. For bias-corrected precipitation time series, thresholding of 1 mm is applied before evaluation to replace values lower than 1 mm by 0. Bias correction outputs from the first and second cross-validation methods are evaluated in terms of both marginal and spatial properties. Analyses of

temporal properties are only provided for outputs from the second cross-validation method, in which calibration and projection periods are divided chronologically and hence do not distort temporal properties, contrary to the first cross-validation method that randomly defines these periods. To assess the potential benefits of considering spatial aspects in the correction procedure, the univariate QQ method (Déqué 2007) is also included in the study as a benchmark.

5.1 MOS vs. PP

5.1.1 Training of MBC-CycleGANs

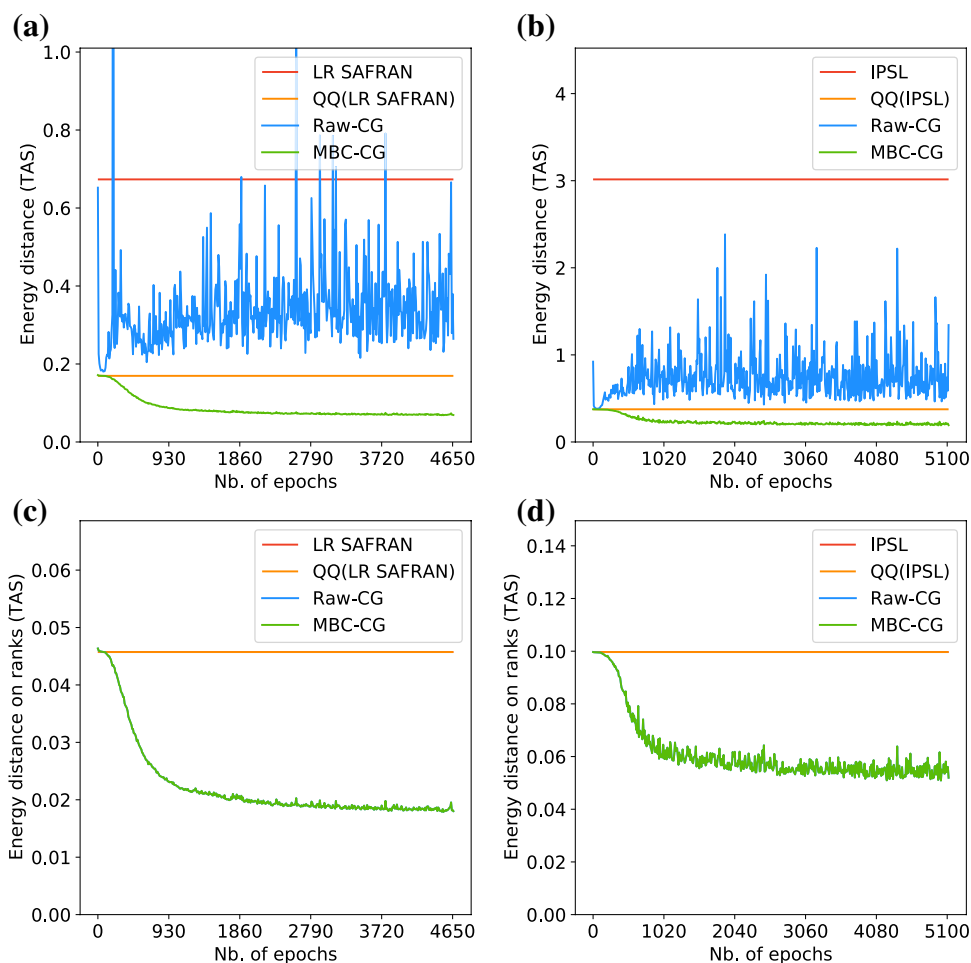
Figure 5 shows energy distances with respect to SAFRAN references for temperature computed on physical values (Fig. 5a, b) and ranks (Fig. 5c, d) for LR SAFRAN, plain IPSL simulations, 1d-QQ, and MBC-CycleGAN (MBC-CG) outputs during the training on the calibration period. In addition, results for Raw-CycleGAN (Raw-CG) are presented. Differences between Raw-CG and MBC-CG only lie in their marginal properties: while Raw-CG corresponds to the outputs obtained from the CycleGAN after denormalization at

the end of Step 3, MBC-CG is the combination of the spatial structure from Raw-CG and univariate properties from QQ outputs (see the flowchart provided in Fig. 3). The results for precipitation are presented in Fig. S1 of the Supplement.

Clearly, Fig. 5a, b show large energy distances computed on physical values of temperature for LR SAFRAN and IPSL datasets, indicating some biases on spatial structures for those dataset with respect to SAFRAN references. Adjusting marginal properties with the univariate QQ method reduces values of energy distance computed on physical values, highlighting the influence of marginal properties on spatial features. Correction of the spatial dependence structure provided by MBC-CG occurs relatively quickly, with energy distances on physical variables reduced by 2 compared to QQ after approximately 1000 epochs for both PP and MOS approaches. However, for Raw-CG, marginal properties generated by the inverse pointwise min-max normalization do not seem to improve values of energy distances, which justifies the post-processing of univariate properties adopted in the MBC-CycleGAN method with the Schaake Shuffle.

Figure 5c, d show that computing energy distances on ranks for temperature removes the influence of univariate

Fig. 5 Values of the energy distances with respect to SAFRAN reference for temperature computed on **a, b** physical values and **c, d** ranks during the training of MBC-CycleGAN. Results are shown for the different datasets involved in **a, c** the Perfect Prognosis approach and **b, d** the MOS approach. Please note that results of QQ and low-resolution SAFRAN (resp. IPSL) for ranks are the same. Red and orange lines are therefore superimposed in **c** (resp. **d**). This remark also applies for Raw-CycleGAN (blue line) and MBC-CycleGAN (green line)



properties on spatial features. Energy distances for both LR SAFRAN and IPSL with their respective QQ corrections are indeed the same (Fig. 5c). The same remark holds for MBC-CG and Raw-CG energy distances on ranks that have, by construction, similar spatial dependence structures. As explained in Sect. 3.3.3, the CycleGAN model that minimizes the energy distance on ranks of MBC-CycleGAN outputs is selected.

For precipitation (Fig. S1), the same conclusions hold, indicating a relative ability of the CycleGAN to adjust spatial dependence structure of precipitation fields. Nevertheless, contrary to temperature, one should remark that energy distances on ranks are different for LR SAFRAN, IPSL and their respective QQ corrections (Figs. S1c, d), which is specific to precipitation variables that can contain several null values for dry events. Indeed, ranks are computed here such that, when tied values are encountered, the minimum value of rank is attributed to each tied value. The combination of the correction with the QQ method and the thresholding for precipitation below 1 mm could modify the frequency of dry events, which could result in obtaining different rank structures, and hence, mechanically, different energy distances with respect to SAFRAN references. This mechanism is also obtained between MBC-CG and Raw-CG (Figs. S1c, d), that present different energy distances due to the difference of dry events.

5.1.2 Univariate distribution properties

Once the CycleGAN models have been selected for both the PP and MOS approaches, the corrections of IPSL simulations can be performed for the projection period. First, bias-corrected data are evaluated in terms of univariate statistics. For temperature and precipitation, differences of mean values between the bias corrected data and the SAFRAN references are computed at each grid cell. For temperature mean, absolute differences are computed, while for precipitation variables having absolute zeros, relative mean differences are more appropriate. Maps of differences with respect to the reference—for IPSL simulations and the bias-corrected data—are displayed in Fig. 6 for both temperature and precipitation. The mean absolute error (MAE) with respect to the reference dataset is also reported on each map. For more results on marginal properties, maps of standard deviation relative differences for both physical variables are also provided in Fig. S2 of the Supplement.

For both temperature and precipitation, the maps for the IPSL model (Fig. 6c, d) present large values of mean differences with respect to the SAFRAN map (Fig. 6a, b) and highlight the need to adjust univariate properties of simulations. Maps provided by 1d-QQ outputs (Fig. 6e, f) indicate that, as expected, the univariate method globally improves marginal properties at each individual site. In agreement

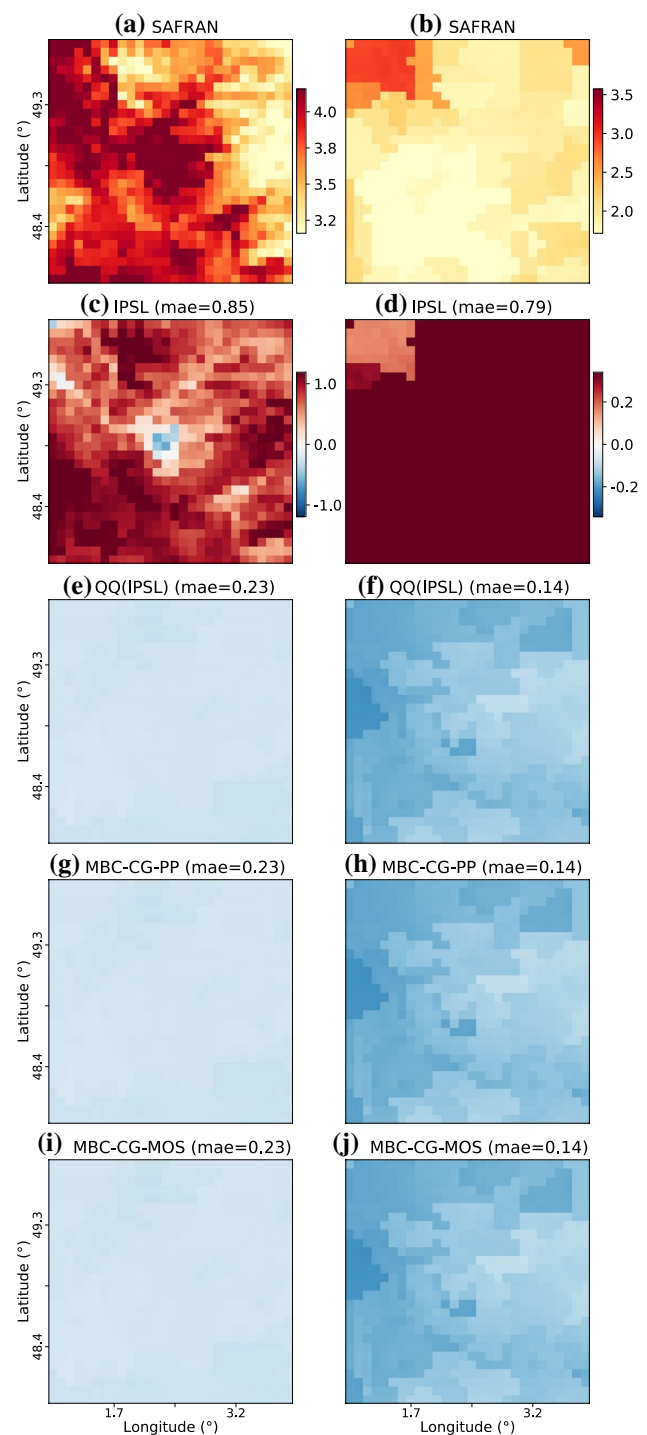


Fig. 6 Mean differences for **c, e, g, i** temperature and relative mean differences for **d, f, h, j** precipitation computed at each grid cell between SAFRAN reference and the different datasets (plain IPSL, QQ, MBC-CycleGAN-PP and MBC-CycleGAN-MOS outputs) during winter over the projection period. Note that the color scales between panels **c, e, g, i** and **d, f, h, j** are not the same to better emphasize intensities of values for the two physical variables. Maps of daily mean for SAFRAN references are also shown for **a** temperature and **b** precipitation

with the properties of the marginal/dependence MBC methods, maps for MBC-CG for PP (MBC-CG-PP, Fig. 6g, h) and MOS (MBC-CG-MOS, Fig. 6i, j) are exactly the same as those from the 1d-QQ method. Indeed, by construction, the univariate distribution properties are identical between QQ and MBC-CycleGAN outputs, regardless of the spatial correlation adjustments. Although MBC-CG-PP and MBC-CG-MOS do not use the same data for the training of the CycleGAN to adjust spatial features, same marginals are taken from the QQ outputs of IPSL data, which results in obtaining the same univariate properties between the three corrections.

5.1.3 Spatial correlations

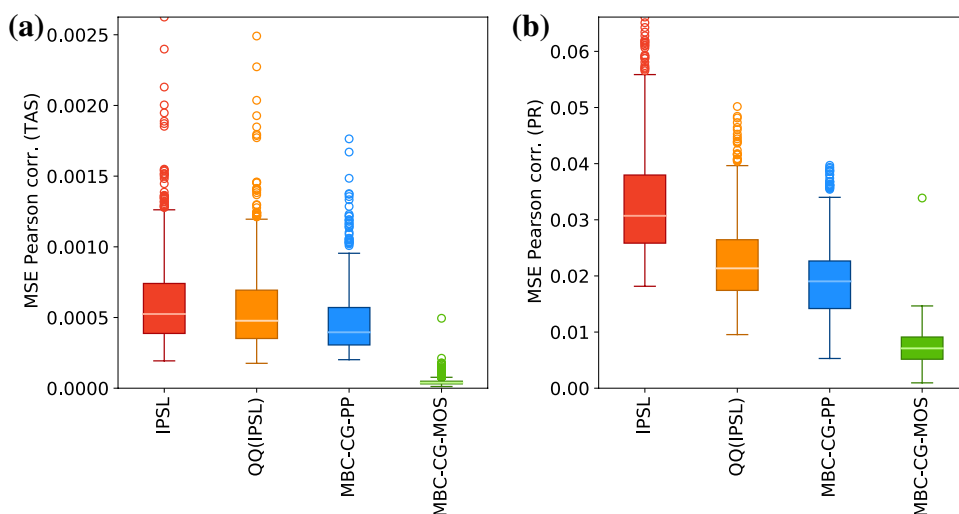
Quality of the corrections in terms of spatial correlations is now assessed. For each grid cell, spatial dependencies are evaluated for temperature and precipitation by computing Pearson pairwise correlations between the cell of interest and each of the remaining 783 grid cells over the region of Paris for the different climate datasets. The biases of these 783 spatial Pearson correlations are then summarized by computing the Mean Squared Error (MSE) with the corresponding 783 correlations computed for the references. By computing the MSE values for each grid cell, 784 MSE values are obtained for each climate dataset and can be intercompared from one dataset to another. Figure 7 shows the boxplots of the MSE values obtained for both temperature and precipitation for the plain IPSL simulations and BC outputs. For both variables, the boxplots for the IPSL simulations indicate strong values of MSE with respect to SAFRAN references. For QQ outputs, only slight reductions of MSE of spatial correlations are observed compared to those from IPSL, indicating that QQ globally conserves the spatial structure of the IPSL model. This result could have been expected, as, for each site, the univariate QQ method

does not modify (too much) rank sequences of the simulated time series. The slight improvement of spatial statistics, which is greater for precipitation (Fig. 7b) than temperature (Fig. 7a), is in fact mainly attributable to the correction of univariate properties provided by the QQ method. Concerning MBC-CycleGAN, the PP and MOS approaches display different performances in adjusting the spatial properties of simulations. Boxplots of MSE for MBC-CG-MOS indicate clear improvements of spatial correlations with respect to QQ outputs for both temperature and, to a lesser extent, precipitation. However, results for MBC-CG-PP show less pronounced improvements, suggesting a failure for the MBC-CG-PP approach to adjust spatial properties. This difference of performance for the PP approach indicates that, although CycleGAN models are able to learn the spatial relationships between large-scale predictors (LR SAFRAN) and local-scale predictands (SAFRAN) during the training of the algorithm, as previously shown in Figs. 5 and S1, these relationships do not prove to be suited for adjusting IPSL simulations. Indeed, simulated large-scale predictors seem here to present too large biases with respect to LR SAFRAN to make the CycleGAN fitted in a PP context applicable to the IPSL simulations. Hence, the perfect-prognosis approach should be discarded in our context of bias correction of climate simulations. Therefore, in the following, only the MOS approach of MBC-CG is further investigated.

5.2 MBC-CycleGAN in the nonstationary context

In the following, analyses are presented for the application of the MBC-CycleGAN method with the MOS approach in a nonstationary context using the second cross-validation method. Results for the correction of the three datasets - IPSL, IPSLbis and LR SAFRAN - with different changes in marginal and dependence properties between the calibration and projection periods are provided.

Fig. 7 Boxplots of mean squared errors of Pearson spatial correlations computed at each grid cell for **a** temperature and **b** precipitation over the projection period. Results are shown for plain IPSL, QQ, MBC-CycleGAN-PP and MBC-CycleGAN-MOS outputs



5.2.1 Univariate distribution properties

Similarly to the first cross-validation method, univariate properties are evaluated using mean differences computed at each grid cell. Figure 8 shows, for the bias-corrected outputs from the three bias correction exercises, the maps of temperature mean differences with respect to SAFRAN references. Maps for precipitation relative mean differences are presented in Fig. S6 of the Supplement. For information purposes only, standard deviation relative mean differences for temperature and precipitation are also displayed in Figs. S7 and S8, respectively.

For temperature, values of IPSL and IPSLbis mean differences (Fig. 8b, c) are high, indicating strong biases of temperature mean with respect to the SAFRAN reference dataset (Fig. 8a), although less pronounced for IPSLbis. This was somehow expected since IPSLbis data are specifically constructed to mimic the SAFRAN changes in terms of marginal (and dependence) properties. It results here in having IPSLbis temperature means closer to those from SAFRAN reference for the projection period. Map for LR SAFRAN (Fig. 8d) shows small differences with the reference. Clear improvements of the temperature mean are provided by the QQ method for each of the bias correction exercises (Fig. 8e–g). Nevertheless, quite interestingly, QQ method provides less pronounced improvements for IPSL data (Fig. 8e), suggesting a degrading effect on results of correction when changes of marginal properties between calibration and projection periods for the climate data to be corrected are not in agreement with those from the references. With regard to the performances of the MBC methods, MBC-CycleGAN presents exactly the same results as the QQ method (Fig. 8h–j), in agreement with the marginal/dependence MBC properties. For Spatial- R^2D^2 ($S-R^2D^2$), very slight modifications of the marginal mean values provided by QQ are observed (Fig. 8k–m), due to the use of the multivariate conditioning to adjust spatial dependence structure (Vrac and Thao 2020). Concerning Spatial-dOTC (S -dOTC), the corrected outputs for IPSLbis (Fig. 8o) and LR SAFRAN (Fig. 8p) present results similar to those obtained for QQ and MBC-CycleGAN. However, it is worth mentioning that, for the correction of IPSL, S -dOTC (Fig. 8n) slightly improves marginal properties ($MAE=0.37$) compared to those obtained from QQ outputs ($MAE=0.42$).

For precipitation relative mean differences (Fig. S6), the same conclusions hold for each (M)BC method, indicating no particular influence of the variable to correct on the results of the marginal statistics adjustment.

5.2.2 Spatial correlations

We now evaluate the ability of MBC-CycleGAN to adjust spatial dependence. First, as for the Sect. 5.1, we compute

MSE of spatial Pearson correlations for both temperature and precipitation. Figure 9 displays the results with box-plots for the different datasets to correct and their adjusted outputs. Scatterplots of MSE values with respect to QQ outputs are presented in Fig. S9 to better assess the potential benefits of using MBC methods relative to univariate ones. For temperature (Fig. 9a), the positive values of MSE for IPSL suggest biases with respect to the SAFRAN references, illustrating the necessity to correct spatial properties of the model before using it in subsequent analyses. For IPSLbis, MSE values are slightly smaller, but still indicates strong differences of spatial correlations with respect to the references. The difference of results between IPSL and IPSLbis highlights that discrepancies of changes with the references can potentially have a non-negligible effect on spatial properties; in fact, reducing those discrepancies as it is done with the generation of IPSLbis leads here to reduce biases in spatial correlations. Concerning LR SAFRAN, MSE values are small, suggesting that upscaling the reference dataset deteriorates only slightly its spatial structure. By simply correcting univariate distributions, the three QQ outputs do not present a particular improvement of temperature MSE values. Clear improvements of the spatial correlation structures are provided by the MBC-CycleGAN method for the adjustment of IPSL, IPSLbis and LR SAFRAN, although some differences of performances are observed between the three corrected outputs. Temperature MSE values are indeed closer to 0 for the correction of LR SAFRAN than for the correction of IPSLbis and IPSL, for which similar results are obtained.

Concerning Spatial- R^2D^2 , the corrections of IPSL and IPSLbis provide major improvements in adjusting the spatial correlations. In particular, better results are obtained for the correction of IPSLbis. However, with regard to the Spatial- R^2D^2 outputs with LR SAFRAN, the benefits provided by R^2D^2 are smaller, as not all of the spatial correlations are improved. This result can better be seen in Fig. S9e. This contrasted performance for the R^2D^2 method appears in the context of the correction of LR SAFRAN that already presents small spatial biases with respect to SAFRAN references. The correction obtained for LR SAFRAN suggests that the R^2D^2 method is too constrained by the selected conditioning to find an appropriate collection of analogues for the projection period of this specific dataset.

For Spatial-dOTC outputs, results present low MSEs values for each bias correction exercise, indicating that spatial correlations are satisfyingly corrected by this method. Nevertheless, the adjustments are slightly better for the corrected output of IPSL than for those for IPSLbis, which may be confusing here. Indeed, as dOTC is specifically designed to take into account the changes of the data to adjust in the correction procedure, better results for IPSLbis, for which changes of spatial correlations are in line with those from SAFRAN references, would have been expected. The great

Fig. 8 Mean differences for temperature with SAFRAN reference for BC methods using as inputs **b, e, h, k, n** IPSL, **c, f, i, l, o** IPSLbis and **d, g, j, m, p** LR SAFRAN data. Results are shown during winter over the projection period for IPSL, IPSLbis, LR SAFRAN, QQ, MBC-CycleGAN, Spatial-R²D² and Spatial-dOTC datasets. The map of daily mean for SAFRAN references is also shown for temperature (**a**)

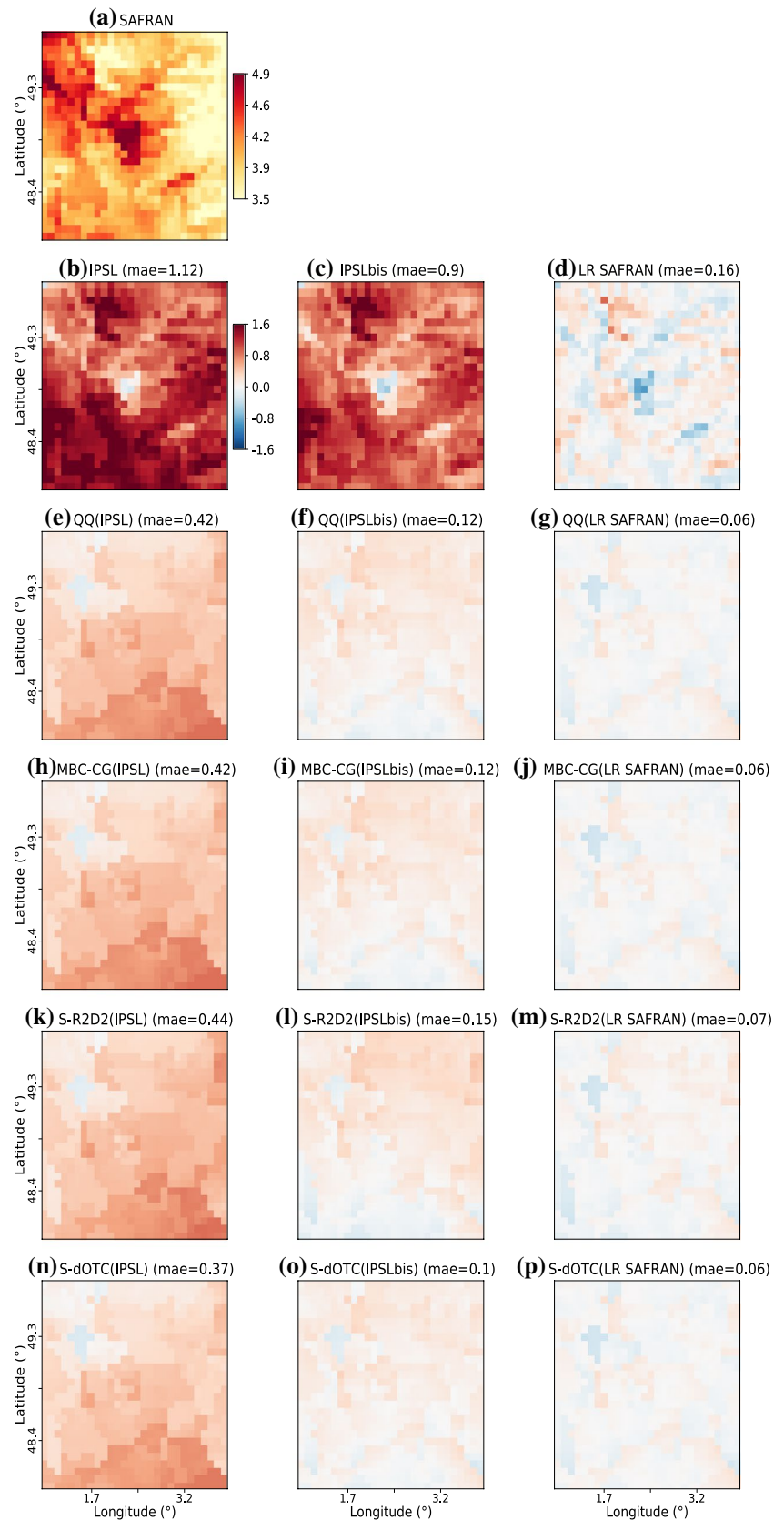
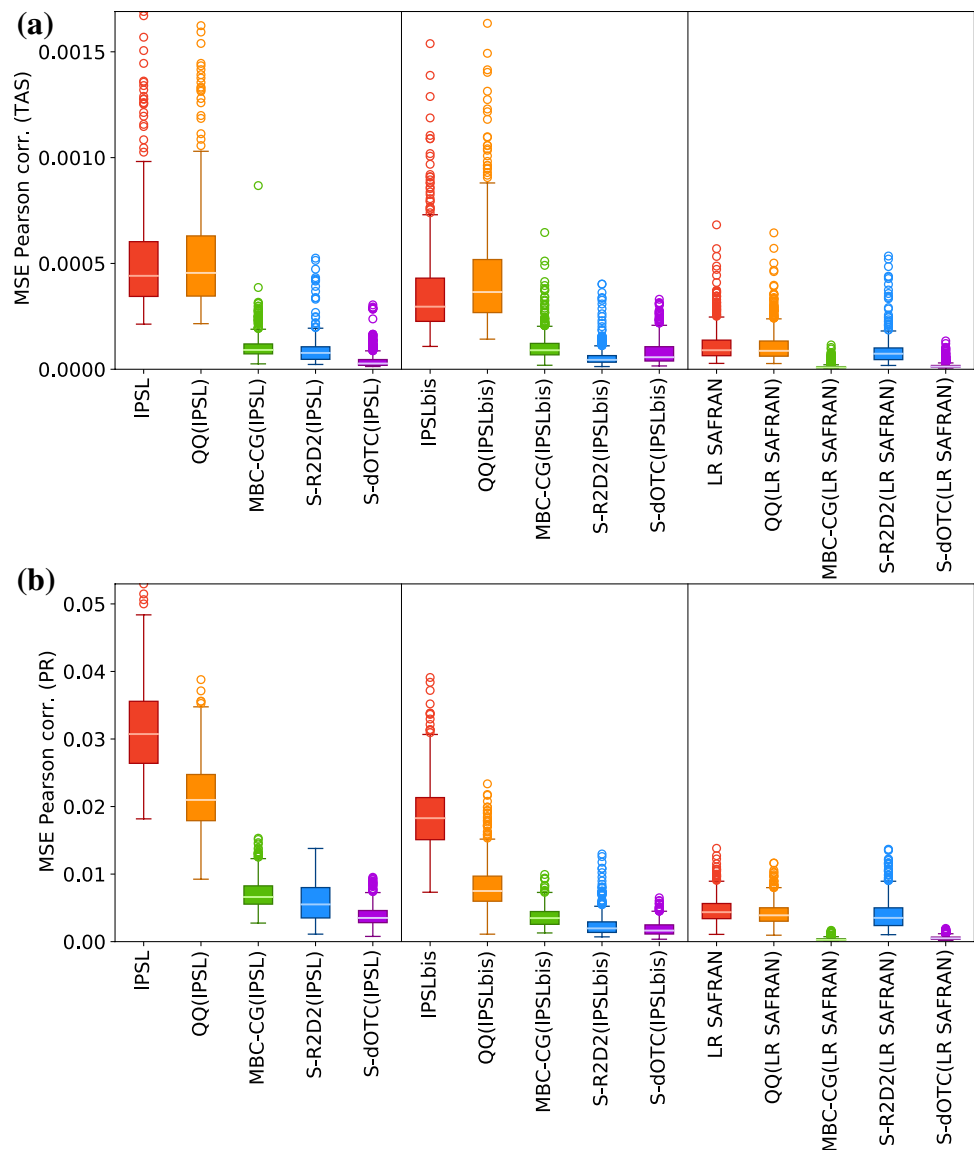


Fig. 9 Boxplots of mean squared errors of Pearson spatial correlations computed at each grid cell for **a** temperature and **b** precipitation over the projection period. Results are shown for IPSL, IPSLbis, LR SAFRAN, QQ, MBC-CycleGAN, Spatial-R²D² and Spatial-dOTC datasets



performance of dOTC to correct spatial correlations for IPSL could be due to the fact that, as explained in Appendix 4, IPSL simulated changes for temperature are not in total disagreement with those from SAFRAN, and hence there is no strong discrepancy of changes affecting the corrections.

For precipitation (Fig. 9b), the same conclusions as those drawn for temperature hold. Nevertheless, quite interestingly, IPSL and IPSLbis data present even larger differences of MSE values. This shows the effects on spatial correlations of the strong discrepancies of precipitation changes between the IPSL model and the references observed in Appendix 4: reducing this discrepancy of marginal and spatial changes with IPSLbis decreases significantly the biases on spatial correlations. In contrast with temperature, these differences of spatial correlations for precipitation between IPSL and IPSLbis are significant enough to spread itself in

the bias-corrected outputs: for each of the BC methods, the corrected outputs for IPSLbis present systematically lower MSE values compared to the corrections of IPSL.

To better assess spatial structure adjustments brought by MBCs, the calculation of energy distances between the bias-corrected time series and the references are performed for each physical variable according to two different multivariate distributions:

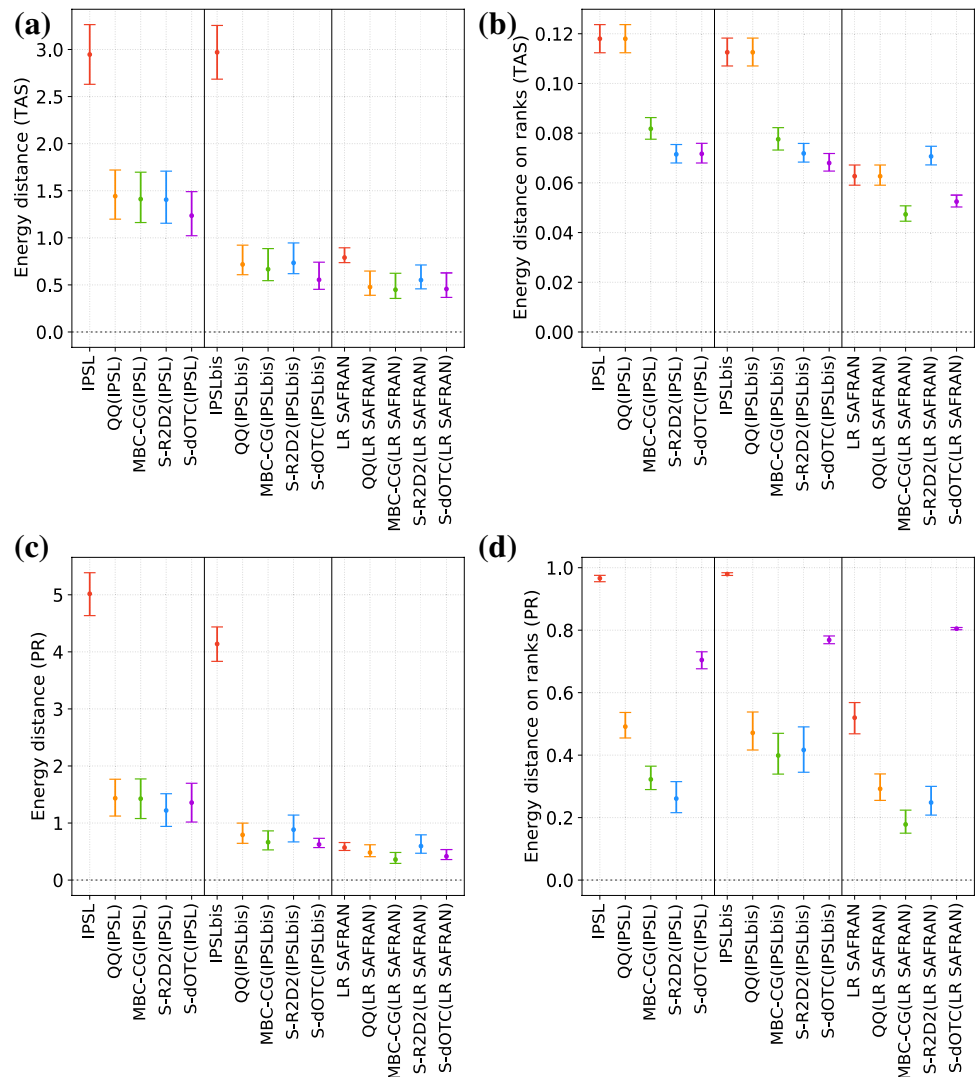
- on values of the physical variable directly over the whole region of Paris to assess differences of spatial properties (i.e., including both the marginals and their dependence);
- on ranks of the physical variable over the whole region of Paris to assess differences of spatial dependence structures (i.e., without the influence of marginal properties).

Values of energy distances are estimated using a bootstrap method. It consists for each dataset in (i) sampling (with replacement) daily fields, (ii) computing the energy distance on the bootstrapped dataset, and (iii) repeating the previous two steps 1000 times to construct the bootstrap sampling distribution. From this bootstrap sampling, distribution is deduced by the bootstrap estimator (mean of the 1000 energy distances obtained) and a 90% bootstrap sampling interval to provide uncertainty bands of the estimated distance. Results for temperature and precipitation are displayed in Fig. 10. The closer the values of the energy distances are to 0, the closer the spatial properties of the outputs are to the one of the reference data.

For temperature, the two estimators of energy distances on physical values (Fig. 10a) and ranks (Fig. 10b) for IPSL and IPSLbis data are quite high compared to those for LR SAFRAN, which is in agreement with the differences of spatial properties already observed between these datasets and the references in Fig. 9. For the three QQ outputs, while

energy distances on physical values are lower (Fig. 10a), similar energy distances on ranks as those from the dataset to correct are obtained (Fig. 10b). It highlights again that, although the QQ method adjusts the univariate distributions, it is not supposed to modify rank sequence of time series, and therefore spatial dependence structures, during the correction procedure. With regard to the three MBC methods for the correction of IPSL, dOTC performs slightly better on raw values (Fig. 10a) than MBC-CycleGAN and R²D², for which comparable results are obtained. For energy distances computed on ranks (Fig. 10b), dOTC and R²D² produce similar results. Slightly poorer performances of MBC-CycleGAN are obtained compared to the two other MBC methods, although strongly improving the spatial dependence structures of IPSL simulations. Note that, while bootstrap sampling intervals of energy distances on temperature values are overlapping for the three MBC methods, it is less the case for energy distances on temperature ranks, thereby permitting to determine with more confidence the best method for

Fig. 10 Values of the estimated energy distances with respect to the reference SAFRAN for temperature (a, b) and precipitation (c, d) computed on physical values (a, c) and ranks (b, d) during the projection period. Results are presented for IPSL, IPSLbis, LR SAFRAN, QQ, MBC-CycleGAN, Spatial-R²D² and Spatial-dOTC outputs. Estimates are evaluated using a bootstrap method (1000 replicates) that independently samples with replacement the daily fields from datasets. Note that same sequences of random days (i.e., same sampled days) are used to estimate values of energy distance for the different datasets. Error bars shows 90% bootstrap sampling intervals



the adjustment of spatial dependence properties. However, it must be mentioned that results of energy distances between the three MBCs are very close. Consequently, differences in performances between MBCs might not be significant. Concerning the correction of IPSLbis, best performances are provided by dOTC for both multivariate distributions. For multivariate distributions with raw values, MBC-CycleGAN is second best, while being third for rank dependence structure. This swap of performances between raw values and ranks for MBC-CycleGAN and R^2D^2 must be analyzed with caution as differences of estimated energy distances between the two MBC methods are again very small and thus might not be significant. This swap can however be explained by both the strong influence of marginal properties on energy distances and the slight deterioration of marginal properties provided by R^2D^2 compared to the QQ outputs, already mentioned in Sect. 5.2.1. For the corrections of LR SAFRAN, MBC-CycleGAN performs best and dOTC second best, with a more significant difference of performance for estimated energy distances evaluated on rank values (Fig. 10b).

For precipitation (Fig. 10c, d), conclusions similar to those obtained for temperature can be drawn for IPSL, IPSLbis and LR SAFRAN outputs. However, conclusions are slightly different for QQ and the MBCs. As already explained in Sect. 5.1, QQ modifies the frequency of dry events and consequently changes the rank dependence structure of precipitation, which results here in an improvement of spatial energy distances on ranks for the 1d-QQ corrections of IPSL, IPSLbis and LR SAFRAN. Concerning the performances of the three MBCs for IPSL, R^2D^2 performs best on energy distances for both raw values and ranks, while MBC-CycleGAN produces reasonable results, in particular for the adjustment of the rank dependence structure of precipitation. The dOTC method produces results that are clearly unsatisfactory concerning the rank dependence structure of precipitation. Instead of improving the rank dependence structure, dOTC correction strongly degrades it. This underperformance is in fact due to the presence of too many wet events in the corrections provided by dOTC (not shown) compared to the references, which mechanically largely affects the quality of its rank dependence structure for precipitation. For the same reason, this underperformance on precipitation rank dependence structure is also observed for the adjustments of IPSLbis and LR SAFRAN with dOTC. For IPSLbis, estimated energy distances on ranks are similar between MBC-CycleGAN and R^2D^2 . Note here that similar values of energy distances do not necessarily imply that their spatial dependence structures are similar. Concerning LR SAFRAN corrections, MBC-CycleGAN again outperforms both dOTC and R^2D^2 algorithms according to estimated energy distances on raw values and ranks.

5.2.3 Temporal structure

In this section, bias-corrected data are evaluated relative to temporal properties. As a reminder, MBC-CycleGAN and dOTC methods have been specifically implemented to only adjust marginal and spatial properties of climate simulations. Similarly, the R^2D^2 algorithm is applied to adjust marginal and spatial features but, contrary to the two other methods, it also takes into account (part of) the temporal dependence properties through the multivariate conditioning chosen for its implementation, as previously explained in Sect. 4. In theory, this choice of conditioning dimensions allows R^2D^2 to partially recover temporal properties of the reference dataset (Vrac and Thao 2020). Adjusting spatial coherence necessarily modifies the rank sequences of the initial time series during the correction procedure (e.g., Vrac 2018). It is hence interesting to quantify how strong those modifications are depending on the MBC method, whether temporal properties are taken into account in the correction procedure or not. Evaluation of temporal properties is performed by computing 1-d lag Pearson autocorrelations (AR1) at each grid cell for both temperature and precipitation. The resulting maps of differences with respect to SAFRAN references for the different BC outputs are presented in Fig. 11 (resp. Fig. S10) for temperature (resp. precipitation).

For temperature, IPSL shows relatively low values of AR1 differences (Fig. 11b), indicating that temporal properties for temperature are relatively in line with those from the SAFRAN references (Fig. 11a). A similar differences map is provided by IPSLbis outputs (Fig. 11c). In fact, IPSLbis temporal properties are inherited from IPSL outputs: even in a high-dimensional context, the two-step procedure—and in particular, the matrix-recorrelation technique—used to construct IPSLbis from IPSL does not lead to a strong modification of temporal properties. This result on temporal properties of data preprocessed with this matrix-recorrelation technique is consistent with the conclusions obtained in François et al. (2020) for a MBC method (MRec) using the same matrix-recorrelation. For LR SAFRAN outputs (Fig. 11d), values of AR1 differences are very close to 0, highlighting that the upscaling step used to construct LR SAFRAN data does not strongly modify the temporal properties of the initial SAFRAN reference dataset, which was expected by construction. Difference maps for temperature from QQ outputs (Fig. 11e–g) are relatively similar to those from the three datasets to adjust, respectively. However, for the three MBC methods used to adjust spatial dependence structure, modifications of temporal properties for temperature are not equivalent. With regard to MBC-CycleGAN and dOTC outputs (Fig. 11h, i, j, n, o and p), temporal statistics are close to that from the QQ outputs. It hence suggests that both MBC-CycleGAN and dOTC algorithms, although correcting the spatial features, perform little changes of the

Fig. 11 Differences of order 1 Pearson autocorrelation for temperature with SAFRAN reference for BC methods using as inputs **(b, e, h, k, n)** IPSL, **(c, f, i, l, o)** IPSLbis and **(d, g, j, m, p)** LR SAFRAN data. Results are shown during winter over the projection period for IPSL, IPSLbis, LR SAFRAN, QQ, MBC-CycleGAN, Spatial-R²D² and Spatial-dOTC datasets. The map of order 1 Pearson autocorrelation for SAFRAN references is also shown for temperature **(a)**

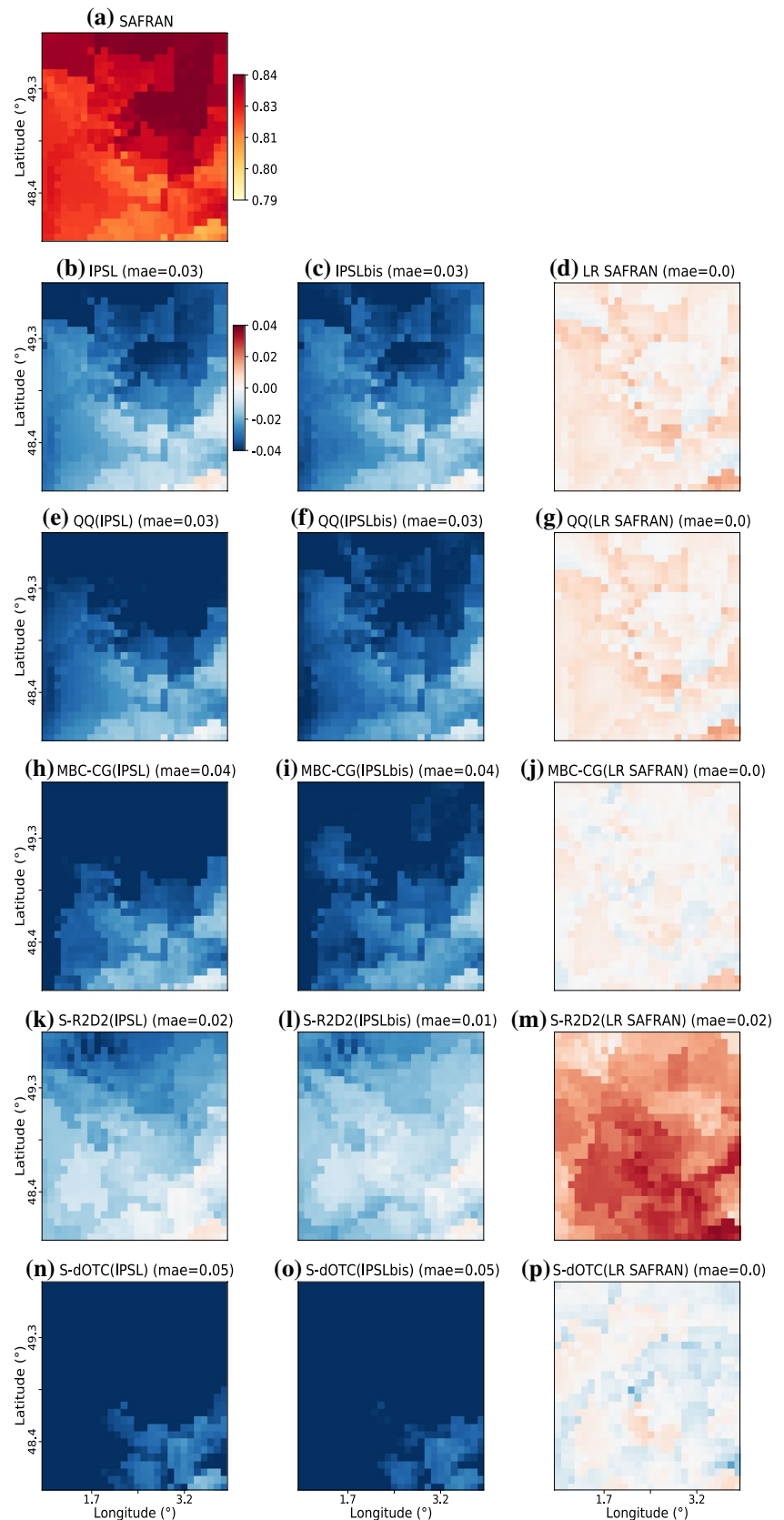


Table 1 Summary of attributes of the different climate data to correct

Climate data	Marginal prop.	Spatial prop.	Changes of marginal prop.	Changes of spatial prop.
IPSL model	From raw IPSL	From raw IPSL	Potentially not in line with SAFRAN	Potentially not in line with SAFRAN
LR SAFRAN	~ Same as SAFRAN	~ Same as SAFRAN	~ In line with SAFRAN	~ In line with SAFRAN
IPSLbis	~ Same as IPSL model	~ Same as IPSL model	~ In line with SAFRAN	~ In line with SAFRAN

temporal sequencing of the time series to correct. For MBC-CycleGAN, this is partly explained by the fact that, within the CycleGAN procedure, input maps from QQ outputs are transformed to outputs with improved spatial features, whilst not modifying too much the initial input image. It hence results in partially preserving the temporal properties of the QQ outputs used as inputs of the CycleGAN while providing improvements of the spatial representation. This particular point is thereafter discussed in greater details. Concerning R^2D^2 outputs, different results are obtained depending on the dataset to correct. For the correction of both IPSL and IPSLbis (Fig. 11k, l), R^2D^2 provides small improvements of temporal properties of temperature, which illustrates that, by including lags in the conditional dimensions, R^2D^2 is able to improve—in addition to spatial properties—temporal structure of climate datasets. However, for the correction of LR SAFRAN (Fig. 11m), a deterioration of AR1 temperature differences is obtained with respect to initial LR SAFRAN data (Fig. 11d). This result can be linked with the previously mentioned contrasted performances of the R^2D^2 method to adjust LR SAFRAN dataset in Subsect. 5.2.2.

For precipitation (Fig. S10), same conclusions hold for IPSL, IPSLbis and LR SAFRAN outputs. However, contrary to temperature, 1d-QQ corrections of IPSL and IPSLbis (Figs. S10e, f) show a pronounced improvement of temporal properties for precipitation, highlighting the potential influence of marginal properties of precipitation time series on its autocorrelation values. Moreover, the improvements of temporal properties of temperature provided by R^2D^2 for the corrections of IPSL and IPSLbis are no longer observed for precipitation (Fig. S10k, l). Instead, temporal properties with unexpected behaviors are obtained, potentially due to the difficulty of R^2D^2 to correct physical variables with events occurring at local scale, such as precipitation (Vrac and Thao 2020). It can also be due to the choice of the conditioning information made in R^2D^2 . As a reminder, it is indeed the rank structure of simulated precipitation (resp. temperature) that serves as a conditioning to generate Spatial- R^2D^2 outputs for precipitation (resp. temperature). As temporal properties (including rank sequences) of precipitation time series are not well simulated by IPSL model (Fig. S10b) compared to temperature (Fig. 11b), it potentially affects the quality of the corrections—and its temporal properties—provided by Spatial- R^2D^2 for precipitation. This highlights the

importance of choosing a relevant conditioning dimension for the implementation of R^2D^2 (Vrac and Thao 2020).

To illustrate the fact that MBC-CycleGAN performs little changes of the temporal sequencing of the inputs to adjust, we compare corrected daily maps from LR SAFRAN with those from the references. As the LR SAFRAN dataset is temporally matching the SAFRAN dataset by construction, classic forecast statistics such as Root Mean Square Error (RMSE) can indeed be interesting to assess the performances of MBC methods. Table 2 shows, for temperature and precipitation, the RMSE values with respect to SAFRAN references for the different BC outputs of LR SAFRAN. For temperature, the RMSE value between daily maps of the reference and the LR SAFRAN dataset is around 0.36. Slight improvement in terms of RMSE is provided by the QQ method (RMSE = 0.31). As the QQ method preserves the temporal sequencing of the times series to correct, this improvement is only due to the correction of marginal properties. The MBC-CycleGAN method presents better results (RMSE = 0.23), permitting to state with more confidence that, while adjustment of spatial dependence structure are performed, it modifies only slightly the temporal sequencing of the times series to correct. For R^2D^2 outputs, the RMSE value is quite large (RMSE=1.51), suggesting a strong modification of temporal properties. It can be linked with the underperformance of R^2D^2 already observed in Fig. 11m for the correction of LR SAFRAN. Concerning dOTC outputs, the RMSE value (= 0.42) is slightly higher than those observed for LR SAFRAN and QQ outputs. It suggests that the influence of the correction of univariate distributions and spatial dependence on temporal properties

Table 2 RMSE values between the reference SAFRAN and the different climate datasets in rows for temperature and precipitation during winter over the projection period

Physical variable	LR SAFRAN	QQ	MBC-CG	S- R^2D^2	S-dOTC
TAS	0.36	0.31	<u>0.23</u>	1.51	0.42
PR	0.75	0.73	<u>0.51</u>	3.41	1.03

As LR SAFRAN dataset is temporally matching the SAFRAN references, results are presented for LR SAFRAN and its MBC corrections only. For each physical variable, the best performing method is underlined

provided by dOTC is strong enough to affect its ability to provide appropriate forecasts at a daily scale. For precipitation, the same conclusions hold for the different BC outputs. To better illustrate the results from Table 2, two animations presenting the successive daily temperature and precipitation maps generated by MBC-CycleGAN for the correction of LR SAFRAN, as well as the corresponding daily maps from the references and the different BC methods, are provided as supplementary materials.

6 Conclusion, discussion and future work

6.1 Conclusions

Climate simulations biases are typically corrected with univariate BC methods, adjusting one physical variable and one location at a time, and thus spatial dependencies remain uncorrected. In this study, MBC-CycleGAN, an adaptation of the CycleGAN approach (Zhu et al. 2017) used to train image-to-image translation models, was presented, allowing for the adjustment of not only univariate distributions but also spatial dependence structures of climate simulations. The new suggested MBC method takes advantage of convolutional neural networks with simple architecture that are trained in competition to adjust spatial properties of simulated variables. The MBC-CycleGAN method was tested by adjusting temperature and precipitation time series from IPSL simulations with respect to the SAFRAN dataset over the region of Paris using two different cross-validation methods. The first cross-validation, that defines randomly calibration and projection periods, allows to test the new methodology in a stationary context. We took advantage of this first cross-validation method to compare two post-processing schemes (PP and MOS) approaches that differ in the statistical relationships the MBC-CycleGAN model learns to adjust spatial dependences. The MOS approach that considers biases to refer to systematic distributional differences between references and simulated climate variables was found to be more appropriate for the implementation of the MBC-CycleGAN method and was chosen to be applied for the rest of the study. The second cross-validation method, that defines chronologically calibration and projection periods, was then used to evaluate the ability of the MBC-CycleGAN method to adjust climate datasets in a non-stationary context. As IPSL simulations and SAFRAN references present different marginal and spatial changes between calibration and projection periods, two additional climate datasets (LR SAFRAN and IPSLbis) with changes that are in line with the references were specifically constructed and adjusted, allowing to better assess the quality of the corrections provided by the new method depending on the statistical biases of the data to be corrected. A wide range of

metrics has been used to evaluate bias adjustment outputs with references and initial climate data and assess the corrections of univariate distributions, spatial correlations and temporal properties. In addition to the 1d-QQ method, two state-of-the-art MBC (R^2D^2 and dOTC) methods have been implemented and used as benchmarks to better evaluate the influence of nonstationary properties on the results of the MBC-CycleGAN method. The results indicate that all the (M)BC methods implemented in this study generally present similar corrections of univariate distributions. Regarding spatial properties, the benefits of using MBC methods are clear compared to the 1d-QQ method. The MBC-CycleGAN method produced reasonable adjustments of spatial correlations with respect to R^2D^2 and dOTC methods for both temperature and precipitation and the three different climate datasets to adjust. Concerning the temporal aspect, the MBC-CycleGAN method is not designed to correct this specific statistical property and tends to conserve the temporal sequencing of the time series to correct. Combined with the corrections of spatial features, this property has proved to be particularly interesting for the applications of MBC-CycleGAN when the data to correct temporally match the references (e.g., as for LR SAFRAN and SAFRAN dataset, see Sect. 5.2.2). The proposed method indeed outperformed all the others (M)BC alternatives for the correction of LR SAFRAN by generally presenting both spatial and temporal statistics closer to those from the references. Concerning nonstationary properties, it has been found that changes of both marginal and spatial properties between the calibration and projection periods of the climate data to adjust can have a non-negligible effect on the quality of corrections from the MBC-CycleGAN algorithm, and more generally from all (M)BC outputs. In a general way, better results are obtained for the corrections of simulations with changes that are in agreement with those from the references, whether the MBCs make the assumption of nonstationarity of marginal properties and dependence structures or not.

6.2 Discussion and perspectives

In this study, the development of the MBC-CycleGAN method was mainly intended as a proof of concept, in order to test if GANs can be used for multivariate bias correction of climate simulations. Although bringing results with comparable performances of correction to that of well-established MBC methods, several avenues can be considered for the improvement of the proposed algorithm.

First, in order to remain in a context of proof of concept, a simple architecture of neural networks with a small number of convolutional layers has been considered for the discriminators and generators constituting the MBC-CycleGAN method. In the same idea, a classic formulation of the CycleGAN procedure—as initially described in Zhu

et al. (2017)—has been used with a binary-cross entropy loss function for the adversarial training (Eq. 1). Improving the training performances of GANs through more advanced architectures and optimization techniques is an active area of research (e.g., Salimans et al. 2016; Arjovsky et al. 2017; Karras et al. 2018, among others). A first natural step to potentially improve results would be to opt for a more sophisticated CycleGAN model. For example, it can be done by adding more layers in the neural network architectures of both generators and discriminators to potentially capture more complex spatial relationships for the correction of climate simulations. Also, modifying the initial adversarial loss functions (L_{GAN} in Eq. 1), as proposed in Arjovsky et al. (2017), would be interesting as it could permit to improve the stability of the learning and can prevent from mode collapse issues. However, although progress is constantly increasing concerning GANs, it is well-known that this particular class of neural networks can be more difficult to train than classical neural networks (e.g., Wu et al. 2020). The possibilities of modifications of the parameters defining a CycleGAN model are numerous, and a priori do not guarantee to improve the overall performance of the CycleGAN for the specific application of bias correction. Testing the different possibilities goes way beyond the scope of the present study and is left for future work.

Second, it has to be noted that our method, by combining the 1d-QQ method and the CycleGAN approach to adjust both marginal and spatial properties, is not designed to specifically account for any simulated changes for future periods. For marginal properties, other 1d-BC methods that are able to account for potential changes of univariate CDFs from the calibration to the projection period (e.g., CDF-t or QDM, Vrac et al. 2012; Cannon et al. 2015) can of course be employed instead of QQ, as long as they do not modify (too much) rank sequence of temperature and precipitation time series and thus do not distort the convergence of the CycleGAN procedure. Concerning changes of spatial properties, the CycleGAN approach as implemented in this study is based on the key assumption that the conditional distributions $X|Y$ and $Y|X$ are the same in the training (i.e., calibration) and test (i.e., projection) datasets. It results in our context in making a strong assumption on copula stationarity between present and future periods. Although spatial dependence structures can be considered to be stable in time as imposed by physical laws over a specific region of interest (e.g., Vrac 2018), it can not be generalized to each of the physical variables and regions. For example, more concentrated spatial rainfall events are expected with higher temperatures in the future (Guinard et al. 2015; Wasko et al. 2016). Therefore, should the changes in spatial properties in the simulations between calibration and projection periods be reproduced in the correction? By comparing our results obtained with different levels of nonstationarity in the model

evolution and with two well-established MBCs based on copula stationarity (R^2D^2) and nonstationarity (dOTC) for future periods, we shed light on how the nonstationary properties of the simulations are taken into account by the different multivariate BC methods. The benefits of considering MBC methods assuming copula nonstationarity for the correction of such climate dataset are not always as clear-cut as expected compared to MBC methods assuming copula stationarity. This raises the question of whether developing MBC methods assuming copula nonstationarity is justified, i.e., whether it is worth striving for developing complicated statistical methods that consider the simulated evolution of copula in the correction procedure, and, in the end, do not produce drastically better results than MBCs assuming copula stationarity. In practice, accounting for nonstationarity of simulations in bias correction procedures still remains an open question which needs to be answered on a case-by-case basis. Developing new MBC methods that are specifically able to reproduce these simulated changes in the correction is of course an important perspective but the application of such methods would be inappropriate as long as the changes from climate simulations for future periods have not been first identified as relevant.

Third, the MBC-CycleGAN method has been developed to correct spatial correlations of climate simulations for each physical variable separately, and thus does neither consider the adjustment of inter-variable correlations nor temporal structure. A possible extension of the initial method can be the consideration of inter-variable and/or temporal correlations by providing to the CycleGAN model images with not only one but several channels of the different physical variables to correct. For example, for the adjustment of inter-variable correlations between temperature and precipitation, concatenated images of daily temperature and precipitation maps in an array of dimension $2 \times 28 \times 28$ can be provided as inputs to the adversarial neural network. Similarly, adjusting temporal correlations could be considered by adding channels with lagged versions of the physical variable. Using images with additional channels would imply to change, at least, the neural network architecture by replacing 2d-convolutional neural networks with 3d-ones to allow the CycleGAN model to consider inter-channels correlations. However, as adding additional channels can potentially make the training of the CycleGAN more complicated, it is likely that others changes relative to the architecture of neural networks and optimization techniques would be required, as those mentioned previously.

Fourth, according to the results for the correction of the references at large-scale (LR SAFRAN), MBC-CycleGAN showed greater improvements of both spatial and temporal statistics compared to the other MBC methods. These promising results suggest that MBC-CycleGAN can be used directly in downscaling applications, a practice that is not

initially recommended with univariate quantile mapping techniques (Maraun 2013; Gutmann et al. 2014). Although producing reasonable results of adjustments for temperature and precipitation spatial distributions of IPSL and IPSLbis datasets, the outperformance of MBC-CycleGAN observed for the correction of LR SAFRAN is not obtained for these climate outputs. A possible reason explaining why the performances of MBC-CycleGAN differ between these three exercises of correction concerns the importance of the distributional differences between the inputs and target dataset considered. Indeed, unsupervised image-to-image translation algorithms such as CycleGAN can present difficulties to map two random variables \mathbf{X} and \mathbf{Y} with probability distributions that exhibit strong differences (Gokaslan et al. 2019; Royer et al. 2020). As LR SAFRAN presents smaller bias with the references than IPSL and IPSLbis data, outstanding results are obtained for the correction of LR SAFRAN with MBC-CycleGAN, while more moderate quality results are produced for IPSL and IPSLbis. Improving the MBC-CycleGAN algorithm such that it is able to produce satisfactory results even when distributions with very strong (marginal and spatial) differences are considered is of great interest to allow its use for operational purposes.

Fifth, in this study, particular precautions have been taken to prevent overfitting during training of CycleGAN networks, such as including a regularization technique called “dropout” in both generators and discriminators architectures (see Appendix B for further details), or verifying that the performances of MBC-CycleGAN on projection periods are not deteriorated along training (not shown). These precautions permit to apply with confidence MBC-CycleGAN algorithms on projection periods. The issue of overfitting raises the question of the generalization capability of statistical models, and how they cope with new (and unseen) data. In most of the study, calibration and projection periods have been defined chronologically for the 1979–2016 period, and one can argue that small differences in terms of spatial properties are obtained between the two periods. Assessing the performances of the MBC-CycleGAN algorithm for the adjustment of climate projections with very different spatial structures remains an interesting perspective. For example, this could be done by adapting the methodology used for the generation of IPSLbis to generate alternative climate simulations for the projection period with strong spatial changes, and apply the pretrained CycleGAN neural network used for the correction of IPSL in this study.

Finally, as implemented in this study, the proposed MBC-CycleGAN algorithm produces a single correction (output) for a given input. Although essential in climate applications, uncertainty quantification of MBC-CycleGAN outputs is not estimated here. An interesting possibility of extension to model uncertainty of corrected outputs would be to introduce some stochasticity into the correction procedure by

giving to the generators not only daily maps to adjust but also vectors of random noises. Then, for a given daily map, it would produce an ensemble of plausible corrections. The spread between the ensemble members would represent the uncertainty associated with the multivariate bias correction.

We hope that this study serves as a starting point for the use of GANs for multivariate bias correction of climate simulations. One of the main advantages of using MBC-CycleGAN is that adjustment is performed images by images, i.e. maps by maps. If well trained, discriminators somehow guarantee that individual generated maps produced by generators are realistic with respect to references, while daily maps with strong statistical artefacts are rejected. This is not the case for the other MBC methods such as R^2D^2 or dOTC, that provide corrected simulations with appropriate distributional statistics without being particularly constrained to generate realistic daily maps. Providing corrections with realistic maps at a daily scale can be useful for the scientific community working on climate change impacts, e.g., in hydrology, for which daily spatial features are of major concern.

Appendix A: Details on the MBC-CycleGAN method

Let consider the correction of a random variable, denoted \mathbf{X} (e.g., biased climate simulations outputs) with respect to a reference random variable, denoted \mathbf{Y} . In our study, \mathbf{X} and \mathbf{Y} live in dimension $28 \times 28 = 784$ dimensions. We denote \mathbf{X}^0 and \mathbf{X}^1 the random variables to correct from climate simulations during the calibration and projection period, respectively. Similarly, \mathbf{Y}^0 is considered as the random variable of references for the calibration period. The goal of any BC methods is to infer future unobserved data \mathbf{Y}^1 from the reference variable \mathbf{Y}^0 during calibration, and the variables from model simulations for calibration (\mathbf{X}^0) and projection (\mathbf{X}^1) periods.

In practice, BC methods are applied to correct samples $(\mathbf{x}_1^0, \dots, \mathbf{x}_n^0)$ and $(\mathbf{x}_1^1, \dots, \mathbf{x}_n^1)$ from the random variables \mathbf{X}^0 and \mathbf{X}^1 , with respect to a sample $(\mathbf{y}_1^0, \dots, \mathbf{y}_n^0)$ from the random variable \mathbf{Y}^0 . For example, 1d-bias corrections of $(\mathbf{x}_1^0, \dots, \mathbf{x}_n^0)$ and $(\mathbf{x}_1^1, \dots, \mathbf{x}_n^1)$ with the QQ method can be denoted $(\mathbf{q}\mathbf{q}_1^0, \dots, \mathbf{q}\mathbf{q}_n^0)$ and $(\mathbf{q}\mathbf{q}_1^1, \dots, \mathbf{q}\mathbf{q}_n^1)$. As explained in Sect. 3, the CycleGAN approach within the MBC-CycleGAN methodology is applied between 1d-QQ outputs and references. Hence, two generators $G_{\mathbf{Q}\mathbf{Q} \rightarrow \mathbf{Y}}$ and $G_{\mathbf{Y} \rightarrow \mathbf{Q}\mathbf{Q}}$ are considered, as well as two discriminators $D_{\mathbf{Q}\mathbf{Q}}$ and $D_{\mathbf{Y}}$. The different steps constituting the MBC-CycleGAN method are described in an algorithmic way as follows:

Algorithm 1 MBC-CycleGAN training algorithm. In all experiments presented in the paper, a batch size $m = 32$ is used.

Require: α_{disc} - the learning rate for the discriminators, α_{gen} - the learning rate for the generators, m - the batch size.

Require: $\delta_{\mathbf{Q}\mathbf{Q}}$, $\delta_{\mathbf{Y}}$ - the initial discriminators' parameters. $\theta_{G_{\mathbf{Q}\mathbf{Q}\rightarrow\mathbf{Y}}}$, $\theta_{G_{\mathbf{Y}\rightarrow\mathbf{Q}\mathbf{Q}}}$ - the initial generators' parameters.

$(\mathbf{x}_1^0, \dots, \mathbf{x}_n^0)$ a sample of the random variable \mathbf{X}^0 .

$(\mathbf{x}_1^1, \dots, \mathbf{x}_p^1)$ a sample of the random variable \mathbf{X}^1 .

$(\mathbf{y}_1^0, \dots, \mathbf{y}_n^0)$ a sample of the random variable \mathbf{Y}^0 .

Ensure: $(\mathbf{z}_1^0, \dots, \mathbf{z}_n^0)$ and $(\mathbf{z}_1^1, \dots, \mathbf{z}_p^1)$ the corrections of $(\mathbf{x}_1^0, \dots, \mathbf{x}_n^0)$ and $(\mathbf{x}_1^1, \dots, \mathbf{x}_p^1)$.

- 1: Compute $(\mathbf{q}\mathbf{q}_1^0, \dots, \mathbf{q}\mathbf{q}_n^0)$ and $(\mathbf{q}\mathbf{q}_1^1, \dots, \mathbf{q}\mathbf{q}_p^1)$ the 1d-bias corrections of $(\mathbf{x}_1^0, \dots, \mathbf{x}_n^0)$ and $(\mathbf{x}_1^1, \dots, \mathbf{x}_p^1)$ using the quantile-mapping method (Déqué 2007).
- 2: Compute $(\widetilde{\mathbf{q}\mathbf{q}}_1^0, \dots, \widetilde{\mathbf{q}\mathbf{q}}_n^0)$ and $(\widetilde{\mathbf{q}\mathbf{q}}_1^1, \dots, \widetilde{\mathbf{q}\mathbf{q}}_p^1)$, the point-wise min-max normalizations of $(\mathbf{q}\mathbf{q}_1^0, \dots, \mathbf{q}\mathbf{q}_n^0)$ and $(\mathbf{q}\mathbf{q}_1^1, \dots, \mathbf{q}\mathbf{q}_p^1)$ with the range of $(\mathbf{y}_1^0, \dots, \mathbf{y}_n^0)$.
- 3: Compute $(\widetilde{\mathbf{y}}_1^0, \dots, \widetilde{\mathbf{y}}_n^0)$, the point-wise min-max normalization of $(\mathbf{y}_1^0, \dots, \mathbf{y}_n^0)$.
- 4: **while** $\widehat{\mathcal{E}}$, the estimated energy distance on ranks between $(\mathbf{y}_1^0, \dots, \mathbf{y}_n^0)$ and $(\mathbf{z}_1^0, \dots, \mathbf{z}_n^0)$, has not converged **do**
- 5: Sample $\{\widetilde{\mathbf{q}\mathbf{q}}_i^0\}_{i=1}^{m/2}$ a batch from the dataset $(\widetilde{\mathbf{q}\mathbf{q}}_1^0, \dots, \widetilde{\mathbf{q}\mathbf{q}}_n^0)$.
- 6: Sample $\{\widetilde{\mathbf{y}}_i^0\}_{i=1}^{m/2}$ a batch from the dataset $(\widetilde{\mathbf{y}}_1^0, \dots, \widetilde{\mathbf{y}}_n^0)$.
- 7: Generate "fake" samples $\{\widetilde{\mathbf{y}}_i^{0, fake}\}_{i=1}^{m/2}$:
 $\forall i \in [1, \dots, m/2], \widetilde{\mathbf{y}}_i^{0, fake} = G_{\mathbf{Q}\mathbf{Q}\rightarrow\mathbf{Y}}(\widetilde{\mathbf{q}\mathbf{q}}_i^0)$.
- 8: Generate "fake" samples $\{\widetilde{\mathbf{q}\mathbf{q}}_i^{0, fake}\}_{i=1}^{m/2}$:
 $\forall i \in [1, \dots, m/2], \widetilde{\mathbf{q}\mathbf{q}}_i^{0, fake} = G_{\mathbf{Y}\rightarrow\mathbf{Q}\mathbf{Q}}(\widetilde{\mathbf{y}}_i^0)$.
- 9: Update $\delta_{\mathbf{Q}\mathbf{Q}}$, using Adam optimizer and the learning rate α_{disc} , by computing the adversarial loss function (Eq. 1) and its gradients with the samples $\{\widetilde{\mathbf{q}\mathbf{q}}_i^0\}_{i=1}^{m/2}$ and $\{\widetilde{\mathbf{q}\mathbf{q}}_i^{0, fake}\}_{i=1}^{m/2}$. The adversarial loss function must be maximized.
- 10: Update $\delta_{\mathbf{Y}}$, using Adam optimizer and the learning rate α_{disc} , by computing the adversarial loss function (Eq. 1) and its gradients with the samples $\{\widetilde{\mathbf{y}}_i^0\}_{i=1}^{m/2}$ and $\{\widetilde{\mathbf{y}}_i^{0, fake}\}_{i=1}^{m/2}$. The adversarial loss function must be maximized.
- 11: Compute the full loss function (Eq. 6) and its gradients with respect to the parameters $\theta_{G_{\mathbf{Q}\mathbf{Q}\rightarrow\mathbf{Y}}}$ and $\theta_{G_{\mathbf{Y}\rightarrow\mathbf{Q}\mathbf{Q}}}$ of the generators $G_{\mathbf{Q}\mathbf{Q}\rightarrow\mathbf{Y}}$ and $G_{\mathbf{Y}\rightarrow\mathbf{Q}\mathbf{Q}}$.
- 12: Update the parameters $\theta_{G_{\mathbf{Q}\mathbf{Q}\rightarrow\mathbf{Y}}}$ and $\theta_{G_{\mathbf{Y}\rightarrow\mathbf{Q}\mathbf{Q}}}$ by minimizing the full loss function, using Adam optimizer, according to its gradients and the learning rate α_{gen} .
- 13: Compute $(\widetilde{\mathbf{s}}_1^0, \dots, \widetilde{\mathbf{s}}_n^0)$ the normalized data with a corrected spatial dependence structure for the calibration period:
 $\forall i \in [1, \dots, n], \widetilde{\mathbf{s}}_i^0 = G_{\mathbf{Q}\mathbf{Q}\rightarrow\mathbf{Y}}(\widetilde{\mathbf{q}\mathbf{q}}_i^0)$.
- 14: Reorder each 1d-bias corrected dimension from the dataset $(\mathbf{q}\mathbf{q}_1^0, \dots, \mathbf{q}\mathbf{q}_n^0)$ according to its rank structure in the dataset $(\widetilde{\mathbf{s}}_1^0, \dots, \widetilde{\mathbf{s}}_n^0)$ with the Schaake Shuffle method (Clark et al 2004) to obtain $(\mathbf{z}_1^0, \dots, \mathbf{z}_n^0)$, the bias correction of $(\mathbf{x}_1^0, \dots, \mathbf{x}_n^0)$.
- 15: Compute the estimated energy distance on ranks $\widehat{\mathcal{E}}$ evaluated between $(\mathbf{y}_1^0, \dots, \mathbf{y}_n^0)$ and $(\mathbf{z}_1^0, \dots, \mathbf{z}_n^0)$.
- 16: **end while**
- 17: Compute $(\widetilde{\mathbf{s}}_1^1, \dots, \widetilde{\mathbf{s}}_p^1)$ the normalized data with a corrected spatial dependence structure for the projection period:
 $\forall i \in [1, \dots, p], \widetilde{\mathbf{s}}_i^1 = G_{\mathbf{Q}\mathbf{Q}\rightarrow\mathbf{Y}}(\widetilde{\mathbf{q}\mathbf{q}}_i^1)$.
- 18: Reorder each 1d-bias corrected dimension from the dataset $(\mathbf{q}\mathbf{q}_1^1, \dots, \mathbf{q}\mathbf{q}_p^1)$ according to its rank structure in the dataset $(\widetilde{\mathbf{s}}_1^1, \dots, \widetilde{\mathbf{s}}_p^1)$ with the Schaake Shuffle method to obtain $(\mathbf{z}_1^1, \dots, \mathbf{z}_p^1)$, the bias correction of $(\mathbf{x}_1^1, \dots, \mathbf{x}_p^1)$.

Appendix B: Details on the simple architecture of neural networks used in MBC-CycleGAN

The simple neural network architectures used for the discriminators and generators constituting the MBC-CycleGAN method in this study are described with more details in this appendix.

Appendix B.1: Architecture of the generators

As explained in Sect. 3.3.2, skip connections are used in the architecture of the generators to ease the training process. Skip connections permit to provide information to a given layer that comes not only from the direct previous layer, but also from other upstream convolution layers in the architecture. Skipping over layers permits to avoid vanishing gradients issues, which is a problem that can make the network hard to train. All layers except the first one have leaky rectified linear unit (leaky-ReLu) activation functions defined as: $y = \begin{cases} x & \text{if } x \geq 0, \\ \alpha x & \text{otherwise,} \end{cases}$ with $\alpha = 0.2$. Dropout regularization, that refers to ignoring neurons chosen at random during training, is used after the second and third 2D convolutional layers to prevent from overfitting (e.g., Srivastava et al. 2014). The probability used for dropout is 0.4. A summary of the simple neural network architecture used for the generators is described below in Table 3.

Appendix B.2: Architecture of the discriminators

A summary of the simple neural network architecture used for the discriminators is described below in Table 4.

Appendix C: Methodology for the generation of IPSLbis

For the generation of IPSLbis data, a two-step procedure is developed to construct, from IPSL data, climate data that present marginal and spatial changes that are in line with those from references between the calibration and projection periods. In order to stay with comparable changes as those from LR SAFRAN, LR SAFRAN changes are reproduced. We recall that, for the calibration period, IPSL and IPSLbis data are strictly identical. The two-step procedure is only used to produce alternative climate data for the projection period.

Appendix C.1: Marginal changes with CDF-t

The first step of the procedure consists in producing time series for the projection period of IPSLbis by taking into

account marginal changes of LR SAFRAN with the 1d-BC named CDFt (Vrac et al. 2012). Initially, CDF-t is a version of univariate quantile mapping method designed to correct at each individual grid cell marginal properties of climate simulations outputs during the calibration and the projection period according to the data from the reference observed during calibration. CDF-t, by defining a specific transfer function, has been conceived to take into account the potential simulated changes of univariate distributions from the calibration to the projection period in order to produce the adjusted data such that the marginal changes are in line with those from the simulations. While, traditionally, this quantile-mapping approach is used to find, in a bias correction context, a mathematical transformation allowing to go from simulations to references, we here applied CDF-t to go from “large scale” references (LR SAFRAN) to simulations for future periods. By proceeding this way, the produced time series are projected distributions in the domain of IPSL simulations that have been obtained while taking into account the potential evolution of CDFs of the LR SAFRAN dataset between the calibration and projection periods. By concatenating times series from IPSL for the calibration period and those obtained from the CDF-t method for the projection period, new climate times series are obtained, presenting marginal distributions changes in line with those from references.

Appendix C.2: Spatial changes with a matrix-recorrelation technique

The second step consists in deriving a spatial dependence structure for the projection period such that spatial changes of LR SAFRAN are reproduced. To do so, we take advantage of a matrix-recorrelation technique used for the MBC method presented in Bárdossy and Pegram (2012) to impose to climate data a specific spatial dependence structure for the projection period. Our methodology is summarized in Table 5. It consists in first projecting individually each variable of both IPSL simulations and LR SAFRAN during calibration and projection periods to the univariate normal distribution with a Gaussian quantile mapping method. This “Gaussianization” step is particularly suited for variables with mixed distributions such as precipitation (composed of wet and dry events). Computing Pearson correlation matrices on such Gaussianized data instead of raw data permits to better describe its dependence structure. Thus, Pearson correlation matrices of the different Gaussianized data are computed. They are respectively denoted as $C_{I,C}$, $C_{I,P}$, $C_{I,C}^{(bis)}$, $C_{I,P}^{(bis)}$, $C_{S,C}$, $C_{S,P}$ for IPSL during calibration, IPSL during projection, IPSLbis during calibration, IPSLbis during projection, LR SAFRAN during calibration and LR SAFRAN during projection. Additionally, let $r_{I,C}$, $r_{I,P}$, $r_{I,C}^{(bis)}$, $r_{I,P}^{(bis)}$, $r_{S,C}$, $r_{S,P}$ denote one of their entry. Note that by construction, $C_{I,C}$

Table 3 The architecture of the generators used in the MBC-CycleGAN network

Layer	Layer name	Filter	Stride size	Output size
1	Input layer	n.a.	n.a.	$28 \times 28 \times 1$
2	Conv2D	$3 \times 3 \times 64$	(1, 1)	$28 \times 28 \times 64$
3	Conv2D	$3 \times 3 \times 128$	(2, 2)	$14 \times 14 \times 128$
	Leaky ReLU + Dropout	n.a.	n.a.	$14 \times 14 \times 128$
4	Conv2D	$3 \times 3 \times 256$	(2, 2)	$7 \times 7 \times 256$
	Leaky ReLU + Dropout	n.a.	n.a.	$7 \times 7 \times 256$
5	Conv2DTranspose	$4 \times 4 \times 128$	(2, 2)	$14 \times 14 \times 128$
	Skip connection (Layer 3) + Leaky ReLU	n.a.	n.a.	$14 \times 14 \times 128$
6	Conv2DTranspose	$4 \times 4 \times 64$	(2, 2)	$28 \times 28 \times 64$
	Skip connection (Layer 2) + Leaky ReLU	n.a.	n.a.	$28 \times 28 \times 64$
7	Conv2D	$1 \times 1 \times 1$	(1, 1)	$28 \times 28 \times 1$
	Skip connection (Layer 1) + Leaky ReLU	n.a.	n.a.	$28 \times 28 \times 1$

is the same as $C_{I,C}^{(bis)}$ and that $C_{I,P}^{(bis)}$ is unknown. Assessing the changes of LR SAFRAN spatial correlations between calibration and projection periods is now required to derive the spatial dependence structure of IPSLbis for the projection period. A simple approach to determine $r_{I,P}^{(bis)}$, the correlation of the Gaussianized data of IPSLbis for projection, would be to compute it based on the difference of correlations from Gaussianized LR SAFRAN data such as $r_{I,P}^{(bis)} = r_{I,C} + r_{S,P} - r_{S,C}$. However, computing $r_{I,P}^{(bis)}$ this way can lead to obtain correlation values that are out of range, i.e. being greater than 1 or less than -1, which is not appropriate.

From Bárdossy and Pegram (2012), given $r_{I,C}, r_{S,C}, r_{S,P}$, one can derive $r_{I,P}$ using Fisher-Z transformation (Fisher 1915) as following:

$$r_{I,P}^{(bis)} = \frac{\frac{(1+r_{S,P})}{(1+r_{S,C})}(1+r_{I,C}) - \frac{(1-r_{S,P})}{(1-r_{S,C})}(1-r_{I,C})}{\frac{(1+r_{S,P})}{(1+r_{S,C})}(1+r_{I,C}) + \frac{(1-r_{S,P})}{(1-r_{S,C})}(1-r_{I,C})}. \tag{7}$$

Fisher-Z transformation permits to transform a bounded random variable to another random variable that can be assumed to be Normal, and for which additive correction can be performed (see Mehrotra and Sharma (2019) for the derivation of Eq. 7). By deriving this way all the new correlation coefficients, the potential changes in correlations in the Gaussianized LR SAFRAN data are preserved and the Pearson correlation matrix for Gaussianized IPSLbis during the projection period is obtained.

Table 4 The architecture of the discriminators used in the MBC-CycleGAN network.

Layer	Layer name	Filter	Stride size	Output size
1	Input layer	n.a.	n.a.	$28 \times 28 \times 1$
2	Conv2D	$3 \times 3 \times 64$	(2, 2)	$14 \times 14 \times 64$
	Leaky ReLU + Dropout	n.a.	n.a.	$14 \times 14 \times 64$
3	Conv2D	$3 \times 3 \times 128$	(2, 2)	$7 \times 7 \times 128$
	Leaky ReLU + Dropout	n.a.	n.a.	$7 \times 7 \times 128$
4	Flatten	n.a.	n.a.	6272
5	Dense + sigmoid	1	n.a.	1

Now that the Pearson correlation matrix, $C_{I,P}^{(bis)}$, is computed, a combination of “decorrelation” and “redecorrelation” steps using decompositions of correlation matrices through singular value decomposition (SVD, Beltrami 1873; Jordan 1874a, b; Stewart 1993) is applied on the Gaussianized data of IPSL during projection period, forcing its Pearson correlation matrix to be exactly the same as the Pearson correlation matrix, $C_{I,P}^{(bis)}$. The new dependence structure for IPSLbis is obtained. Finally, a reordering of time series from CDF-t outputs according to this new dependence structure is performed using the Schaake Shuffle method to obtain IPSLbis data for the projection period.

Appendix D: Spatial correlation changes analysis

We present a spatial changes analysis to provide a better picture of the properties of the climate data in terms of changes between the calibration and projection periods. As a reminder, IPSLbis data are generated using the two-step procedure described in Appendix 3 such that its marginal and dependence changes are in line with those from LR SAFRAN (and therefore SAFRAN) for the projection period. Fig. S3 displays scatterplots of differences between Spearman spatial correlations of temperature and precipitation evaluated for all pairwise combinations of sites, computed for the calibration (1979–2005) and the projection (2006–2016) period, respectively. Scatterplots compares differences of Spearman correlation with respect to those from LR SAFRAN. It permits one to visually verify if changes in the spatial dependence structure are in line to those from references at large-scale. Using rank correlation here permits to measure in isolation the spatial dependence between two sites rid of their marginal properties. Figures for the analysis of marginal changes— in particular, mean and standard deviation changes—are also displayed in Figs. S4 and S5 for information purposes only. Results on univariate

Table 5 Summary of the different steps used to construct the spatial dependence structure of IPSLbis

Methodology step	IPSL calib. proj.	IPSLbis proj.	LR SAFRAN calib. proj.
0. Input data	$I_C I_P$		$S_C S_P$
1. Gaussianization	$G_{I,C} G_{I,P}$		$G_{S,C} G_{S,P}$
2. Pearson corr.	$C_{I,C} C_{I,P}$	$C_{I,P}^{(bis)}$ (see Eq. 7)	$C_{S,C} C_{S,P}$
3. Square root matrix		U (for $C_{I,P}^{(bis)}$)	
4. Inverse square root		T (for $C_{I,P}$)	
5. Decorr.-recorr.		$G_{I,P}^{(bis)} =$ $G_{I,P} \times T \times U$	
6. Reorder marginals		$I_P^{(bis)}$	

properties can be briefly summarized as such: changes in marginal properties from SAFRAN references (resp. IPSL model) are in agreement (resp. disagreement) with those from LR SAFRAN for both temperature and precipitation. For IPSLbis, the application of the CDF-t method permits to obtain marginal changes for both temperature and precipitation similar to those from LR SAFRAN. Concerning spatial properties, as expected, changes in spatial correlations from SAFRAN references are (partially) in agreement with those from LR SAFRAN for both temperature (Fig. S3a) and precipitation (Fig. S3d). Concerning changes in the IPSL simulations, simulated changes of spatial correlations for temperature (Fig. S3b) are globally in line with those from LR SAFRAN, highlighting the ability of the climate model to provide appropriate temperature changes in spatial structure between the calibration and the projection periods. However, conclusions are quite different for precipitation, for which simulated changes are not in agreement at all with those from the reference at large scale (Fig. S3e). Hence, IPSL model presents discrepancy of changes for precipitation with respect to LR SAFRAN (and thus, SAFRAN references), that could potentially affect the quality of the correction depending on how MBC-CycleGAN accounts for these changes in its correction procedure. Concerning the results for IPSLbis, changes for both temperature (Fig. S3c) and precipitation (Fig. S3f) are similar to those from LR SAFRAN, confirming that the two-step methodology used to impose to IPSL specific changes of spatial correlations is appropriate here.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00382-021-05869-8>.

Acknowledgements This work was granted access to the HPC resources of IDRIS under the allocation 20XX-[AD011011646] made by GENCI. MV acknowledges support from the CoCliServ project,

which is part of ERA4CS, an ERA-NET initiated by JPI Climate and cofunded by the European Union.

Author contributions MV had the initial idea of the study and its structure, which was enriched by all coauthors. BF made all computations and figures, with help from ST. BF wrote the first draft of the article, with inputs, corrections and additional writing contributions from MV and ST.

Funding This research has been supported by the CoCliServ project, which is part of ERA4CS, an ERA-NET initiated by JPI Climate and cofunded by the European Union.

Availability of data and material The IPSL-CM5A-MR model data simulations as part of the CMIP5 climate model simulations can be downloaded through the Earth System Grid Federation portals. Instructions to access the data are available here: <https://pcmdi.llnl.gov/mips/cmip5/data-access-getting-started.html>, last access: 06 September 2020, (PCMDI, 1989). The SAFRAN reanalysis dataset is available upon request to the French National Centre for Meteorological Research (CNRM, Météo-France CNRS).

Declaration

Conflicts of interest/Competing interests The authors have no conflicts of interest to declare that are relevant to the content of this article.

Code availability The code for MBC-CycleGAN is publicly available at https://github.com/bastien-francois/MBC_CycleGAN. The R package for R^2D^2 is available at <https://github.com/thaos/R2D2> (Vrac and Thao 2020). dOTC is publicly available at <https://github.com/yrobink/SBCK> (Robin et al. 2019).

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

References

- Arjovsky M, Chintala S, Bottou L (2017) Wasserstein GAN. [arXiv:1701.07875](https://arxiv.org/abs/1701.07875)
- Baño-Medina J, Manzanar R, Gutiérrez JM (2020) Configuration and intercomparison of deep learning neural models for statistical downscaling. *Geosci Model Dev* 13(4):2109–2124. <https://doi.org/10.5194/gmd-13-2109-2020>
- Bárdossy A, Pegram G (2012) Multiscale spatial recorrelation of RCM precipitation to produce unbiased climate change scenarios over large areas and small. *Water Resour Res* 48:9502. <https://doi.org/10.1029/2011WR011524>
- Bartok B, Tobin I, Vautard R, Vrac M, Jin X, Levavasseur G, Denvil S, Dubus L, Parey S, Michelangeli PA, Troccoli A, Saint-Drenan YM (2019) A climate projection dataset tailored for the European

- energy sector. *Clim Serv* 16(100):138. <https://doi.org/10.1016/j.cliser.2019.100138>
- Bates B, Kundzewicz Z, Wu S, Burkett V, Doell P, Gwary D, Hanson C, Heij B, Jiménez B, Kaser G, Kitoh A, Kovats S, Kumar P, Magadza C, Martino D, Mata L, Medany M, Miller K, Arnell N (2008) Climate change and water. Technical Paper of the Intergovernmental Panel on Climate Change. Tech. rep, The Intergovernmental Panel on Climate Change
- Beltrami E (1873) Sulle funzioni bilineari. *Giornale Mat Uso degli Stud Delle Univ* 11:98–106
- Berg P, Feldmann H, Panitz HJ (2012) Bias correction of high resolution regional climate model data. *J Hydrol* 448–449:80–92. <https://doi.org/10.1016/j.jhydrol.2012.04.026>
- Bhatia S, Jain A, Hooi B (2020) ExGAN: adversarial generation of extreme samples. [arXiv:2009.08454](https://arxiv.org/abs/2009.08454)
- Bihlo A (2020) A generative adversarial network approach to (ensemble) weather prediction. [arXiv:2006.07718](https://arxiv.org/abs/2006.07718)
- Caminade C, Kovats S, Rocklov J, Tompkins AM, Morse AP, Colón-González FJ, Stenlund H, Martens P, Lloyd SJ (2014) Impact of climate change on global malaria distribution. *Proc Natl Acad Sci USA* 111(9):3286–3291. <https://doi.org/10.1073/pnas.1302089111>
- Cannon AJ (2018) Multivariate quantile mapping bias correction: an N-dimensional probability density function transform for climate model simulations of multiple variables. *Clim Dyn* 50(1):31–49. <https://doi.org/10.1007/s00382-017-3580-6>
- Cannon A, Sobie S, Murdock T (2015) Bias correction of gcm precipitation by quantile mapping: how well do methods preserve changes in quantiles and extremes? *J Clim* 28(17):6938–6959. <https://doi.org/10.1175/JCLI-D-14-00754.1>
- Cattiaux J, Douville H, Peings Y (2013) European temperatures in CMIP5: origins of present-day biases and future uncertainties. *Clim Dyn* 41:2889–2907. <https://doi.org/10.1007/s00382-013-1731-y>
- Chapman WE, Subramanian AC, Delle Monache L, Xie SP, Ralph FM (2019) Improving atmospheric river forecasts with machine learning. *Geophys Res Lett* 46(17–18):10627–10635. <https://doi.org/10.1029/2019GL083662>
- Christensen JH, Boberg F, Christensen OB, Lucas-Picher P (2008) On the need for bias correction of regional climate change projections of temperature and precipitation. *Geophys Res Lett* 35(20):L20709. <https://doi.org/10.1029/2008GL035694>
- Clark M, Gangopadhyay S, Hay L, Rajagopalan B, Wilby R (2004) The Schaake shuffle: a method for reconstructing space-time variability in forecasted precipitation and temperature fields. *J Hydrometeor* 5(1):243–262
- Defrance D, Ramstein G, Charbit S, Vrac M, Famién AM, Sultan B, Swingedouw D, Dumas C, Gemenne F, Alvarez-Solas J, Vanderlinden JP (2017) Consequences of rapid ice sheet melting on the Sahelian population vulnerability. *Proc Natl Acad Sci USA* 114(25):6533–6538. <https://doi.org/10.1073/pnas.1619358114>
- Dekens L, Parey S, Grandjacques M, Dacunha-Castelle D (2017) Multivariate distribution correction of climate model outputs: a generalization of quantile mapping approaches: multivariate distribution correction of climate model outputs. *Environmetrics* 28:e2454. <https://doi.org/10.1002/env.2454>
- Denton E, Chintala S, Szlam A, Fergus R (2015) Deep generative image models using a laplacian pyramid of adversarial networks. [arXiv:1506.05751](https://arxiv.org/abs/1506.05751)
- Déqué M (2007) Frequency of precipitation and temperature extremes over France in an anthropogenic scenario: model results and statistical correction according to observed values. *Glob Planet Change* 57(1):16–26. <https://doi.org/10.1016/j.gloplacha.2006.11.030>
- Dufresne JL, Foujols MA, Denvil S, Caubel A, Marti O, Aumont O, Balkanski Y, Bekki S, Bellenger H, Benshila R, Bony S, Bopp L, Braconnot P, Brockmann P, Cadule P, Cheruy F, Codron F, Cozic A, Cugnet D, de Noblet N, Duvel JP, Ethé C, Fairhead L, Fichet T, Flavoni S, Friedlingstein P, Grandpeix JY, Guez L, Guilyardi E, Hauglustaine D, Hourdin F, Idelkadi A, Ghattas J, Joussaume S, Kageyama M, Krinner G, Labetoulle S, Lahellec A, Lefebvre MP, Lefevre F, Levy C, Li ZX, Lloyd J, Lott F, Madec G, Mancip M, Marchand M, Masson S, Meurdesoif Y, Mignot J, Musat I, Parouty S, Polcher J, Rio C, Schulz M, Swingedouw D, Szopa S, Talandier C, Terray P, Viovy N, Vuichard N (2013) Climate change projections using the IPSL-CM5 Earth System Model: from CMIP3 to CMIP5. *Clim Dyn* 40(9):2123–2165. <https://doi.org/10.1007/s00382-012-1636-1>
- Eden J, Widmann M, Grawe D, Rast S (2012) Skill, correction, and downscaling of GCM-simulated precipitation. *J Clim* 25:3970–3984. <https://doi.org/10.1175/JCLI-D-11-00254.1>
- Fisher RA (1915) Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika* 10(4):507–521
- François B, Vrac M, Cannon AJ, Robin Y, Allard D (2020) Multivariate bias corrections of climate simulations: which benefits for which losses? *Earth Syst Dyn* 2020:1–41. <https://doi.org/10.5194/esd-2020-10>
- Gagne DJ II, Christensen HM, Subramanian AC, Monahan AH (2020) Machine learning for stochastic parameterization: generative adversarial networks in the Lorenz ‘96 model. *J Adv Model Earth Syst* 12(3):e2019MS001896. <https://doi.org/10.1029/2019MS001896>
- Gan Z, Chen L, Wang W, Pu Y, Zhang Y, Liu H, Li C, Carin L (2017) Triangle generative adversarial networks. [arXiv:1709.06548](https://arxiv.org/abs/1709.06548)
- Gauthier J (2014) Conditional generative adversarial nets for convolutional face generation. In: *Class Project for Stanford CS231N: convolutional neural networks for visual recognition*, Winter semester vol. 5, p 2
- Gokaslan A, Ramanujan V, Ritchie D, Kim KI, Tompkin J (2019) Improving shape deformation in unsupervised image-to-image translation. [arXiv:1808.04325](https://arxiv.org/abs/1808.04325)
- Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2014) Generative adversarial nets. *Adv Neural Inf Process Syst*. <https://doi.org/10.1145/3422622>
- Gudmundsson L, Bremnes JB, Haugen JE, Engen-Skaugen T (2012) Technical note: downscaling RCM precipitation to the station scale using statistical transformations—a comparison of methods. *Hydrol Earth Syst Sci* 16(9):3383–3390. <https://doi.org/10.5194/hess-16-3383-2012>
- Guinard K, Mailhot A, Caya D (2015) Projected changes in characteristics of precipitation spatial structures over North America. *Int J Climatol* 35:596–612. <https://doi.org/10.1002/joc.4006>
- Guo Q, Chen J, Zhang X, Shen M, Chen H, Guo S (2019) A new two-stage multivariate quantile mapping method for bias correcting climate model outputs. *Clim Dyn* 53(5):3603–3623. <https://doi.org/10.1007/s00382-019-04729-w>
- Gutmann E, Pruiett T, Clark M, Brekke L, Arnold J, Raff D, Rasmussen R (2014) An intercomparison of statistical downscaling methods used for water resource assessments in the United States. *Water Resour Res* 50:7167–7186. <https://doi.org/10.1002/2014WR015559>
- Haddad Z, Rosenfeld D (1997) Optimality of empirical Z-R relations. *Q J R Meteor Soc* 123(541):1283–1293. <https://doi.org/10.1002/qj.49712354107>
- He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: *2016 IEEE Conference on computer vision and pattern recognition (CVPR)*, pp 770–778. <https://doi.org/10.1109/CVPR.2016.90>
- Hnilica J, Hanel M, Puš V (2017) Multisite bias correction of precipitation data from regional climate models. *Int J Climatol* 37:2934–2946. <https://doi.org/10.1002/joc.4890>

- IPCC (2014) Climate change 2014: synthesis report. In: Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]. IPCC, Geneva, Switzerland, p 151. <https://www.ipcc.ch/report/ar5/syr/>
- Isola P, Zhu JY, Zhou T, Efros AA (2017) Image-to-image translation with conditional adversarial networks. In: 2017 IEEE Conference on computer vision and pattern recognition (CVPR), pp 5967–5976. <https://doi.org/10.1109/CVPR.2017.632>
- Jordan C (1874a) Mémoire sur les formes bilinéaires. *J Math Pures Appl* 19(Deuxième Série):35–54
- Jordan C (1874b) Sur la réduction des formes bilinéaires. *C R Acad Sci Paris* 78(Deuxième Série):614–617
- Karras T, Aila T, Laine S, Lehtinen J (2018) Progressive growing of GANs for improved quality, stability, and variation. [arXiv:1710.10196](https://arxiv.org/abs/1710.10196)
- Kim T, Cha M, Kim H, Lee JK, Kim J (2017) Learning to discover cross-domain relations with generative adversarial networks. [arXiv:1703.05192](https://arxiv.org/abs/1703.05192)
- Kingma DP, Ba J (2017) Adam: a method for stochastic optimization. [arXiv:1412.6980](https://arxiv.org/abs/1412.6980)
- Lecun Y, Bengio Y (1995) Convolutional networks for images, speech, and time-series. In: Arbib MA (ed) *The handbook of brain theory and neural networks*. MIT Press, Cambridge, MA, pp 255–258
- Leinonen J, Berne A (2020) Unsupervised classification of snowflake images using a generative adversarial network and *K*-medoids classification. *Atmos Meas Tech* 13(6):2949–2964. <https://doi.org/10.5194/amt-13-2949-2020>
- Leinonen J, Nerini D, Berne A (2020) Stochastic super-resolution for downscaling time-evolving atmospheric fields with a generative adversarial network. *IEEE Trans Geosci Remote Sens*. <https://doi.org/10.1109/TGRS.2020.3032790>
- Liu Y, Racah E, Prabhat, Correa J, Khosrowshahi A, Lavers D, Kunkel K, Wehner M, Collins W (2016) Application of deep convolutional neural networks for detecting extreme weather in climate datasets. [arXiv:1605.01156](https://arxiv.org/abs/1605.01156)
- Mao X, Li Q, Xie H, Lau RYK, Wang Z, Smolley SP (2017) Least squares generative adversarial networks. [arXiv:1611.04076](https://arxiv.org/abs/1611.04076)
- Maraun D (2013) Bias correction, quantile mapping, and downscaling: revisiting the inflation issue. *J Clim* 26(6):2137–2143. <https://doi.org/10.1175/JCLI-D-12-00821.1>
- Maraun D (2016) Bias correcting climate change simulations—a critical review. *Curr Clim Chang Rep* 2:211–220. <https://doi.org/10.1007/s40641-016-0050-x>
- Maraun D, Wetterhall F, Ireson AM, Chandler RE, Kendon EJ, Widmann M, Brienen S, Rust HW, Sauter T, Themeßl M, Venema VKC, Chun KP, Goodess CM, Jones RG, Onof C, Vrac M, Thiele-Eich I (2010) Precipitation downscaling under climate change: recent developments to bridge the gap between dynamical models and the end user. *Rev Geophys*. <https://doi.org/10.1029/2009RG000314>
- Marti O, Braconnot P, Dufresne J-L, Bellier J, Benshila R, Bony S, Brockmann P, Cadule P, Caubel A, Codron F, de Noblet N, Denvil S, Fairhead L, Fichefet T, Foujols M-A, Friedlingstein P, Goosse H, Grandpeix J-Y, Guilyardi E, Hourdin F, Idelkadi A, Kageyama M, Krinner G, Lévy C, Madec G, Mignot J, Musat I, Swingedouw D, Talandier C (2010) Key features of the IPSL ocean atmosphere model and its sensitivity to atmospheric resolution. *Clim Dyn* 34:1–26. <https://doi.org/10.1007/S00382-009-0640-6>
- Mehrotra R, Sharma A (2016) A multivariate quantile-matching bias correction approach with auto- and cross-dependence across multiple time scales: implications for downscaling. *J Clim* 29(10):3519–3539. <https://doi.org/10.1175/JCLI-D-15-0356.1>
- Mehrotra R, Sharma A (2019) A resampling approach for correcting systematic spatiotemporal biases for multiple variables in a changing climate. *Water Resour Res* 55(1):754–770. <https://doi.org/10.1029/2018WR023270>
- Menick J, Kalchbrenner N (2018) Generating high fidelity images with subscale pixel networks and multidimensional upscaling. [arXiv:1812.01608](https://arxiv.org/abs/1812.01608)
- Mirza M, Osindero S (2014) Conditional generative adversarial nets. [arXiv:1411.1784](https://arxiv.org/abs/1411.1784)
- Mueller B, Seneviratne S (2014) Systematic land climate and evapotranspiration biases in CMIP5 simulations. *Geophys Res Lett* 41:128–134. <https://doi.org/10.1002/2013GL058055>
- Muerth MJ, Gauvin St-Denis B, Ricard S, Velázquez JA, Schmid J, Minville M, Caya D, Chaumont D, Ludwig R, Turcotte R (2013) On the need for bias correction in regional climate scenarios to assess climate change impacts on river runoff. *Hydrol Earth Syst Sci* 17(3):1189–1204. <https://doi.org/10.5194/hess-17-1189-2013>
- Nahar J, Johnson F, Sharma A (2018) Addressing spatial dependence bias in climate model simulations—an independent component analysis approach. *Water Resour Res* 54(2):827–841. <https://doi.org/10.1002/2017WR021293>
- Nguyen H, Mehrotra R, Sharma A (2019) Correcting systematic biases across multiple atmospheric variables in the frequency domain. *Clim Dyn* 52:1283–1298. <https://doi.org/10.1007/s00382-018-4191-6>
- Piani C, Haerter J (2012) Two dimensional bias correction of temperature and precipitation copulas in climate models. *Geophys Res Lett* 39(L20):401. <https://doi.org/10.1029/2012GL053839>
- Racah E, Beckham C, Maharaj T, Kahou SE, Prabhat, Pal C (2017) ExtremeWeather: a large-scale climate dataset for semi-supervised detection, localization, and understanding of extreme weather events. [arXiv:1612.02095](https://arxiv.org/abs/1612.02095)
- Radford A, Metz L, Chintala S (2016) Unsupervised representation learning with deep convolutional generative adversarial networks. [arXiv:1511.06434](https://arxiv.org/abs/1511.06434)
- Ramirez-Villegas J, Challinor A, Thornton P, Jarvis A (2013) Implications of regional improvement in global climate models for agricultural impact research. *Environ Res Lett* 8(024):018. <https://doi.org/10.1088/1748-9326/8/2/024018>
- Randall D, Wood R, Bony S, Colman R, Fichefet T, Fyfe J, Kattsov V, Pitman A, Shukla J, Srinivasan J, Ronald S, Sumi A, Taylor K (2007) *Climate models and their evaluation*. Cambridge University Press, Cambridge, pp 589–662
- Reichler T, Kim J (2008) how well do coupled models simulate today's climate? *Bull Am Meteorol Soc* 89:303–311. <https://doi.org/10.1175/BAMS-89-3-303>
- Reichstein M, Camps-Valls G, Stevens B, Jung M, Denzler J, Carvalhais N, Prabhat M (2019) Deep learning and process understanding for data-driven Earth system science. *Nature* 566:195–204. <https://doi.org/10.1038/s41586-019-0912-1>
- Robin Y, Vrac M, Naveau P, Yiou P (2019) Multivariate stochastic bias corrections with optimal transport. *Hydrol Earth Syst Sci* 23(2):773–786. <https://doi.org/10.5194/hess-23-773-2019>
- Rodrigues ER, Oliveira I, Cunha RLF, Netto MAS (2018) DeepDownscale: a deep learning strategy for high-resolution weather forecast. In: 2018 IEEE 14th International Conference on e-Science (e-Science), pp 415–422. <https://doi.org/10.1109/eScience.2018.00130>
- Roth K, Lucchi A, Nowozin S, Hofmann T (2017) Stabilizing training of generative adversarial networks through regularization. [arXiv:1705.09367](https://arxiv.org/abs/1705.09367)
- Royer A, Bousmalis K, Gouws S, Bertsch F, Mosseri I, Cole F, Murphy K (2020) XGAN: unsupervised image-to-image translation for many-to-many mappings. Springer International Publishing, pp 33–49. https://doi.org/10.1007/978-3-030-30671-7_3
- Räty O, Räisänen J, Bosshard T, Donnelly C (2018) Intercomparison of univariate and joint bias correction methods in changing climate from a hydrological perspective. *Climate* 6:33. <https://doi.org/10.3390/cli6020033>

- Salimans T, Goodfellow I, Zaremba W, Cheung V, Radford A, Chen X (2016) Improved techniques for training GANs. [arXiv:1606.03498](https://arxiv.org/abs/1606.03498)
- Scher S, Messori G (2018) Predicting weather forecast uncertainty with machine learning. *Q J R Meteorol Soc* 144(717):2830–2841. <https://doi.org/10.1002/qj.3410>
- Scher S, Messori G (2019) Weather and climate forecasting with neural networks: using general circulation models (GCMs) with different complexity as a study ground. *Geosci Model Dev* 12(7):2797–2809. <https://doi.org/10.5194/gmd-12-2797-2019>
- Scher S, Peßenteiner S (2020) Technical note: temporal disaggregation of spatial rainfall fields with generative adversarial networks. *Hydrol Earth Syst Sci* 2020:1–23. <https://doi.org/10.5194/hess-2020-464>
- Schmidhuber J (2015) Deep learning in neural networks: an overview. *Neural Netw* 61:85–117. <https://doi.org/10.1016/j.neunet.2014.09.003>
- Shi X, Chen Z, Wang H, Yeung DY, Wong W, Woo W (2015) Convolutional LSTM network: a machine learning approach for precipitation nowcasting. [arXiv:1506.04214](https://arxiv.org/abs/1506.04214)
- Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R (2014) Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res* 15(56):1929–1958
- Stewart GW (1993) On the early history of the singular value decomposition. *SIAM Rev* 35(4):551–566. <https://doi.org/10.1137/1035134>
- Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A (2015) Going deeper with convolutions. In: 2015 IEEE Conference on computer vision and pattern recognition (CVPR), pp 1–9. <https://doi.org/10.1109/CVPR.2015.7298594>
- Székelly G, Rizzo M (2004) Testing for equal distributions in high dimension. *InterStat* 5:1249–1272
- Székelly G, Rizzo M (2013) Energy statistics: a class of statistics based on distances. *J Stat Plan Inference* 143:1249–1272. <https://doi.org/10.1016/j.jspi.2013.03.018>
- Teutschbein C, Seibert J (2012) Bias correction of regional climate model simulations for hydrological climate-change impact studies: review and evaluation of different methods. *J Hydrol* 456:12–29. <https://doi.org/10.1016/j.jhydrol.2012.05.052>
- Tong Y, Gao X, Han Z, Xu Y, Xu Y, Giorgi F (2020) Bias correction of temperature and precipitation over China for RCM simulations using the QM and QDM methods. *Clim Dyn*. <https://doi.org/10.1007/s00382-020-05447-4>
- Tramblay Y, Ruelland D, Somot S, Bouaicha R, Servat E (2013) High-resolution Med-CORDEX regional climate model simulations for hydrological impact studies: a first evaluation of the ALADIN-Climate model in Morocco. *Hydrol Earth Syst Sci* 17(10):3721–3739. <https://doi.org/10.5194/hess-17-3721-2013>
- Van Loon A, Gleeson T, Clark J, van Dijk A, Stahl K, Hannaford J, Di Baldassarre G, Teuling A, Tallaksen L, Uijlenhoet R, Hannah D, Sheffield J, Svoboda M, Verbeiren B, Wagener T, Rangelcroft S, Wanders N, Van Lanen H (2016) Drought in the anthropocene. *Nat Geosci* 9:89–91. <https://doi.org/10.1038/ngeo2646>
- Vandal T, Kodra E, Ganguly S, Michaelis A, Nemani R, Ganguly AR (2017) DeepSD: generating high resolution climate change projections through single image super-resolution. In: Proceedings of the 23rd ACM SIGKDD International Conference on knowledge discovery and data mining, pp 1663–1672. <https://doi.org/10.1145/3097983.3098004>
- Vidal JP, Martin E, Franchistéguy L, Baillon M, Soubeyrou JM (2010) A 50-year high-resolution atmospheric reanalysis over France with the Safran system. *Int J Climatol* 30(11):1627–1644. <https://doi.org/10.1002/joc.2003>
- Vigaud N, Vrac M, Caballero Y (2013) Probabilistic downscaling of GCM scenarios over southern India. *Int J Climatol* 33:1248–1263. <https://doi.org/10.1002/joc.3509>
- Vorogushyn S, Bates PD, de Bruijn K, Castellarin A, Kreibich H, Priest S, Schröter K, Bagli S, Blöschl G, Domeneghetti A, Gouldby B, Klijn F, Lammersen R, Neal JC, Ridder N, Terink W, Viavattene C, Viglione A, Zanardo S, Merz B (2018) Evolutionary leap in large-scale flood risk assessment needed. *WIREs Water* 5(2):e1266. <https://doi.org/10.1002/wat2.1266>
- Vrac M (2018) Multivariate bias adjustment of high-dimensional climate simulations: the rank resampling for distributions and dependences (R^2D^2) bias correction. *Hydrol Earth Syst Sci* 22(6):3175–3196. <https://doi.org/10.5194/hess-22-3175-2018>
- Vrac M, Thao S (2020) R^2D^2 v2.0: accounting for temporal dependences in multivariate bias correction via analogue ranks resampling. *Geosci Model Dev* 2020:1–29. <https://doi.org/10.5194/gmd-2020-132>
- Vrac M, Drobinski P, Merlo A, Herrmann M, Lavaysse C, Li L, Somot S (2012) Dynamical and statistical downscaling of the French Mediterranean climate: uncertainty assessment. *Nat Hazards Earth Syst Sci* 12(9):2769–2784. <https://doi.org/10.5194/nhess-12-2769-2012>
- Vrac M, Noël T, Vautard R (2016) Bias correction of precipitation through singularity stochastic removal: because occurrences matter. *J Geophys Res Atmos* 121:5237–5258. <https://doi.org/10.1002/2015JD024511>
- Wang J, Liu Z, Foster I, Chang W, Kettimuthu R, Kotamarthi R (2021) Fast and accurate learned multiresolution dynamical downscaling for precipitation. [arXiv:2101.06813](https://arxiv.org/abs/2101.06813)
- Wasko C, Sharma A, Westra S (2016) Reduced spatial extent of extreme storms at higher temperatures. *Geophys Res Lett* 43(8):4026–4032. <https://doi.org/10.1002/2016GL068509>
- Wheeler T, von Braun J (2013) Climate change impacts on global food security. *Science* 341(6145):508–513. <https://doi.org/10.1126/science.1239402>
- Wilcke RAI, Mendlik T, Gobiet A (2013) Multi-variable error correction of regional climate models. *Clim Change* 120:871–887. <https://doi.org/10.1007/s10584-013-0845-x>
- Wilks DS (2006) *Statistical methods in the atmosphere science*. Academic Press
- Wu JL, Kashinath K, Albert A, Chirila D, Prabhat Xiao H (2020) Enforcing statistical constraints in generative adversarial networks for modeling chaotic dynamical systems. *J Comput Phys* 406(109):209. <https://doi.org/10.1016/j.jcp.2019.109209>
- Xie Y, Franz E, Chu M, Thuerey N (2018) TempoGAN: a temporally coherent, volumetric GAN for super-resolution fluid flow. *ACM Trans Graph*. <https://doi.org/10.1145/3197517.3201304>
- Xu CY (1999) From GCMs to river flow: a review of downscaling methods and hydrologic modelling approaches. *Prog Phys Geogr* 23:229–249. <https://doi.org/10.1177/030913339902300204>
- Yi Z, Zhang H, Tan P, Gong M (2017) DualGAN: unsupervised dual learning for image-to-image translation. In: 2017 IEEE International Conference on computer vision (ICCV), pp 2868–2876. <https://doi.org/10.1109/ICCV.2017.310>
- Yoo D, Kim N, Park S, Paek AS, Kweon IS (2016) Pixel-level domain transfer. [arXiv:1603.07442](https://arxiv.org/abs/1603.07442)
- Zhu JY, Park T, Isola P, Efros AA (2017) unpaired image-to-image translation using cycle-consistent adversarial networks. [arXiv:1703.10593](https://arxiv.org/abs/1703.10593)
- Zscheischler J, Westra S, Hurk B, Seneviratne S, Ward P, Pitman A, AghaKouchak A, Bresch D, Leonard M, Wahl T, Zhang X (2018) Future climate risk from compound events. *Nat Clim Change*. <https://doi.org/10.1038/s41558-018-0156-3>
- Zscheischler J, Fischer E, Lange S (2019) The effect of univariate bias adjustment on multivariate hazard estimates. *Earth Syst Dyn* 10:31–43. <https://doi.org/10.5194/esd-10-31-2019>