

Adjusting spatial dependence of climate model outputs with Cycle-Consistent Adversarial Networks

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

LSCE: Laboratoire des sciences du climat et de l'environnement <https://orcid.org/0000-0002-9471-026X>

Soulivanh Thao

LSCE: Laboratoire des sciences du climat et de l'environnement

Mathieu Vrac

LSCE: Laboratoire des sciences du climat et de l'environnement

Research Article

Keywords: bias correction, spatial dependence, post-processing, climate simulations, generative adversarial networks, model output statistics

Posted Date: March 17th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-299929/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Climate Dynamics on July 30th, 2021. See the published version at <https://doi.org/10.1007/s00382-021-05869-8>.

1 **Adjusting spatial dependence of climate model**
2 **outputs with Cycle-Consistent Adversarial Networks**

3 **Bastien François · Soulivanh Thao ·**
4 **Mathieu Vrac**

5
6 Received: date / Accepted: date

7 **Abstract** Climate model outputs are commonly corrected using statistical
8 univariate bias correction methods. Most of the time, those 1d-corrections do
9 not modify the ranks of the time series to be corrected. This implies that bi-
10 ases in the spatial or inter-variable dependences of the simulated variables are
11 not adjusted. Hence, over the last few years, some multivariate bias correction
12 (MBC) methods have been developed to account for inter-variable structures,
13 inter-site ones, or both. As proof-of-concept, we propose to adapt a computer
14 vision technique used for Image-to-Image translation tasks (CycleGAN) for
15 the adjustment of spatial dependence structures of climate model projections.
16 The proposed algorithm, named MBC-CycleGAN, aims to transfer simulated
17 maps (seen as images) with inappropriate spatial dependence structure from
18 climate model outputs to more realistic images with spatial properties similar
19 to the observed ones. For evaluation purposes, the method is applied to adjust
20 maps of temperature and precipitation from climate simulations through two
21 cross-validation approaches. The first one is designed to assess two different
22 post-processing schemes (Perfect Prognosis and Model Output Statistics). The
23 second one assesses the influence of nonstationary properties of climate simu-
24 lations on the performance of MBC-CycleGAN to adjust spatial dependences.
25 Results are compared against a popular univariate bias correction method,

Bastien François

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

Tel.: +33-169086000

Fax: +33-169087716

E-mail: bastien.francois@lsce.ipsl.fr

Soulivanh Thao

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

Mathieu Vrac

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

26 a “quantile-mapping” method, which ignores inter-site dependencies in the
27 correction procedure, and two state-of-the-art multivariate bias correction al-
28 gorithms aiming to adjust spatial correlation structure. In comparison with
29 these alternatives, the MBC-CycleGAN algorithm reasonably corrects spatial
30 correlations of climate simulations for both temperature and precipitation,
31 encouraging further research on the improvement of this approach for multi-
32 variate bias correction of climate model projections.

33 **Keywords** bias correction · spatial dependence · post-processing · climate
34 simulations · generative adversarial networks · model output statistics

35 1 Introduction

36 With ongoing climate change, mitigation and adaptation strategies have to
37 be anticipated by decision makers in order to reduce potential future conse-
38 quences of climate change on human societies and activities (IPCC 2014). Such
39 consequences are commonly assessed through climate change impact studies,
40 for instance in hydrology (e.g., Bates et al 2008), agronomy (e.g., Wheeler and
41 von Braun 2013) or epidemiology (e.g., Caminade et al 2014). They rely on im-
42 pact model simulations, the quality of which highly depends on the reliability
43 of the climate information used as inputs (e.g., Muerth et al 2013; Ramirez-
44 Villegas et al 2013). Besides observations, global and regional climate models
45 (GCM and RCM) are the major tools to understand the climate system and its
46 evolutions in the future (Randall et al 2007; Reichler and Kim 2008). However,
47 despite considerable improvements in climate modelling, climate simulations
48 often remain biased compared to observations: even for the current climate,
49 key statistical features such as mean, variance or the dependence structures
50 between physical variables or between sites can differ from those calculated for
51 observational references (e.g., Eden et al 2012; Cattiaux et al 2013; Mueller
52 and Seneviratne 2014). Consequently, biases are expected to be present in
53 climate projections for future periods, making bias correction an often un-
54 avoidable data pre-processing step for impact studies (e.g., Christensen et al
55 2008; Maraun et al 2010; Teutschbein and Seibert 2012).

56 In the recent years, many statistical bias correction (BC) methods have
57 been developed that aim to correct (selected features of) the distribution of
58 climate variables. The idea of statistical bias correction is to find a mathe-
59 matical transformation that makes climate simulations have similar statistical
60 properties as a reference dataset over the historical period, and then apply
61 this transformation for the modeled projection. Such transformations may be
62 determined with statistical models based on either perfect prognosis (PP) or
63 model output statistics (MOS) approaches (Maraun et al 2010). The PP ap-
64 proach consists in determining the statistical link between a variable of interest
65 from references (predictand) and one or several observed variables (predictors)
66 occurring at the same time. Simultaneous values of predictand and predictors
67 are indeed required to implement the PP approach and learn the (synchronous)
68 relationships between them. By applying these relationships to predictors from

69 climate simulations, this approach implicitly makes the assumption that these
70 predictors are realistically simulated (Wilks 2006). In the MOS approach, ob-
71 served and simulated variables are not considered to be synchronized in time,
72 and biases relate to differences in some statistics (such as means or variances)
73 or in distributions between references and modeled climate variables. Adjust-
74 ments can be made to the simulated mean (e.g., Delta method, Xu 1999),
75 variance (e.g., simple scaling adjustment, Berg et al 2012) and also all mo-
76 ments of higher order and percentiles (e.g., “quantile-mapping”, Haddad and
77 Rosenfeld 1997; Déqué 2007; Gudmundsson et al 2012). In particular, quantile-
78 mapping technique has received a keen interest since it permits for adjusting
79 not only the mean and variance but also the whole distribution of climate
80 variables. It has been conducive to the development of many variants (e.g.,
81 Vrac et al 2012; Trambly et al 2013; Cannon et al 2015; Vrac et al 2016), and
82 applied for various studies (e.g., Vigaud et al 2013; Defrance et al 2017; Bartok
83 et al 2019; Tong et al 2020). However, such BC methods are designed to only
84 correct statistical aspects of univariate distributions. Simulated variables are
85 indeed adjusted separately for each physical variable at each specific location.
86 Thus, potential biases in the spatial dependence structure of modeled variables
87 are not corrected (e.g., Wilcke et al 2013), which can generate corrections with
88 inappropriate multivariate situations and can affect subsequent analyses that
89 depend on spatial characteristics of climate variables (e.g., Zscheischler et al
90 2019). For instance, this can occur with flood risk assessment, that depends
91 on spatial (and temporal) properties of precipitation, soil moisture and river
92 flow (Vorogushyn et al 2018) or with drought-related impacts, that depend on
93 complex interaction of natural and anthropogenic processes (Van Loon et al
94 2016). It is hence crucial to provide end users with bias corrections of climate
95 simulations that present not only relevant 1-dimensional information at each
96 individual site but also appropriate spatial representation.

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

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

- 112 – the “marginal/dependence” correction approach, that consists of MBC
113 methods adjusting in two distinct steps, i.e. separately, marginal distribu-

- 114 tions and multivariate dependencies of climate simulations (e.g., Bárdossy
115 and Pegram 2012; Mehrotra and Sharma 2016; Hnilica et al 2017; Nahar
116 et al 2018; Cannon 2018; Nguyen et al 2019; Guo et al 2019; Vrac and Thao
117 2020).
- 118 – the “successive conditional” category, made up of MBC methods perform-
119 ing successive univariate corrections of climate variables conditionally on
120 the previously adjusted ones (e.g., Piani and Haerter 2012; Dekens et al
121 2017).
 - 122 – the “all-in-one” correction approach, that adjusts directly the whole sta-
123 tistical distribution (i.e. both univariate and multivariate properties) of
124 climate simulations at the same time (e.g., Robin et al 2019).

125 Based on this categorization, François et al (2020) performed an intercompar-
126 ison and critical review of MBC methods. It presents a global picture of the
127 performances of MBCs in terms of multivariate adjustments of climate simu-
128 lations, as well as the different assumptions and statistical techniques used.

129 In parallel, i.e., in contexts other than bias correction, over the last decades,
130 machine learning techniques have emerged as a promising approach to model
131 highly nonlinear and complex relationships between statistical variables. Ma-
132 jor improvements have been obtained with Deep Learning models (see the
133 overview of Schmidhuber 2015), which have proved to be efficient to extract
134 high-level feature information from various datasets. In particular, convolu-
135 tional neural networks (CNNs, see e.g., Lecun and Bengio 1995) showed that
136 they can capture with great performances complex spatial structures. Initially
137 developed for computer vision problems (e.g., Szegedy et al 2015; He et al
138 2016), they found numerous applications in climate sciences: for instance for
139 weather forecast prediction uncertainty (Scher and Messori 2018), emulations
140 of atmospheric dynamics (Shi et al 2015; Scher and Messori 2019; Chapman
141 et al 2019), detection of extreme weather events (Liu et al 2016; Rcah et al
142 2017) and statistical downscaling (Vandal et al 2017; Rodrigues et al 2018;
143 Baño-Medina et al 2020). A recent overview of Deep Learning applications for
144 Earth system science is offered by Reichstein et al (2019).

145 Recently, a new class of artificial neural networks, named Generative Ad-
146 versarial Networks (GANs; Goodfellow et al 2014), has led to tremendous in-
147 terests due to their ability to infer high dimensional probability distributions.
148 Initially, this machine-learning method has been developed for estimating the
149 distribution of images from a target dataset, with the aim of sampling new
150 (and unseen) images from this distribution. GANs, implemented with deep
151 convolutional neural networks, have achieved impressive results in computer
152 vision problems (e.g., Radford et al 2016) and are a subject of active research to
153 improve computing architectures (e.g., Salimans et al 2016; Karras et al 2018;
154 Menick and Kalchbrenner 2018) and optimization techniques (e.g., Mao et al
155 2017; Arjovsky et al 2017; Roth et al 2017). Conditional formulations of GANs
156 have also been developed, for which additional information, such as class labels
157 or images, can serve as inputs to condition the generation of the new images
158 (e.g., Mirza and Osindero 2014; Gauthier 2014; Denton et al 2015; Kim et al

159 2017; Isola et al 2017). In particular, image-conditional GANs permit to per-
160 form image-to-image translation tasks by learning how to map the statistical
161 distribution of one set of images (source dataset) to the statistical distribution
162 of another set (target dataset). Depending on the correspondence between im-
163 ages of the source and target datasets, different versions of image-conditional
164 GANs have been developed. When all the images are paired (i.e., there is a
165 known one-to-one correspondence between every images of the source and tar-
166 get datasets), conditional GANs are trained by supervised learning (Yoo et al
167 2016; Isola et al 2017). When only a few images are paired, semi-supervised
168 is used (Gan et al 2017) and when all points are unpaired, only unsupervised
169 learning can be applied (Kim et al 2017; Yi et al 2017; Zhu et al 2017). Due to
170 the stochastic and high-dimensionality nature of many physical processes of
171 the Earth system, GANs and conditional GANs are particularly appealing for
172 atmospheric science problems. Recently, they have been used for various Earth-
173 science related applications: for instance for statistical downscaling (Leinonen
174 et al 2020; Wang et al 2021), temporal disaggregation of spatial rainfall fields
175 (Scher and Peßenteiner 2020), sampling of extreme values (Bhatia et al 2020),
176 modelling of chaotic dynamical systems (e.g., Xie et al 2018; Wu et al 2020),
177 classification of snowflake images (Leinonen and Berne 2020), weather fore-
178 casting (Bihlo 2020) and stochastic parameterization in geophysical models
179 (Gagne II et al 2020).

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

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

The paper is organized as follows: Section 2 presents the model and reference data used, and Section 3 describes the MBC-CycleGAN algorithm. Then, Section 4 displays the experimental setup used in this study, and results are provided in Section 5. Conclusions, discussions and perspectives for future research are finally proposed in Section 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 $8 \text{ km} \times 8 \text{ km}$ 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.

3 Methodology

3.1 GAN

In its most basic formulation, a generative adversarial network consists of two neural networks that are trained conjointly: 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,

243 such as images of the target dataset (i.e., references), and hence takes its values
 244 in a high-dimensional space. We assume to have at hand samples $\mathbf{y}_1, \dots, \mathbf{y}_n$
 245 drawn according to the density $\mathbb{P}_{\mathbf{Y}}$ on \mathbb{R}^d . The generator, denoted G , is a
 246 function from $\mathbb{R}^{d'}$ to \mathbb{R}^d and is intended to be applied to a d' -dimensional ran-
 247 dom variable \mathbf{W} , usually multivariate Gaussian random noise (with $d' \ll d$),
 248 such that the random variable $G(\mathbf{W})$ follows the law of \mathbf{Y} , i.e. $\mathbb{P}_{\mathbf{Y}} = \mathbb{P}_{G(\mathbf{W})}$.
 249 Let $\mathbf{w}_1, \dots, \mathbf{w}_n$ be a sample drawn from the distribution of \mathbf{W} . To train the
 250 generator G , the discriminator $D_{\mathbf{Y}}$, that is a function from \mathbb{R}^d to $[0, 1]$, is
 251 used as complex loss function (Goodfellow et al 2014). This neural network
 252 is a binary classifier that returns the probability that a given observation, or
 253 image, comes from $\mathbb{P}_{\mathbf{Y}}$. The discriminator is trained in a supervised way to
 254 return maximal probability values on the reference images \mathbf{y}_i and minimal
 255 values on the artificially generated images $G(\mathbf{w}_i)$. Conversely, the goal of the
 256 generator is to “fool” the discriminator by making the distribution of $G(\mathbf{w}_i)$
 257 as indistinguishable as possible from that of \mathbf{y}_i , i.e., making difficult for the
 258 discriminator to determine that a sample $G(\mathbf{w}_i)$ comes from a distribution
 259 different from $\mathbb{P}_{\mathbf{Y}}$. Generator and discriminator are trained in turns and are in
 260 competition (i.e. “adversarial training”) to improve themselves until it reaches
 261 an optimal equilibrium state.

262 The original formulation of GANs explained above is unconditional: the
 263 generator G only takes as input noise vectors \mathbf{w}_i to produce new samples that
 264 are drawn from the target distribution $\mathbb{P}_{\mathbf{Y}}$. The idea of conditional GANs (e.g.,
 265 Goodfellow et al 2014; Mirza and Osindero 2014) is to add some information
 266 as inputs to direct the generation. By conditioning the generation on an input
 267 image, the generator is able to generate a corresponding output image, render-
 268 ing the conditional GANs appropriate for image-to-image translation tasks
 269 (e.g., Isola et al 2017).

270 3.2 CycleGAN for unsupervised image-to-image translation

271 CycleGAN (Zhu et al 2017) is a particular image-conditional GANs that is
 272 commonly used for unsupervised image-to-image translation. In the origi-
 273 nal application, CycleGAN has been applied with great success to transform
 274 photographs into the styles of master paintings by modifying colour infor-
 275 mation (i.e., RGB colour channels and/or spatial features of colours) of the
 276 photographs. Instead of the random noise \mathbf{W} , we introduce another random
 277 variable \mathbf{X} , with probability distribution $\mathbb{P}_{\mathbf{X}}$, living in the same dimensional
 278 space as \mathbf{Y} (i.e., \mathbb{R}^d). This random variable \mathbf{X} characterizes the images of the
 279 source dataset (i.e., biased simulations to correct). The CycleGAN approach
 280 consists in learning a mapping (i.e., a generator) $G_{\mathbf{X} \rightarrow \mathbf{Y}} : \mathbb{R}^d \rightarrow \mathbb{R}^d$ such that
 281 the random variable $G_{\mathbf{X} \rightarrow \mathbf{Y}}(\mathbf{X})$ follows the law of \mathbf{Y} (i.e., $\mathbb{P}_{\mathbf{Y}} = \mathbb{P}_{G_{\mathbf{X} \rightarrow \mathbf{Y}}(\mathbf{X})}$).
 282 In addition to samples $\mathbf{y}_1, \dots, \mathbf{y}_n$, we assume to have at hand image samples
 283 $\mathbf{x}_1, \dots, \mathbf{x}_n$ drawn according to density $\mathbb{P}_{\mathbf{X}}$ on \mathbb{R}^d . Similarly as unconditional
 284 GANs, the mapping $G_{\mathbf{X} \rightarrow \mathbf{Y}}$ is learned using an adversarial loss, i.e. with a dis-
 285 criminator $D_{\mathbf{Y}}$ 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:

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

$G_{\mathbf{X} \rightarrow \mathbf{Y}}$ 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 $D_{\mathbf{Y}}$:

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

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 the optimization process of GANs fails to converge, leading to mapping all input images \mathbf{x}_i to the same image, let say \mathbf{y}_k . 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:

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

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

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

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

$$\begin{aligned} L(G_{\mathbf{X} \rightarrow \mathbf{Y}}, G_{\mathbf{Y} \rightarrow \mathbf{X}}, D_{\mathbf{X}}, D_{\mathbf{Y}}) = & L_{GAN}(G_{\mathbf{X} \rightarrow \mathbf{Y}}, D_{\mathbf{Y}}) + L_{GAN}(G_{\mathbf{Y} \rightarrow \mathbf{X}}, D_{\mathbf{X}}) \\ & + \lambda_{cyc} L_{cyc}(G_{\mathbf{X} \rightarrow \mathbf{Y}}, G_{\mathbf{Y} \rightarrow \mathbf{X}}) \\ & + \lambda_{id} L_{id}(G_{\mathbf{X} \rightarrow \mathbf{Y}}, G_{\mathbf{Y} \rightarrow \mathbf{X}}), \end{aligned} \quad (5)$$

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

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

327 Although estimating the inverse mapping $G_{\mathbf{Y} \rightarrow \mathbf{X}}$ is not necessarily the initial
 328 goal of many image-to-image translation problems, its use to constrain the
 329 optimization problem has been found to be crucial in an unsupervised context
 330 for the convergence of the algorithm and the estimation of the desired mapping
 331 $G_{\mathbf{X} \rightarrow \mathbf{Y}}$. Illustrations of the adversarial, cycle-consistent and identity losses
 332 within the CycleGAN architecture are given in Fig. 2.

333 3.3 The MBC-CycleGAN approach

334 3.3.1 Adaptation of CycleGAN for MBC

335 The main idea of the proposed methodology, named MBC-CycleGAN, is to
 336 adapt the CycleGAN approach so that it turns daily maps of a simulated
 337 variable with spatial features inappropriate compared to a reference dataset,
 338 to more realistic maps. Here, MBC-CycleGAN is developed in the context of
 339 the ‘‘marginal/dependence’’ MBC category, i.e., correcting separately marginal
 340 distributions and dependence relationships. In addition to marginal distribu-
 341 tions, we consider the adjustment of spatial dependence structures. The algo-
 342 rithm is trained on a historical period (i.e., calibration) for which both climate
 343 simulations and reference datasets are available. Once the adversarial neural
 344 network has converged, adjustment of climate simulations over a projection pe-
 345 riod (e.g., a future time period) is performed using the pretrained algorithm.
 346 The MBC-CycleGAN proceeds as follows:

- 347 1. As MBC-CycleGAN belongs to the marginal/dependence category, univariate
348 distributions of modeled climate variables are first corrected indepen-
349 dently using a univariate BC method for both calibration and projection
350 periods. In this study, the quantile-quantile (QQ) mapping method is used
351 (Déqué 2007).
- 352 2. Then, quantile-quantile and reference data over the calibration time pe-
353 riod are transformed to belong to $[0, 1]$ using a pointwise min-max normal-
354 ization. For each grid cell, the minimum and maximum values from the
355 reference during the calibration are taken to compute the normalization.
356 The resulting daily maps are then given to a CycleGAN model to learn
357 the transfer between the two distributions of images. Generators and dis-
358 criminators are trained until the spatial distribution of the corrected maps
359 stops improving. More details about the criteria used to evaluate spatial
360 distributions are presented thereafter.
- 361 3. Once the CycleGAN model has been trained for the calibration period,
362 the same pointwise normalization is performed for quantile-quantile data
363 over the projection period, i.e., using the same minimum and maximum
364 values from the reference during the calibration period. Normalized daily
365 maps from quantile-quantile data in the projection period are translated
366 in the normalized reference domain using the pretrained adversarial neural
367 network. Then, the corrected outputs obtained are rescaled to physical
368 values by applying the inverse of the pointwise min-max normalization
369 used.
- 370 4. Finally, by taking advantage of the Schaake Shuffle technique (Clark et al
371 2004), quantile-quantile data for the projection period obtained from Step
372 1 are reordered such that the rank structure of the data obtained from Step
373 3 is reproduced. This shuffling technique, already employed in a few mul-
374 tivariate bias correction methods (e.g., Vrac 2018; Cannon 2018; Mehro-
375 tra and Sharma 2019), permits here to obtain bias-corrected data with
376 marginal properties from quantile-quantile outputs and rank dependence
377 structure from CycleGAN outputs.

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

381 3.3.2 Network Architecture

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

388 Additionally, in this paper, we only present results obtained with a simple
389 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 Section 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 B.

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. 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 B1 and B2 in Appendix B.

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, 2013) is

434 used. This metric, already used in the bias correction literature (e.g., Cannon
 435 2018), permits to measure the statistical discrepancy between two multivariate
 436 distributions that are potentially in high dimension. Given two k -multivariate
 437 independent random vectors \mathbf{P} and \mathbf{Q} with multivariate probability distribu-
 438 tions μ and ν respectively, the energy distance \mathcal{E} between the two distributions
 439 is:

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

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

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

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

447 During training, computations of energy distances are performed every
 448 10 epochs, i.e. each time that the CycleGAN has worked 10 times through
 449 the entire training dataset. Estimated energy distances $\hat{\mathcal{E}}$ are calculated on
 450 multivariate distributions of ranks between references and bias-corrected data.
 451 It permits to assess along the training the performance of the method to correct
 452 the whole spatial dependence structure of climate simulations. Computing
 453 energy distance using ranks instead of raw values allows the removal of the
 454 influence of univariate properties on the spatial relationships. The CycleGAN
 455 model that minimizes the energy distance on ranks during training is chosen
 456 for the correction of the projection period. Training 1000 epochs takes ~ 4
 457 hours on a single NVIDIA Tesla V100 GPU.

458 4 Design of experiments

459 For evaluation purposes, the proposed MBC-CycleGAN method is applied
 460 to adjust climate simulations outputs with SAFRAN data as references. Bias
 461 correction is performed on separate seasons in order to preserve seasonal prop-
 462 erties. In the following, for sake of clarity, only the winter results are presented.
 463 Data are available for the 1979-2016 period (i.e, 3420 winter days), and need to
 464 be divided into a calibration period and a projection period to train and evalu-
 465 ate our algorithm. In accordance with common practices in machine learning,
 466 the 1979-2016 period is split as follows: 70% (2394 days) as training dataset

467 and 30% (1026 days) as evaluation dataset. In this study, two different cross-
468 validation methods - that differ in how calibration and projection periods are
469 constructed - are used to evaluate our methodology.

470 4.1 Model Output Statistics (MOS) vs. Perfect Prog (PP)

471 The first cross-validation method consists in drawing randomly the days that
472 define the calibration and projection periods. As these periods are drawn ran-
473 domly, the potential climate change signal present in the data during the 1979-
474 2016 period vanishes. Hence, for this cross-validation method, no changes in
475 marginal and dependence properties are expected between the calibration and
476 projection periods, allowing for the assessment of the method in a stationary
477 context. We take advantage of this first stationary cross-validation technique
478 to apply our method in both PP and MOS post-processing schemes for the ad-
479 justment of IPSL climate simulations. Implementing and evaluating both the
480 PP and MOS approaches in such a validation context permits to determine
481 which approach is better suited in our context of bias correction of climate
482 simulations. For the MOS approach, MBC-CycleGAN is applied directly to
483 IPSL data according to the 4 steps already described in Section 3.3. Concern-
484 ing the implementation of the PP approach, the same procedure is applied but
485 the CycleGAN model is trained in a slightly different way. Indeed, as already
486 explained in Section 1, a PP approach consists in establishing the statistical
487 relationships between large-scale predictors and local-scale predictands from
488 observational or reanalysis data (including for the predictors) before applying
489 them to climate model data. Hence, large-scale predictors temporally match-
490 ing the SAFRAN dataset are needed to a PP approach. For this purpose, a
491 new climate dataset is constructed for both temperature and precipitation as
492 follows: initial local-scale SAFRAN data with $8 \text{ km} \times 8 \text{ km}$ spatial resolution
493 are upscaled using conservative interpolation on a large-scale grid of 32 km
494 $\times 32 \text{ km}$ spatial resolution. Then, the obtained large-scale data are regrid-
495 ded using bilinear interpolation to the initial grid of SAFRAN, allowing to train
496 CycleGAN. It results in “biased” daily maps of temperature and precipitation
497 (large-scale predictors) of the initial SAFRAN data (local-scale predictands),
498 temporally matching the chronology of the SAFRAN time series. Using these
499 new data – hereafter referred to as “low-resolution (LR) SAFRAN” – a Cycle-
500 GAN model is trained for the implementation of the PP approach by learning
501 the transfer of maps from 1d-BC large-scale predictors (QQ(LR SAFRAN))
502 to maps from local-scale predictands (SAFRAN). This trained model is then
503 used to bias correct IPSL simulations over the projection period and, hence,
504 evaluate the CycleGAN results in a PP context.

505 4.2 Nonstationarity investigation

506 To evaluate the nonstationary behavior of the proposed method, a second
507 cross-validation method is defined, which consists in dividing the 1979-2016 pe-

riod 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 and SAFRAN references are likely to present disagreeing changes in their statistical (marginal and dependence) properties between calibration and projection periods, 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

553 and projection periods are in line with those from the reference. In order
554 to ease the comparison of results with the first bias correction exercise, we
555 forced IPSLbis to have the same changes as LR SAFRAN. This is reached
556 by using a two-step procedure that takes advantage of a nonstationary quan-
557 tile mapping technique for marginal changes (CDF-t, Vrac et al 2012) and a
558 matrix-recorrelation technique for dependence changes (Bárdossy and Pegram
559 2012). More details about the generation of the IPSLbis data can be found in
560 Appendix C and a detailed evaluation of the evolution of statistical proper-
561 ties of the different dataset between the calibration and projection period is
562 provided in Appendix D. In particular, results presented in Appendix D indi-
563 cate that, as expected, changes in spatial structures from SAFRAN references
564 are (globally) in agreement with those from LR SAFRAN for both tempera-
565 ture and precipitation. However, concerning changes in spatial structures for
566 IPSL simulations, conclusions are not the same depending on the physical
567 variable. While, for temperature, simulated changes of spatial correlations are
568 partially in line with those from LR SAFRAN, IPSL model presents discrep-
569 ancy of changes for precipitation. Globally, the construction of IPSLbis with
570 the two-step procedure described in Appendix C permits to impose to IPSL
571 data spatial changes for both temperature and precipitation that are in line
572 with those from LR SAFRAN.

573 4.3 Comparisons to existing MBCs: R^2D^2 and dOTC

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

586 R^2D^2 , developed in the context of marginal/dependence category, relies
587 on an analogue-based method that allows to resample ranks from a reference
588 dataset according to some conditioning information and reconstructs depen-
589 dence structure of the simulated time series. The information to condition the
590 analogues can be multivariate by considering, for example, a set of variables to
591 be corrected at a given time t . Conditioning for the ranks resampling can also
592 be extended to ranks sequences, i.e. conditioning by not only one but several
593 lagged time steps. Please note that, for the different implementations of R^2D^2
594 in this study, the multivariate conditioning used includes 4 grid points that
595 cover uniformly the region of interest. In addition, 5 lagged time steps are used

596 for the conditioning, as it has been found to stabilize the R^2D^2 method (not
597 shown). Also, the QQ method is used to correct the marginal properties for
598 R^2D^2 outputs.

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

608 Both R^2D^2 and dOTC methods are applied according to the spatial-
609 dimensional configuration (hereinafter referred to as “Spatial-”), where all the
610 784 time series for a particular physical variable are corrected jointly. While
611 R^2D^2 assumes spatial dependence structures (i.e., the rank correlations, or
612 copulas) to be stable in time, the dOTC method makes the hypothesis of non-
613 stationarity of the dependence structure between the calibration and the pro-
614 jection periods, which allows for taking into account the changes of the model
615 (e.g., due to climate change) in the bias correction procedure. Intercomparing
616 the results from both Spatial- R^2D^2 and Spatial-dOTC for adjusting spatial
617 dependence structure of climate simulations with those from MBC-CycleGAN
618 allows to better assess how the proposed method performs in a nonstationary
619 context.

620 5 Results

621 In this section, analyses are presented for the winter season (December, Jan-
622 uary and February) only. CycleGAN models are trained during the calibration
623 period and selected such that energy distances on ranks are minimized. All
624 evaluations are performed on the projection period for the corrected outputs
625 obtained from the two cross-validation methods and results are compared to
626 those from the reference dataset. For bias-corrected precipitation time series,
627 thresholding of 1 mm is applied before evaluation to replace values lower than
628 1 mm by 0. Bias correction outputs from the first and second cross-validation
629 methods are evaluated in terms of both marginal and spatial properties. Anal-
630 yses of temporal properties are only provided for outputs from the second
631 cross-validation method, in which calibration and projection periods are di-
632 vided chronologically and hence do not distort temporal properties, contrary
633 to the first cross-validation method that randomly defines these periods. To
634 assess the potential benefits of considering spatial aspects in the correction
635 procedure, the univariate QQ method (Déqué 2007) is also included in the
636 study as a benchmark.

637 5.1 MOS vs. PP

638 5.1.1 Training of MBC-CycleGANs

639 Fig. 5 shows energy distances with respect to SAFRAN references for tempera-
640 ture computed on physical values (Figs. 5a and b) and ranks (Figs. 5c and
641 d) for LR SAFRAN, plain IPSL simulations, 1d-QQ, and MBC-CycleGAN
642 (MBC-CG) outputs during the training on the calibration period. In addition,
643 results for Raw-CycleGAN (Raw-CG) are presented. Differences between
644 Raw-CG and MBC-CG only lie in their marginal properties: while Raw-CG
645 corresponds to the outputs obtained from the CycleGAN after denormaliza-
646 tion at the end of Step 3, MBC-CG is the combination of the spatial structure
647 from Raw-CG and univariate properties from QQ outputs (see the flowchart
648 provided in Fig. 3). The results for precipitation are presented in Fig. S1 of
649 the Supplement.

650 Clearly, Figs. 5a and b show large energy distances computed on physical
651 values of temperature for LR SAFRAN and IPSL datasets, indicating some
652 biases on spatial structures for those dataset with respect to SAFRAN refer-
653 ences. Adjusting marginal properties with the univariate QQ method reduces
654 values of energy distance computed on physical values, highlighting the in-
655 fluence of marginal properties on spatial features. Correction of the spatial
656 dependence structure provided by MBC-CG occurs relatively quickly, with
657 energy distances on physical variables reduced by 2 compared to QQ after
658 approximately 1000 epochs for both PP and MOS approaches. However, for
659 Raw-CG, marginal properties generated by the inverse pointwise min-max nor-
660 malization do not seem to improve values of energy distances, which justifies
661 the post-processing of univariate properties adopted in the MBC-CycleGAN
662 method with the Schaake Shuffle.

663 Figs. 5c and d show that computing energy distances on ranks for tempera-
664 ture removes the influence of univariate properties on spatial features. Energy
665 distances for both LR SAFRAN and IPSL with their respective QQ correc-
666 tions are indeed the same (Fig. 5c). The same remark holds for MBC-CG and
667 Raw-CG energy distances on ranks that have, by construction, similar spatial
668 dependence structures. As explained in subsection 3.3.3, the CycleGAN model
669 that minimizes the energy distance on ranks of MBC-CycleGAN outputs is se-
670 lected.

671 For precipitation (Fig. S1), the same conclusions hold, indicating a relative
672 ability of the CycleGAN to adjust spatial dependence structure of precipitation
673 fields. Nevertheless, contrary to temperature, one should remark that energy
674 distances on ranks are different for LR SAFRAN, IPSL and their respective
675 QQ corrections (Figs. S1c and d), which is specific to precipitation variables
676 that can contain several null values for dry events. Indeed, ranks are computed
677 here such that, when tied values are encountered, the minimum value of rank
678 is attributed to each tied value. The combination of the correction with the
679 QQ method and the thresholding for precipitation below 1 mm could modify
680 the frequency of dry events, which could result in obtaining different rank

681 structures, and hence, mechanically, different energy distances with respect to
682 SAFRAN references. This mechanism is also obtained between MBC-CG and
683 Raw-CG (Figs. S1c and d), that present different energy distances due to the
684 difference of dry events.

685 5.1.2 Univariate distribution properties

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

699 For both temperature and precipitation, the maps for the IPSL model
700 (Figs. 6c and d) present large values of mean differences with respect to the
701 SAFRAN map (Figs. 6a and b) and highlight the need to adjust univariate
702 properties of simulations. Maps provided by 1d-QQ outputs (Figs. 6e and f)
703 indicate that, as expected, the univariate method globally improves marginal
704 properties at each individual site. In agreement with the properties of the
705 marginal/dependence MBC methods, maps for MBC-CG for PP (MBC-CG-
706 PP, Figs. 6g and h) and MOS (MBC-CG-MOS, Figs. 6i and j) are exactly the
707 same as those from the 1d-QQ method. Indeed, although MBC-CG-PP and
708 MBC-CG-MOS do not use the same data for the training of the CycleGAN
709 to adjust spatial features, same marginals are taken from the QQ outputs of
710 IPSL data, which results in obtaining the same univariate properties between
711 the two corrections.

712 5.1.3 Spatial correlations

713 Quality of the corrections in terms of spatial correlations is now assessed. For
714 each grid cell, spatial dependencies are evaluated for temperature and precipi-
715 tation by computing Pearson pairwise correlations between the cell of interest
716 and each of the remaining 783 grid cells over the region of Paris for the dif-
717 ferent climate datasets. The biases of these 783 spatial Pearson correlations
718 are then summarized by computing the Mean Squared Error (MSE) with the
719 corresponding 783 correlations computed for the references. By computing the
720 MSE values for each grid cell, 784 MSE values are obtained for each climate
721 dataset and can be intercompared from one dataset to another. Fig. 7 shows

722 the boxplots of the MSE values obtained for both temperature and precip-
723 itation for the plain IPSL simulations and BC outputs. For both variables,
724 the boxplots for the IPSL simulations indicate strong values of MSE with re-
725 spect to SAFRAN references. For QQ outputs, only slight reductions of MSE
726 of spatial correlations are observed compared to those from IPSL, indicating
727 that QQ globally conserves the spatial structure of the IPSL model. This result
728 could have been expected, as, for each site, the univariate QQ method does not
729 modify (too much) rank sequences of the simulated time series. The slight im-
730 provement of spatial statistics, which is greater for precipitation (Fig. 7b) than
731 temperature (Fig. 7a), is in fact mainly attributable to the correction of uni-
732 variate properties provided by the QQ method. Concerning MBC-CycleGAN,
733 the PP and MOS approaches display different performances in adjusting the
734 spatial properties of simulations. Boxplots of MSE for MBC-CG-MOS indi-
735 cate clear improvements of spatial correlations with respect to QQ outputs
736 for both temperature and, to a lesser extent, precipitation. However, results
737 for MBC-CG-PP show less pronounced improvements, suggesting a failure
738 for the MBC-CG-PP approach to adjust spatial properties. This difference of
739 performance for the PP approach indicates that, although CycleGAN mod-
740 els are able to learn the spatial relationships between large-scale predictors
741 (LR SAFRAN) and local-scale predictands (SAFRAN) during the training
742 of the algorithm, as previously shown in Figs. 5 and S1, these relationships
743 do not prove to be suited for adjusting IPSL simulations. Indeed, simulated
744 large-scale predictors seem here to present too large biases with respect to LR
745 SAFRAN to make the CycleGAN fitted in a PP context applicable to the IPSL
746 simulations. Hence, the perfect-prognosis approach should be discarded in our
747 context of bias correction of climate simulations. Therefore, in the following,
748 only the MOS approach of MBC-CG is further investigated.

749 5.2 MBC-CycleGAN in the nonstationary context

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

756 5.2.1 Univariate distribution properties

757 Similarly to the first cross-validation method, univariate properties are evalu-
758 ated using mean differences computed at each grid cell. Fig. 8 shows, for the
759 bias-corrected outputs from the three bias correction exercises, the maps of
760 temperature mean differences with respect to SAFRAN references. Maps for
761 precipitation relative mean differences are presented in Fig. S6 of the Supple-
762 ment. For information purposes only, standard deviation relative mean differ-

ences for temperature and precipitation are also displayed in Figs. S7 and S8, respectively.

For temperature, values of IPSL and IPSLbis mean differences (Fig. 8b and 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, f and 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 (Figs. 8h, i and j), in agreement with the marginal/dependence MBC properties. For Spatial-R²D² (S-R²D²), very slight modifications of the marginal mean values provided by QQ are observed (Figs. 8k, l and 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 subsection 5.1, we compute MSE of spatial Pearson correlations for both temperature and precipitation. Fig. 9 displays the results with boxplots 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 values of MSE for IPSL are high, 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 indicate 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

807 the generation of IPSLbis leads here to reduce biases in spatial correlations.
808 Concerning LR SAFRAN, MSE values are small, suggesting that upscaling
809 the reference dataset deteriorates only slightly its spatial structure. By simply
810 correcting univariate distributions, the three QQ outputs do not present a par-
811 ticular improvement of temperature MSE values. Clear improvements of the
812 spatial correlation structures are provided by the MBC-CycleGAN method for
813 the adjustment of IPSL, IPSLbis and LR SAFRAN, although some differences
814 of performances are observed between the three corrected outputs. Tempera-
815 ture MSE values are indeed closer to 0 for the correction of LR SAFRAN than
816 for the correction of IPSLbis and IPSL, for which similar results are obtained.

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

828 For Spatial-dOTC outputs, results present low MSEs values for each bias
829 correction exercise, indicating that spatial correlations are satisfyingly cor-
830 rected by this method. Nevertheless, the adjustments are slightly better for
831 the corrected output of IPSL than for those for IPSLbis, which may be con-
832 fusing here. Indeed, as dOTC is specifically designed to take into account the
833 changes of the data to adjust in the correction procedure, better results for
834 IPSLbis, for which changes of spatial correlations are in line with those from
835 SAFRAN references, would have been expected. The great performance of
836 dOTC to correct spatial correlations for IPSL could be due to the fact that,
837 as explained in Appendix D, IPSL simulated changes for temperature are not
838 in total disagreement with those from SAFRAN, and hence there is no strong
839 discrepancy of changes affecting the corrections.

840 For precipitation (Fig. 9b), the same conclusions as those drawn for tem-
841 perature hold. Nevertheless, quite interestingly, IPSL and IPSLbis data present
842 even larger differences of MSE values. This shows the effects on spatial corre-
843 lations of the strong discrepancies of precipitation changes between the IPSL
844 model and the references observed in Appendix D: reducing this discrepancy
845 of marginal and spatial changes with IPSLbis decreases significantly the bi-
846 ases on spatial correlations. In contrast with temperature, these differences
847 of spatial correlations for precipitation between IPSL and IPSLbis are signifi-
848 cant enough to spread itself in the bias-corrected outputs: for each of the BC
849 methods, the corrected outputs for IPSLbis present systematically lower MSE
850 values compared to the corrections of IPSL.

851 To better assess spatial structure adjustments brought by MBCs, the cal-
852 culation of energy distances between the bias-corrected time series and the

853 references are performed for each physical variable according to two different
 854 multivariate distributions:

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

861 Values of energy distances are estimated using a bootstrap method. It consists
 862 for each dataset in (i) sampling (with replacement) daily fields, (ii) comput-
 863 ing the energy distance on the bootstrapped dataset, and (iii) repeating the
 864 previous two steps 1000 times to construct the bootstrap sampling distribu-
 865 tion. From this bootstrap sampling, distribution is deduced by the bootstrap
 866 estimator (mean of the 1000 energy distances obtained) and a 90% bootstrap
 867 sampling interval to provide uncertainty bands of the estimated distance. Re-
 868 sults for temperature and precipitation are displayed in Fig. 10. The closer the
 869 values of the energy distances are to 0, the closer the spatial properties of the
 870 outputs are to the one of the reference data.

871 For temperature, the two estimators of energy distances on physical values
 872 (Fig. 10a) and ranks (Fig. 10b) for IPSL and IPSLbis data are quite high com-
 873 pared to those for LR SAFRAN, which is in agreement with the differences of
 874 spatial properties already observed between these datasets and the references
 875 in Fig. 9. For the three QQ outputs, while energy distances on physical val-
 876 ues are lower (Fig. 10a), similar energy distances on ranks as those from the
 877 dataset to correct are obtained (Fig. 10b). It highlights again that, although
 878 the QQ method adjusts the univariate distributions, it is not supposed to mod-
 879 ify rank sequence of time series, and therefore spatial dependence structures,
 880 during the correction procedure. With regard to the three MBC methods for
 881 the correction of IPSL, dOTC performs slightly better on raw values (Fig. 10a)
 882 than MBC-CycleGAN and R^2D^2 , for which comparable results are obtained.
 883 For energy distances computed on ranks (Fig. 10b), dOTC and R^2D^2 produce
 884 similar results. Slightly poorer performances of MBC-CycleGAN are obtained
 885 compared to the two other MBC methods, although strongly improving the
 886 spatial dependence structures of IPSL simulations. Note that, while bootstrap
 887 sampling intervals of energy distances on temperature values are overlapping
 888 for the three MBC methods, it is less the case for energy distances on tem-
 889 perature ranks, thereby permitting to determine with more confidence the
 890 best method for the adjustment of spatial dependence properties. However, it
 891 must be mentioned that results of energy distances between the three MBCs
 892 are very close. Consequently, differences in performances between MBCs might
 893 not be significant. Concerning the correction of IPSLbis, best performances are
 894 provided by dOTC for both multivariate distributions. For multivariate dis-
 895 tributions with raw values, MBC-CycleGAN is second best, while being third
 896 for rank dependence structure. This swap of performances between raw values
 897 and ranks for MBC-CycleGAN and R^2D^2 must be analyzed with caution as

898 differences of estimated energy distances between the two MBC methods are
899 again very small and thus might not be significant. This swap can however be
900 explained by both the strong influence of marginal properties on energy dis-
901 tances and the slight deterioration of marginal properties provided by R^2D^2
902 compared to the QQ outputs, already mentioned in subsection 5.2.1. For the
903 corrections of LR SAFRAN, MBC-CycleGAN performs best and dOTC second
904 best, with a more significant difference of performance for estimated energy
905 distances evaluated on rank values (Fig. 10b).

906 For precipitation (Figs. 10c and d), conclusions similar to those obtained
907 for temperature can be drawn for IPSL, IPSLbis and LR SAFRAN outputs.
908 However, conclusions are slightly different for QQ and the MBCs. As already
909 explained in subsection 5.1, QQ modifies the frequency of dry events and conse-
910 quently changes the rank dependence structure of precipitation, which results
911 here in an improvement of spatial energy distances on ranks for the 1d-QQ
912 corrections of IPSL, IPSLbis and LR SAFRAN. Concerning the performances
913 of the three MBCs for IPSL, R^2D^2 performs best on energy distances for both
914 raw values and ranks, while MBC-CycleGAN produces reasonable results, in
915 particular for the adjustment of the rank dependence structure of precipita-
916 tion. The dOTC method produces results that are clearly unsatisfactory con-
917 cerning the rank dependence structure of precipitation. Instead of improving
918 the rank dependence structure, dOTC correction strongly degrades it. This
919 underperformance is in fact due to the presence of too many wet events in
920 the corrections provided by dOTC (not shown) compared to the references,
921 which mechanically largely affects the quality of its rank dependence structure
922 for precipitation. For the same reason, this underperformance on precipitation
923 rank dependence structure is also observed for the adjustments of IPSLbis and
924 LR SAFRAN with dOTC. For IPSLbis, estimated energy distances on ranks
925 are similar between MBC-CycleGAN and R^2D^2 . Note here that similar val-
926 ues of energy distances do not necessarily imply that their spatial dependence
927 structures are similar. Concerning LR SAFRAN corrections, MBC-CycleGAN
928 again outperforms both dOTC and R^2D^2 algorithms according to estimated
929 energy distances on raw values and ranks.

930 5.2.3 Temporal structure

931 In this section, bias-corrected data are evaluated relative to temporal proper-
932 ties. As a reminder, MBC-CycleGAN and dOTC methods have been specifi-
933 cally implemented to only adjust marginal and spatial properties of climate
934 simulations. Similarly, the R^2D^2 algorithm is applied to adjust marginal and
935 spatial features but, contrary to the two other methods, it also takes into ac-
936 count (part of) the temporal dependence properties through the multivariate
937 conditioning chosen for its implementation, as previously explained in Sec-
938 tion 4. In theory, this choice of conditioning dimensions allows R^2D^2 to par-
939 tially recover temporal properties of the reference dataset (Vrac and Thao
940 2020). Adjusting spatial coherence necessarily modifies the rank sequences of
941 the initial time series during the correction procedure (e.g., Vrac 2018). It is

942 hence interesting to quantify how strong those modifications are depending on
 943 the MBC method, whether temporal properties are taken into account in the
 944 correction procedure or not. Evaluation of temporal properties is performed by
 945 computing 1-d lag Pearson autocorrelations (AR1) at each grid cell for both
 946 temperature and precipitation. The resulting maps of differences with respect
 947 to SAFRAN references for the different BC outputs are presented in Fig. 11
 948 (resp. Fig. S10) for temperature (resp. precipitation).

949 For temperature, IPSL shows relatively low values of AR1 differences
 950 (Fig. 11b), indicating that temporal properties for temperature are relatively
 951 in line with those from the SAFRAN references (Fig. 11a). A similar differ-
 952 ences map is provided by IPSLbis outputs (Fig. 11c). In fact, IPSLbis tem-
 953 poral properties are inherited from IPSL outputs: even in a high-dimensional
 954 context, the two-step procedure - and in particular, the matrix-recorrelation
 955 technique - used to construct IPSLbis from IPSL does not lead to a strong
 956 modification of temporal properties. This result on temporal properties of
 957 data preprocessed with this matrix-recorrelation technique is consistent with
 958 the conclusions obtained in François et al (2020) for a MBC method (MRec)
 959 using the same matrix-recorrelation. For LR SAFRAN outputs (Fig. 11d), val-
 960 ues of AR1 differences are very close to 0, highlighting that the upscaling step
 961 used to construct LR SAFRAN data does not strongly modify the temporal
 962 properties of the initial SAFRAN reference dataset, which was expected by
 963 construction. Difference maps for temperature from QQ outputs (Fig. 11e, f
 964 and g) are relatively similar to those from the three datasets to adjust, respec-
 965 tively. However, for the three MBC methods used to adjust spatial dependence
 966 structure, modifications of temporal properties for temperature are not equiv-
 967 alent. With regard to MBC-CycleGAN and dOTC outputs (Fig. 11h, i, j, n,
 968 o and p), temporal statistics are close to that from the QQ outputs. It hence
 969 suggests that both MBC-CycleGAN and dOTC algorithms, although correct-
 970 ing the spatial features, perform little changes of the temporal sequencing of
 971 the time series to correct. For MBC-CycleGAN, this is partly explained by the
 972 fact that, within the CycleGAN procedure, input maps from QQ outputs are
 973 transformed to outputs with improved spatial features, whilst not modifying
 974 too much the initial input image. It hence results in partially preserving the
 975 temporal properties of the QQ outputs used as inputs of the CycleGAN while
 976 providing improvements of the spatial representation. This particular point
 977 is thereafter discussed in greater details. Concerning R^2D^2 outputs, different
 978 results are obtained depending on the dataset to correct. For the correction of
 979 both IPSL and IPSLbis (Figs. 11k and l), R^2D^2 provides small improvements
 980 of temporal properties of temperature, which illustrates that, by including lags
 981 in the conditional dimensions, R^2D^2 is able to improve - in addition to spatial
 982 properties - temporal structure of climate datasets. However, for the correction
 983 of LR SAFRAN (Fig. 11m), a deterioration of AR1 temperature differences is
 984 obtained with respect to initial LR SAFRAN data (Fig. 11d). This result can
 985 be linked with the previously mentioned contrasted performances of the R^2D^2
 986 method to adjust LR SAFRAN dataset in subsection 5.2.2.

987 For precipitation (Fig. S10), same conclusions hold for IPSL, IPSLbis and
988 LR SAFRAN outputs. However, contrary to temperature, 1d-QQ corrections
989 of IPSL and IPSLbis (Figs. S10e and f) show a pronounced improvement of
990 temporal properties for precipitation, highlighting the potential influence of
991 marginal properties of precipitation time series on its autocorrelation values.
992 Moreover, the improvements of temporal properties of temperature provided
993 by R^2D^2 for the corrections of IPSL and IPSLbis are no longer observed for
994 precipitation (Figs. S10k and l). Instead, temporal properties with unexpected
995 behaviors are obtained, potentially due to the difficulty of R^2D^2 to correct
996 physical variables with events occurring at local scale, such as precipitation
997 (Vrac and Thao 2020). It can also be due to the choice of the conditioning
998 information made in R^2D^2 . As a reminder, it is indeed the rank structure of
999 simulated precipitation (resp. temperature) that serves as a conditioning to
1000 generate Spatial- R^2D^2 outputs for precipitation (resp. temperature). As tem-
1001 poral properties (including rank sequences) of precipitation time series are not
1002 well simulated by IPSL model (Fig. S10b) compared to temperature (Fig. 11b),
1003 it potentially affects the quality of the corrections – and its temporal properties
1004 – provided by Spatial- R^2D^2 for precipitation. This highlights the importance
1005 of choosing a relevant conditioning dimension for the implementation of R^2D^2
1006 (Vrac and Thao 2020).

1007 To illustrate the fact that MBC-CycleGAN performs little changes of the
1008 temporal sequencing of the inputs to adjust, we compare corrected daily maps
1009 from LR SAFRAN with those from the references. As the LR SAFRAN dataset
1010 is temporally matching the SAFRAN dataset by construction, classic forecast
1011 statistics such as Root Mean Square Error (RMSE) can indeed be interesting
1012 to assess the performances of MBC methods. Table 2 shows, for tempera-
1013 ture and precipitation, the RMSE values with respect to SAFRAN references
1014 for the different BC outputs of LR SAFRAN. For temperature, the RMSE
1015 value between daily maps of the reference and the LR SAFRAN dataset is
1016 around 0.36. Slight improvement in terms of RMSE is provided by the QQ
1017 method (RMSE=0.31). As the QQ method preserves the temporal sequencing
1018 of the times series to correct, this improvement is only due to the correction
1019 of marginal properties. The MBC-CycleGAN method presents better results
1020 (RMSE=0.23), permitting to state with more confidence that, while adjust-
1021 ment of spatial dependence structure are performed, it modifies only slightly
1022 the temporal sequencing of the times series to correct. For R^2D^2 outputs,
1023 the RMSE value is quite large (RMSE=1.51), suggesting a strong modifica-
1024 tion of temporal properties. It can be linked with the underperformance of
1025 R^2D^2 already observed in Fig. 11m for the correction of LR SAFRAN. Con-
1026 cerning dOTC outputs, the RMSE value (=0.42) is slightly higher than those
1027 observed for LR SAFRAN and QQ outputs. It suggests that the influence of
1028 the correction of univariate distributions and spatial dependence on temporal
1029 properties provided by dOTC is strong enough to affect its ability to provide
1030 appropriate forecasts at a daily scale. For precipitation, the same conclusions
1031 hold for the different BC outputs. To better illustrate the results from Table 2,
1032 two animations presenting the successive daily temperature and precipitation

1033 maps generated by MBC-CycleGAN for the correction of LR SAFRAN, as
1034 well as the corresponding daily maps from the references and the different BC
1035 methods, are provided as supplementary materials.

1036 **6 Conclusion, discussion and future work**

1037 6.1 Conclusions

1038 Climate simulations biases are typically corrected with univariate BC methods,
1039 adjusting one physical variable and one location at a time, and thus spatial
1040 dependencies remain uncorrected. In this study, MBC-CycleGAN, an adapta-
1041 tion of the CycleGAN approach (Zhu et al 2017) used to train image-to-image
1042 translation models, was presented, allowing for the adjustment of not only
1043 univariate distributions but also spatial dependence structures of climate sim-
1044 ulations. The new suggested MBC method takes advantage of convolutional
1045 neural networks with simple architecture that are trained in competition to
1046 adjust spatial properties of simulated variables. The MBC-CycleGAN method
1047 was tested by adjusting temperature and precipitation time series from IPSL
1048 simulations with respect to the SAFRAN dataset over the region of Paris using
1049 two different cross-validation methods. The first cross-validation, that defines
1050 randomly calibration and projection periods, allows to test the new method-
1051 ology in a stationary context. We took advantage of this first cross-validation
1052 method to compare two post-processing schemes (PP and MOS) approaches
1053 that differ in the statistical relationships the MBC-CycleGAN model learns
1054 to adjust spatial dependences. The MOS approach that considers biases to
1055 refer to systematic distributional differences between references and simulated
1056 climate variables was found to be more appropriate for the implementation of
1057 the MBC-CycleGAN method and was chosen to be applied for the rest of the
1058 study. The second cross-validation method, that defines chronologically cali-
1059 bration and projection periods, was then used to evaluate the ability of the
1060 MBC-CycleGAN method to adjust climate datasets in a nonstationary con-
1061 text. As IPSL simulations and SAFRAN references present different marginal
1062 and spatial changes between calibration and projection periods, two additional
1063 climate datasets (LR SAFRAN and IPSLbis) with changes that are in line with
1064 the references were specifically constructed and adjusted, allowing to better
1065 assess the quality of the corrections provided by the new method depending
1066 on the statistical biases of the data to be corrected. A wide range of metrics
1067 has been used to evaluate bias adjustment outputs with references and ini-
1068 tial climate data and assess the corrections of univariate distributions, spatial
1069 correlations and temporal properties. In addition to the 1d-QQ method, two
1070 state-of-the-art MBC (R^2D^2 and dOTC) methods have been implemented and
1071 used as benchmarks to better evaluate the influence of nonstationary proper-
1072 ties on the results of the MBC-CycleGAN method. The results indicate that
1073 all the (M)BC methods implemented in this study generally present similar
1074 corrections of univariate distributions. Regarding spatial properties, the ben-

1075 efits of using MBC methods are clear compared to the 1d-QQ method. The
1076 MBC-CycleGAN method produced reasonable adjustments of spatial corre-
1077 lations with respect to R^2D^2 and dOTC methods for both temperature and
1078 precipitation and the three different climate datasets to adjust. Concerning the
1079 temporal aspect, the MBC-CycleGAN method is not designed to correct this
1080 specific statistical property and tends to conserve the temporal sequencing of
1081 the time series to correct. Combined with the corrections of spatial features,
1082 this property has proved to be particularly interesting for the applications of
1083 MBC-CycleGAN when the data to correct temporally match the references
1084 (e.g., as for LR SAFRAN and SAFRAN dataset, see subsection 5.2.2). The
1085 proposed method indeed outperformed all the others (M)BC alternatives for
1086 the correction of LR SAFRAN by generally presenting both spatial and tem-
1087 poral statistics closer to those from the references. Concerning nonstationary
1088 properties, it has been found that changes of both marginal and spatial prop-
1089 erties between the calibration and projection periods of the climate data to
1090 adjust can have a non-negligible effect on the quality of corrections from the
1091 MBC-CycleGAN algorithm, and more generally from all (M)BC outputs. In
1092 a general way, better results are obtained for the corrections of simulations
1093 with changes that are in agreement with those from the references, whether
1094 the MBCs make the assumption of nonstationarity of marginal properties and
1095 dependence structures or not.

1096 6.2 Discussion and perspectives

1097 In this study, the development of the MBC-CycleGAN method was mainly
1098 intended as a proof of concept, in order to test if GANs can be used for multi-
1099 variate bias correction of climate simulations. Although bringing results with
1100 comparable performances of correction to that of well-established MBC meth-
1101 ods, several avenues can be considered for the improvement of the proposed
1102 algorithm.

1103 First, in order to remain in a context of proof of concept, a simple archi-
1104 tecture of neural networks with a small number of convolutional layers has
1105 been considered for the discriminators and generators constituting the MBC-
1106 CycleGAN method. In the same idea, a classic formulation of the CycleGAN
1107 procedure – as initially described in Zhu et al (2017) – has been used with a
1108 binary-cross entropy loss function for the adversarial training (Eq. 1). Improv-
1109 ing the training performances of GANs through more advanced architectures
1110 and optimization techniques is an active area of research (e.g., Salimans et al
1111 2016; Arjovsky et al 2017; Karras et al 2018, among others). A first natural
1112 step to potentially improve results would be to opt for a more sophisticated
1113 CycleGAN model. For example, it can be done by adding more layers in the
1114 neural network architectures of both generators and discriminators to poten-
1115 tially capture more complex spatial relationships for the correction of climate
1116 simulations. Also, modifying the initial adversarial loss functions (L_{GAN} in
1117 Eq. 1), as proposed in Arjovsky et al (2017), would be interesting as it could

1118 permit to improve the stability of the learning and can prevent from mode
1119 collapse issues. However, although progress is constantly increasing concern-
1120 ing GANs, it is well-known that this particular class of neural networks can be
1121 more difficult to train than classical neural networks (e.g., Wu et al 2020). The
1122 possibilities of modifications of the parameters defining a CycleGAN model are
1123 numerous, and a priori do not guarantee to improve the overall performance
1124 of the CycleGAN for the specific application of bias correction. Testing the
1125 different possibilities goes way beyond the scope of the present study and is
1126 left for future work.

1127 Second, it has to be noted that our method, by combining the 1d-QQ
1128 method and the CycleGAN approach to adjust both marginal and spatial
1129 properties, is not designed to specifically account for any simulated changes
1130 for future periods. For marginal properties, other 1d-BC methods that are able
1131 to account for potential changes of univariate CDFs from the calibration to the
1132 projection period (e.g., CDF-t or QDM, Vrac et al 2012; Cannon et al 2015)
1133 can of course be employed instead of QQ, as long as they do not modify (too
1134 much) rank sequence of temperature and precipitation time series and thus do
1135 not distort the convergence of the CycleGAN procedure. Concerning changes
1136 of spatial properties, the CycleGAN approach as implemented in this study
1137 is based on the key assumption that the conditional distributions $\mathbf{X}|\mathbf{Y}$ and
1138 $\mathbf{Y}|\mathbf{X}$ are the same in the training (i.e., calibration) and test (i.e., projection)
1139 datasets. It results in our context in making a strong assumption on copula
1140 stationarity between present and future periods. Although spatial dependence
1141 structures can be considered to be stable in time as imposed by physical laws
1142 over a specific region of interest (e.g., Vrac 2018), it can not be generalized
1143 to each of the physical variables and regions. For example, more concentrated
1144 spatial rainfall events are expected with higher temperatures in the future
1145 (Guinard et al 2015; Wasko et al 2016). Therefore, should the changes in spa-
1146 tial properties in the simulations between calibration and projection periods be
1147 reproduced in the correction? By comparing our results obtained with different
1148 levels of nonstationarity in the model evolution and with two well-established
1149 MBCs based on copula stationarity (R^2D^2) and nonstationarity (dOTC) for
1150 future periods, we shed light on how the nonstationary properties of the simu-
1151 lations are taken into account by the different multivariate BC methods. The
1152 benefits of considering MBC methods assuming copula nonstationarity for the
1153 correction of such climate dataset are not always as clear-cut as expected com-
1154 pared to MBC methods assuming copula stationarity. This raises the question
1155 of whether developing MBC methods assuming copula nonstationarity is jus-
1156 tified, i.e., whether it is worth striving for developing complicated statistical
1157 methods that consider the simulated evolution of copula in the correction pro-
1158 cedure, and, in the end, do not produce drastically better results than MBCs
1159 assuming copula stationarity. In practice, accounting for nonstationarity of
1160 simulations in bias correction procedures still remains an open question which
1161 needs to be answered on a case-by-case basis. Developing new MBC methods
1162 that are specifically able to reproduce these simulated changes in the correc-
1163 tion 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.

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

1210 outputs is not estimated here. An interesting possibility of extension to model
1211 uncertainty of corrected outputs would be to introduce some stochasticity into
1212 the correction procedure by giving to the generators not only daily maps to
1213 adjust but also vectors of random noises. Then, for a given daily map, it would
1214 produce an ensemble of plausible corrections. The spread between the ensemble
1215 members would represent the uncertainty associated with the multivariate
1216 bias correction.

1217 We hope that this study serves as a starting point for the use of GANs for
1218 multivariate bias correction of climate simulations. One of the main advantages
1219 of using MBC-CycleGAN is that adjustment is performed images by images,
1220 i.e. maps by maps. If well trained, discriminators somehow guarantee that
1221 individual generated maps produced by generators are realistic with respect to
1222 references, while daily maps with strong statistical artefacts are rejected. This
1223 is not the case for the other MBC methods such as R²D² or dOTC, that provide
1224 corrected simulations with appropriate distributional statistics without being
1225 particularly constrained to generate realistic daily maps. Providing corrections
1226 with realistic maps at a daily scale can be useful for the scientific community
1227 working on climate change impacts, e.g., in hydrology, for which daily spatial
1228 features are of major concern.

1229 **Acknowledgements** This work was granted access to the HPC resources of IDRIS under
1230 the allocation 20XX-[AD011011646] made by GENCI. MV acknowledges support from the
1231 CoCliServ project, which is part of ERA4CS, an ERA-NET initiated by JPI Climate and
1232 cofunded by the European Union.

1233 **Declaration**

1234 **Funding**

1235 This research has been supported by the CoCliServ project, which is part
1236 of ERA4CS, an ERA-NET initiated by JPI Climate and cofunded by the
1237 European Union.

1238 **Conflicts of interest/Competing interests**

1239 The authors have no conflicts of interest to declare that are relevant to the
1240 content of this article.

1241 **Availability of data and material**

1242 The IPSL-CM5A-MR model data simulations as part of the CMIP5 cli-
1243 mate model simulations can be downloaded through the Earth System
1244 Grid Federation portals. Instructions to access the data are available here:

1245 <https://pcmdi.llnl.gov/mips/cmip5/data-access-getting-started.html>, last ac-
1246 cess: 06 September 2020, (PCMDI, 1989). The SAFRAN reanalysis dataset
1247 is available upon request to the French National Centre for Meteorological
1248 Research (CNRM, Météo-France CNRS).

1249 Code availability

1250 The code for MBC-CycleGAN is publicly available at
1251 https://github.com/bastien-francois/MBC_CycleGAN. The R package
1252 for R²D² is available at <https://github.com/thaos/R2D2> (Vrac and Thao
1253 2020). dOTC is publicly available at <https://github.com/yrobink/SBCK>
1254 (Robin et al 2019).

1255 Authors' contributions

1256 MV had the initial idea of the study and its structure, which was enriched
1257 by all coauthors. BF made all computations and figures, with help from ST.
1258 BF wrote the first draft of the article, with inputs, corrections and additional
1259 writing contributions from MV and ST.

1260 References

- 1261 Arjovsky M, Chintala S, Bottou L (2017) Wasserstein GAN. 1701.07875
1262 Baño-Medina J, Manzanar R, Gutiérrez JM (2020) Configuration and intercomparison of
1263 deep learning neural models for statistical downscaling. *Geosci Model Dev* 13(4):2109–
1264 2124, DOI 10.5194/gmd-13-2109-2020
1265 Bárdossy A, Pegram G (2012) Multiscale spatial recorelation of RCM precipitation to
1266 produce unbiased climate change scenarios over large areas and small. *Water Resour*
1267 *Res* 48:9502–, DOI 10.1029/2011WR011524
1268 Bartok B, Tobin I, Vautard R, Vrac M, Jin X, Levvasseur G, Denvil S, Dubus L,
1269 Parey S, Michelangeli PA, Troccoli A, Saint-Drenan YM (2019) A climate projec-
1270 tion dataset tailored for the European energy sector. *Clim Serv* 16:100,138, DOI
1271 10.1016/j.cliser.2019.100138
1272 Bates B, Kundzewicz Z, Wu S, Burkett V, Doell P, Gwary D, Hanson C, Heij B, Jiménez
1273 B, Kaser G, Kitoh A, Kovats S, Kumar P, Magadza C, Martino D, Mata L, Medany
1274 M, Miller K, Arnell N (2008) Climate Change and Water. Technical Paper of the In-
1275 tergovernmental Panel on Climate Change. Tech. rep., The Intergovernmental Panel on
1276 Climate Change
1277 Beltrami E (1873) Sulle funzioni bilineari. *Giornale di Matematiche ad Uso degli Studenti*
1278 *Delle Università* 11:98–106
1279 Berg P, Feldmann H, Panitz HJ (2012) Bias correction of high resolution regional climate
1280 model data. *J Hydrol* 448–449:80–92, DOI 10.1016/j.jhydrol.2012.04.026
1281 Bhatia S, Jain A, Hooi B (2020) ExGAN: Adversarial Generation of Extreme Samples.
1282 2009.08454
1283 Bihlo A (2020) A generative adversarial network approach to (ensemble) weather prediction.
1284 2006.07718
1285 Caminade C, Kovats S, Rocklov J, Tompkins AM, Morse AP, Colón-González FJ, Stenlund
1286 H, Martens P, Lloyd SJ (2014) Impact of climate change on global malaria distribution.
1287 *Proc Natl Acad Sci USA* 111(9):3286–3291, DOI 10.1073/pnas.1302089111

- 1288 Cannon A, Sobie S, Murdock T (2015) Bias Correction of GCM Precipitation by Quan-
1289 tile Mapping: How Well Do Methods Preserve Changes in Quantiles and Extremes? *J*
1290 *Climate* 28(17):6938–6959, DOI 10.1175/JCLI-D-14-00754.1
- 1291 Cannon AJ (2018) Multivariate quantile mapping bias correction: an N-dimensional prob-
1292 ability density function transform for climate model simulations of multiple variables.
1293 *Clim Dynam* 50(1):31–49, DOI 10.1007/s00382-017-3580-6
- 1294 Cattiaux J, Douville H, Peings Y (2013) European temperatures in CMIP5: Origins
1295 of present-day biases and future uncertainties. *Clim Dynam* 41:2889–2907, DOI
1296 10.1007/s00382-013-1731-y
- 1297 Chapman WE, Subramanian AC, Delle Monache L, Xie SP, Ralph FM (2019) Improving At-
1298 mospheric River Forecasts With Machine Learning. *Geophys Res Lett* 46(17-18):10,627–
1299 10,635, DOI 10.1029/2019GL083662
- 1300 Christensen JH, Boberg F, Christensen OB, Lucas-Picher P (2008) On the need for bias cor-
1301 rection of regional climate change projections of temperature and precipitation. *Geophys*
1302 *Res Lett* 35(20):L20,709, DOI 10.1029/2008GL035694
- 1303 Clark M, Gangopadhyay S, Hay L, Rajagopalan B, Wilby R (2004) The Schaake Shuffle:
1304 A method for Reconstructing Space–Time Variability in Forecasted Precipitation and
1305 Temperature Fields. *J Hydrometeor* 5(1):243–262
- 1306 Defrance D, Ramstein G, Charbit S, Vrac M, Famien AM, Sultan B, Swingedouw D, Dumas
1307 C, Gemenne F, Alvarez-Solas J, Vanderlinden JP (2017) Consequences of rapid ice sheet
1308 melting on the Sahelian population vulnerability. *Proc Natl Acad Sci USA* 114(25):6533–
1309 6538, DOI 10.1073/pnas.1619358114
- 1310 Dekens L, Parey S, Grandjacques M, Dacunha-Castelle D (2017) Multivariate distribution
1311 correction of climate model outputs: A generalization of quantile mapping approaches:
1312 Multivariate distribution correction of climate model outputs. *Environmetrics* 28:e2454,
1313 DOI 10.1002/env.2454
- 1314 Denton E, Chintala S, Szlam A, Fergus R (2015) Deep Generative Image Models using a
1315 Laplacian Pyramid of Adversarial Networks. 1506.05751
- 1316 Déqué M (2007) Frequency of precipitation and temperature extremes over France in an
1317 anthropogenic scenario: Model results and statistical correction according to observed
1318 values. *Global Planet Change* 57(1):16–26, DOI 10.1016/j.gloplacha.2006.11.030
- 1319 Eden J, Widmann M, Grawe D, Rast S (2012) Skill, Correction, and Downscaling of GCM-
1320 Simulated Precipitation. *J Climate* 25:3970–3984, DOI 10.1175/JCLI-D-11-00254.1
- 1321 Fisher RA (1915) Frequency Distribution of the Values of the Correlation Coefficient in
1322 Samples from an Indefinitely Large Population. *Biometrika* 10(4):507–521
- 1323 François B, Vrac M, Cannon AJ, Robin Y, Allard D (2020) Multivariate bias corrections
1324 of climate simulations: Which benefits for which losses? *Earth Syst Dyn* 2020:1–41,
1325 DOI 10.5194/esd-2020-10
- 1326 Gagne II DJ, Christensen HM, Subramanian AC, Monahan AH (2020) Machine Learning for
1327 Stochastic Parameterization: Generative Adversarial Networks in the Lorenz '96 Model.
1328 *J Adv Model Earth Syst* 12(3):e2019MS001,896, DOI 10.1029/2019MS001896
- 1329 Gan Z, Chen L, Wang W, Pu Y, Zhang Y, Liu H, Li C, Carin L (2017) Triangle Generative
1330 Adversarial Networks. 1709.06548
- 1331 Gauthier J (2014) Conditional generative adversarial nets for convolutional face generation.
1332 Class Project for Stanford CS231N: Convolutional Neural Networks for Visual Recog-
1333 nition, Winter semester 2014(5):2
- 1334 Gokaslan A, Ramanujan V, Ritchie D, Kim KI, Tompkin J (2019) Improving Shape Deform-
1335 ation in Unsupervised Image-to-Image Translation. 1808.04325
- 1336 Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Ben-
1337 gio Y (2014) Generative Adversarial Nets. *Advances in Neural Information Processing*
1338 *Systems* 3, DOI 10.1145/3422622
- 1339 Gudmundsson L, Bremnes JB, Haugen JE, Engen-Skaugen T (2012) Technical Note: Down-
1340 scaling RCM precipitation to the station scale using statistical transformations – a
1341 comparison of methods. *Hydrol Earth Syst Sci* 16(9):3383–3390, DOI 10.5194/hess-16-
1342 3383-2012
- 1343 Guinard K, Mailhot A, Caya D (2015) Projected changes in characteristics of precipitation
1344 spatial structures over North America. *Int J Climatol* 35:596–612, DOI 10.1002/joc.4006

- 1345 Guo Q, Chen J, Zhang X, Shen M, Chen H, Guo S (2019) A new two-stage multivari-
1346 ate quantile mapping method for bias correcting climate model outputs. *Clim Dynam*
1347 53(5):3603–3623, DOI 10.1007/s00382-019-04729-w
- 1348 Gutmann E, Pruitt T, Clark M, Brekke L, Arnold J, Raff D, Rasmussen R (2014) An
1349 intercomparison of statistical downscaling methods used for water resource assessments
1350 in the United States. *Water Resour Res* 50:7167–7186, DOI 10.1002/2014WR015559
- 1351 Haddad Z, Rosenfeld D (1997) Optimality of empirical Z-R relations. *Q J Roy Meteor Soc*
1352 123(541):1283–1293, DOI 10.1002/qj.49712354107
- 1353 He K, Zhang X, Ren S, Sun J (2016) Deep Residual Learning for Image Recognition. In: 2016
1354 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp 770–778,
1355 DOI 10.1109/CVPR.2016.90
- 1356 Hnilica J, Hanel M, Puš V (2017) Multisite bias correction of precipitation data from regional
1357 climate models. *Int J Climatol* 37:2934–2946, DOI 10.1002/joc.4890
- 1358 IPCC (2014) Climate Change 2014: Synthesis Report. Contribution of Working Groups I,
1359 II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate
1360 Change [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]. IPCC, Geneva,
1361 Switzerland, 151 pp
- 1362 Isola P, Zhu JY, Zhou T, Efros AA (2017) Image-to-Image Translation with Conditional Ad-
1363 versarial Networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recog-
1364 nition (CVPR), pp 5967–5976, DOI 10.1109/CVPR.2017.632
- 1365 Jordan C (1874a) Mémoire sur les formes bilinéaires. *J Math Pures Appl* 19(Deuxième
1366 Série):35–54
- 1367 Jordan C (1874b) Sur la réduction des formes bilinéaires. *C R Acad Sci, Paris (Deuxième
1368 Série):614–617*
- 1369 Karras T, Aila T, Laine S, Lehtinen J (2018) Progressive Growing of GANs for Improved
1370 Quality, Stability, and Variation. 1710.10196
- 1371 Kim T, Cha M, Kim H, Lee JK, Kim J (2017) Learning to Discover Cross-Domain Relations
1372 with Generative Adversarial Networks. 1703.05192
- 1373 Kingma DP, Ba J (2017) Adam: A Method for Stochastic Optimization. 1412.6980
- 1374 Lecun Y, Bengio Y (1995) Convolutional Networks for Images, Speech, and Time-Series.
1375 *The Handbook of Brain Theory and Neural Networks* 3361
- 1376 Leinonen J, Berne A (2020) Unsupervised classification of snowflake images using a gener-
1377 ative adversarial network and K -medoids classification. *Atmos Meas Tech* 13(6):2949–
1378 2964, DOI 10.5194/amt-13-2949-2020
- 1379 Leinonen J, Nerini D, Berne A (2020) Stochastic Super-Resolution for Downscaling Time-
1380 Evolving Atmospheric Fields with a Generative Adversarial Network. *IEEE Trans Geosci*
1381 *Remote Sens* pp 1–13, DOI 10.1109/TGRS.2020.3032790
- 1382 Liu Y, Racah E, Prabhat, Correa J, Khosrowshahi A, Lavers D, Kunkel K, Wehner M,
1383 Collins W (2016) Application of Deep Convolutional Neural Networks for Detecting
1384 Extreme Weather in Climate Datasets. 1605.01156
- 1385 Mao X, Li Q, Xie H, Lau RYK, Wang Z, Smolley SP (2017) Least Squares Generative
1386 Adversarial Networks. 1611.04076
- 1387 Maraun D (2013) Bias Correction, Quantile Mapping, and Downscaling: Revisiting the In-
1388 flation Issue. *J Climate* 26(6):2137–2143, DOI 10.1175/JCLI-D-12-00821.1
- 1389 Maraun D (2016) Bias Correcting Climate Change Simulations - A Critical Review. *Curr*
1390 *Clim Chang Reports* 2:211–220, DOI 10.1007/s40641-016-0050-x
- 1391 Maraun D, Wetterhall F, Ireson AM, Chandler RE, Kendon EJ, Widmann M, Brienen S,
1392 Rust HW, Sauter T, Themeßl M, Venema VKC, Chun KP, Goodess CM, Jones RG,
1393 Onof C, Vrac M, Thiele-Eich I (2010) Precipitation downscaling under climate change:
1394 Recent developments to bridge the gap between dynamical models and the end user.
1395 *Rev Geophys* 48(3), DOI 10.1029/2009RG000314
- 1396 Mehrotra R, Sharma A (2016) A Multivariate Quantile-Matching Bias Correction Approach
1397 with Auto- and Cross-Dependence across Multiple Time Scales: Implications for Down-
1398 scaling. *J Climate* 29(10):3519–3539, DOI 10.1175/JCLI-D-15-0356.1
- 1399 Mehrotra R, Sharma A (2019) A Resampling Approach for Correcting Systematic Spa-
1400 tiotemporal Biases for Multiple Variables in a Changing Climate. *Water Resour Res*
1401 55(1):754–770, DOI 10.1029/2018WR023270

- 1402 Menick J, Kalchbrenner N (2018) Generating High Fidelity Images with Subscale Pixel
1403 Networks and Multidimensional Upscaling. 1812.01608
- 1404 Mirza M, Osindero S (2014) Conditional Generative Adversarial Nets. 1411.1784
- 1405 Mueller B, Seneviratne S (2014) Systematic land climate and evapotranspiration biases in
1406 CMIP5 simulations. *Geophys Res Lett* 41:128–134, DOI 10.1002/2013GL058055
- 1407 Muerth MJ, Gauvin St-Denis B, Ricard S, Velázquez JA, Schmid J, Minville M, Caya D,
1408 Chaumont D, Ludwig R, Turcotte R (2013) On the need for bias correction in regional
1409 climate scenarios to assess climate change impacts on river runoff. *Hydrol Earth Syst
1410 Sci* 17(3):1189–1204, DOI 10.5194/hess-17-1189-2013
- 1411 Nahar J, Johnson F, Sharma A (2018) Addressing Spatial Dependence Bias in Climate
1412 Model Simulations—An Independent Component Analysis Approach. *Water Resour Res*
1413 54(2):827–841, DOI 10.1002/2017WR021293
- 1414 Nguyen H, Mehrotra R, Sharma A (2019) Correcting systematic biases across multiple
1415 atmospheric variables in the frequency domain. *Clim Dynam* 52:1283–1298, DOI
1416 10.1007/s00382-018-4191-6
- 1417 Piani C, Haerter J (2012) Two dimensional bias correction of temperature and precipitation
1418 copulas in climate models. *Geophys Res Lett* 39:L20,401, DOI 10.1029/2012GL053839
- 1419 Racah E, Beckham C, Maharaj T, Kahou SE, Prabhat, Pal C (2017) ExtremeWeather: A
1420 large-scale climate dataset for semi-supervised detection, localization, and understand-
1421 ing of extreme weather events. 1612.02095
- 1422 Radford A, Metz L, Chintala S (2016) Unsupervised Representation Learning with Deep
1423 Convolutional Generative Adversarial Networks. 1511.06434
- 1424 Ramirez-Villegas J, Challinor A, Thornton P, Jarvis A (2013) Implications of regional im-
1425 provement in global climate models for agricultural impact research. *Environ Res Lett*
1426 8:024,018, DOI 10.1088/1748-9326/8/2/024018
- 1427 Randall D, Wood R, Bony S, Colman R, Fichefet T, Fyfe J, Kattsov V, Pitman A, Shukla J,
1428 Srinivasan J, Ronald S, Sumi A, Taylor K (2007) *Climate Models and Their Evaluation.*,
1429 Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp
1430 589–662
- 1431 Reichler T, Kim J (2008) How Well Do Coupled Models Simulate Today’s Climate? *Bull
1432 Amer Meteor Soc* 89:303–311, DOI 10.1175/BAMS-89-3-303
- 1433 Reichstein M, Camps-Valls G, Stevens B, Jung M, Denzler J, Carvalhais N, Prabhat M
1434 (2019) Deep learning and process understanding for data-driven Earth system science.
1435 *Nature* 566:195–204, DOI 10.1038/s41586-019-0912-1
- 1436 Robin Y, Vrac M, Naveau P, Yiou P (2019) Multivariate stochastic bias corrections with
1437 optimal transport. *Hydrol Earth Syst Sci* 23(2):773–786, DOI 10.5194/hess-23-773-2019
- 1438 Rodrigues ER, Oliveira I, Cunha RLF, Netto MAS (2018) DeepDownscale: a Deep Learning
1439 Strategy for High-Resolution Weather Forecast. In: 2018 IEEE 14th International
1440 Conference on e-Science (e-Science), pp 415–422, DOI 10.1109/eScience.2018.00130
- 1441 Roth K, Lucchi A, Nowozin S, Hofmann T (2017) Stabilizing Training of Generative Adver-
1442 sarial Networks through Regularization. 1705.09367
- 1443 Royer A, Bousmalis K, Gouws S, Bertsch F, Mosseri I, Cole F, Murphy K (2020) XGAN:
1444 Unsupervised Image-to-Image Translation for Many-to-Many Mappings, Springer Inter-
1445 national Publishing, pp 33–49. DOI 10.1007/978-3-030-30671-7_3
- 1446 Rätty O, Räisänen J, Bosshard T, Donnelly C (2018) Intercomparison of Univariate and
1447 Joint Bias Correction Methods in Changing Climate From a Hydrological Perspective.
1448 *Climate* 6:33, DOI 10.3390/cli6020033
- 1449 Salimans T, Goodfellow I, Zaremba W, Cheung V, Radford A, Chen X (2016) Improved
1450 Techniques for Training GANs. 1606.03498
- 1451 Scher S, Messori G (2018) Predicting weather forecast uncertainty with machine learning.
1452 *Q J R Meteorol Soc* 144(717):2830–2841, DOI 10.1002/qj.3410
- 1453 Scher S, Messori G (2019) Weather and climate forecasting with neural networks: using
1454 general circulation models (GCMs) with different complexity as a study ground. *Geosci
1455 Model Dev* 12(7):2797–2809, DOI 10.5194/gmd-12-2797-2019
- 1456 Scher S, Peßenteiner S (2020) Technical Note: Temporal Disaggregation of Spatial Rainfall
1457 Fields with Generative Adversarial Networks. *Hydrol Earth Syst Sci* 2020:1–23, DOI
1458 10.5194/hess-2020-464

- 1459 Schmidhuber J (2015) Deep learning in neural networks: An overview. *Neural Netw* 61:85 –
1460 117, DOI <https://doi.org/10.1016/j.neunet.2014.09.003>
- 1461 Shi X, Chen Z, Wang H, Yeung DY, Wong W, Woo W (2015) Convolutional LSTM Network:
1462 A Machine Learning Approach for Precipitation Nowcasting. 1506.04214
- 1463 Stewart GW (1993) On the Early History of the Singular Value Decomposition. *SIAM Rev*
1464 35(4):551–566, DOI 10.1137/1035134
- 1465 Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke
1466 V, Rabinovich A (2015) Going Deeper with Convolutions. In: 2015 IEEE Con-
1467 ference on Computer Vision and Pattern Recognition (CVPR), pp 1–9, DOI
1468 10.1109/CVPR.2015.7298594
- 1469 Székely G, Rizzo M (2004) Testing for equal distributions in high dimension. *InterStat* 5
- 1470 Székely G, Rizzo M (2013) Energy statistics: A class of statistics based on distances. *J Stat*
1471 *Plan Inference* 143:1249–1272, DOI 10.1016/j.jspi.2013.03.018
- 1472 Teutschbein C, Seibert J (2012) Bias correction of regional climate model simulations for
1473 hydrological climate-change impact studies: Review and evaluation of different methods.
1474 *J Hydrol* 456:12–29, DOI 10.1016/j.jhydrol.2012.05.052
- 1475 Tong Y, Gao X, Han Z, Xu Y, Xu Y, Giorgi F (2020) Bias correction of temperature and
1476 precipitation over China for RCM simulations using the QM and QDM methods. *Clim*
1477 *Dynam* pp 1432–0894, DOI 10.1007/s00382-020-05447-4
- 1478 Trambly Y, Ruelland D, Somot S, Bouaicha R, Servat E (2013) High-resolution Med-
1479 CORDEX regional climate model simulations for hydrological impact studies: a
1480 first evaluation of the ALADIN-Climate model in Morocco. *Hydrol Earth Syst Sci*
1481 17(10):3721–3739, DOI 10.5194/hess-17-3721-2013
- 1482 Van Loon A, Gleeson T, Clark J, van Dijk A, Stahl K, Hannaford J, Di Baldassarre G,
1483 Teuling A, Tallaksen L, Uijlenhoet R, Hannah D, Sheffield J, Svoboda M, Verbeiren
1484 B, Wagener T, Rangecroft S, Wanders N, Van Lanen H (2016) Drought in the Anthro-
1485 pocene. *Nat Geosci* 9:89–91, DOI 10.1038/ngeo2646
- 1486 Vandal T, Kodra E, Ganguly S, Michaelis A, Nemani R, Ganguly AR (2017) DeepSD:
1487 Generating High Resolution Climate Change Projections through Single Image Super-
1488 Resolution. In: Proceedings of the 23rd ACM SIGKDD International Conference on
1489 Knowledge Discovery and Data Mining, pp 1663–1672, DOI 10.1145/3097983.3098004
- 1490 Vidal JP, Martin E, Franchistéguy L, Baillon M, Soubeyroux JM (2010) A 50-year high-
1491 resolution atmospheric reanalysis over France with the Safran system. *Int J Climatol*
1492 30(11):1627–1644, DOI 10.1002/joc.2003
- 1493 Vignaud N, Vrac M, Caballero Y (2013) Probabilistic downscaling of GCM scenarios over
1494 southern India. *Int J Climatol* 33:1248–1263, DOI 10.1002/joc.3509
- 1495 Vorogushyn S, Bates PD, de Bruijn K, Castellarin A, Kreibich H, Priest S, Schröter K,
1496 Bagli S, Blöschl G, Domeneghetti A, Gouldby B, Klijn F, Lammersen R, Neal JC,
1497 Ridder N, Terink W, Viavattene C, Viglione A, Zanardo S, Merz B (2018) Evolution-
1498 ary leap in large-scale flood risk assessment needed. *WIREs Water* 5(2):e1266, DOI
1499 10.1002/wat2.1266
- 1500 Vrac M (2018) Multivariate bias adjustment of high-dimensional climate simulations: the
1501 Rank Resampling for Distributions and Dependences (R^2D^2) bias correction. *Hydrol*
1502 *Earth Syst Sci* 22(6):3175–3196, DOI 10.5194/hess-22-3175-2018
- 1503 Vrac M, Thao S (2020) R^2D^2 v2.0: Accounting for temporal dependences in multivari-
1504 ate bias correction via analogue ranks resampling. *Geosci Model Dev* 2020:1–29, DOI
1505 10.5194/gmd-2020-132
- 1506 Vrac M, Drobinski P, Merlo A, Herrmann M, Lavaysse C, Li L, Somot S (2012) Dynamical
1507 and statistical downscaling of the French Mediterranean climate: uncertainty assess-
1508 ment. *Nat Hazards Earth Syst Sci* 12(9):2769–2784, DOI 10.5194/nhess-12-2769-2012
- 1509 Vrac M, Noël T, Vautard R (2016) Bias correction of precipitation through Singularity
1510 Stochastic Removal: Because Occurrences matter. *J Geophys Res Atmos* 121:5237–5258,
1511 DOI 10.1002/2015JD024511
- 1512 Wang J, Liu Z, Foster I, Chang W, Kettimuthu R, Kotamarthi R (2021) Fast and accurate
1513 learned multiresolution dynamical downscaling for precipitation. 2101.06813
- 1514 Wasko C, Sharma A, Westra S (2016) Reduced spatial extent of extreme storms at higher
1515 temperatures. *Geophys Res Lett* 43(8):4026–4032, DOI 10.1002/2016GL068509

- 1516 Wheeler T, von Braun J (2013) Climate Change Impacts on Global Food Security. *Science*
1517 341(6145):508–513, DOI 10.1126/science.1239402
- 1518 Wilcke RAI, Mendlik T, Gobiet A (2013) Multi-variable error correction of regional climate
1519 models. *Clim Change* 120:871–887, DOI 10.1007/s10584-013-0845-x
- 1520 Wilks DS (2006) *Statistical Methods in the Atmosphere Science*. Academic Press
- 1521 Wu JL, Kashinath K, Albert A, Chirila D, Prabhat, Xiao H (2020) Enforcing statistical
1522 constraints in generative adversarial networks for modeling chaotic dynamical systems.
1523 *J Comput Phys* 406:109,209, DOI 10.1016/j.jcp.2019.109209
- 1524 Xie Y, Franz E, Chu M, Thuerey N (2018) TempoGAN: A Temporally Coherent, Vol-
1525 umetric GAN for Super-resolution Fluid Flow. *ACM Trans Graph* 37(4), DOI
1526 10.1145/3197517.3201304
- 1527 Xu CY (1999) From GCMs to river flow: A review of downscaling methods and hydrologic
1528 modelling approaches. *Prog Phys Geog* 23:229–249, DOI 10.1177/030913339902300204
- 1529 Yi Z, Zhang H, Tan P, Gong M (2017) DualGAN: Unsupervised Dual Learning for Image-to-
1530 Image Translation. In: 2017 IEEE International Conference on Computer Vision (ICCV),
1531 pp 2868–2876, DOI 10.1109/ICCV.2017.310
- 1532 Yoo D, Kim N, Park S, Paek AS, Kweon IS (2016) Pixel-Level Domain Transfer. 1603.07442
- 1533 Zhu JY, Park T, Isola P, Efros AA (2017) Unpaired Image-to-Image Translation using Cycle-
1534 Consistent Adversarial Networks. In: Proc. IEEE Int. Conf. Comput. Vision (ICCV),
1535 pp 2223–2232
- 1536 Zscheischler J, Westra S, Hurk B, Seneviratne S, Ward P, Pitman A, AghaKouchak A,
1537 Bresch D, Leonard M, Wahl T, Zhang X (2018) Future climate risk from compound
1538 events. *Nat Clim Chang* pp 469–477, DOI 10.1038/s41558-018-0156-3
- 1539 Zscheischler J, Fischer E, Lange S (2019) The effect of univariate bias adjustment on mul-
1540 tivariate hazard estimates. *Earth Syst Dynam* 10:31–43, DOI 10.5194/esd-10-31-2019

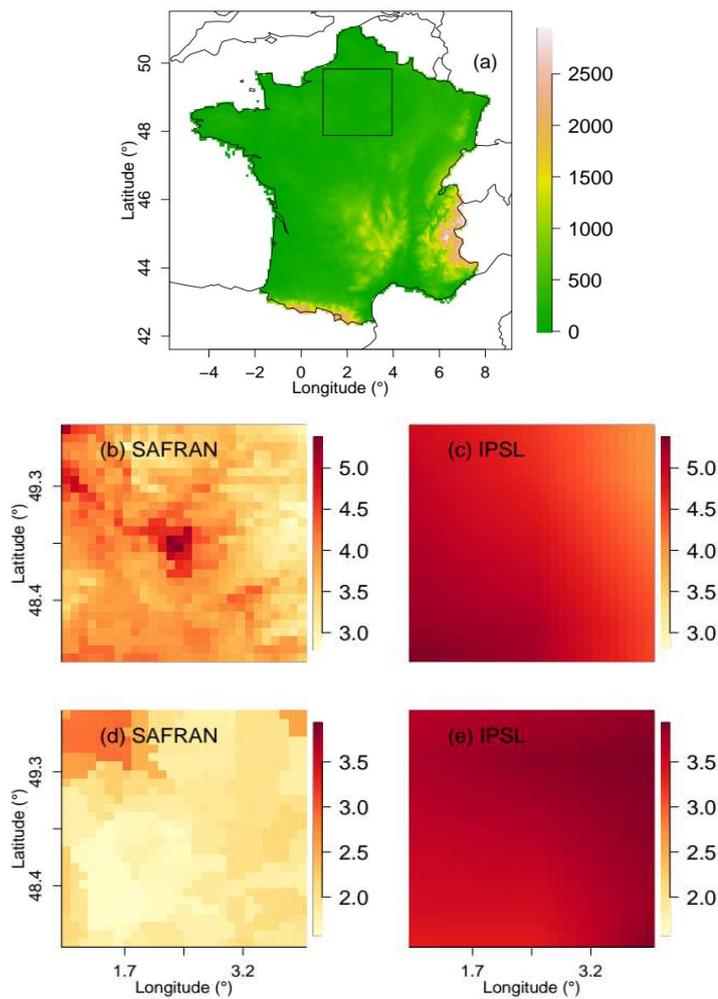


Fig. 1: (a) Topographic map of France with the selected region over Paris in a box, (b-c) temperature and (d-e) precipitation daily mean computed at each grid cell during winter over the 1979-2016 period for Paris. Results are shown for SAFRAN reference and plain IPSL outputs.

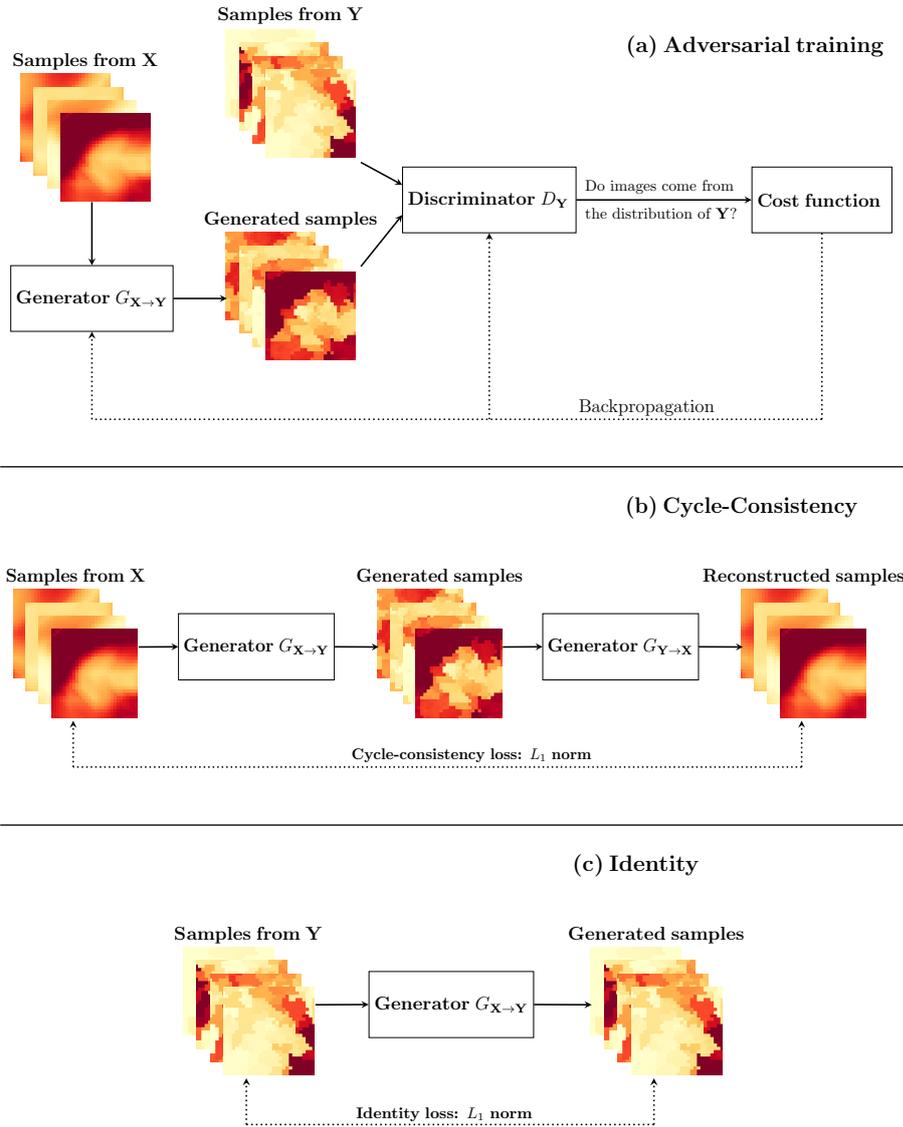


Fig. 2: (a) Illustration of the adversarial training for the mapping function $G_{\mathbf{X} \rightarrow \mathbf{Y}}$, associated with the adversarial discriminator $D_{\mathbf{Y}}$. $D_{\mathbf{Y}}$ encourages $G_{\mathbf{X} \rightarrow \mathbf{Y}}$ to generate outputs that are indistinguishable from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

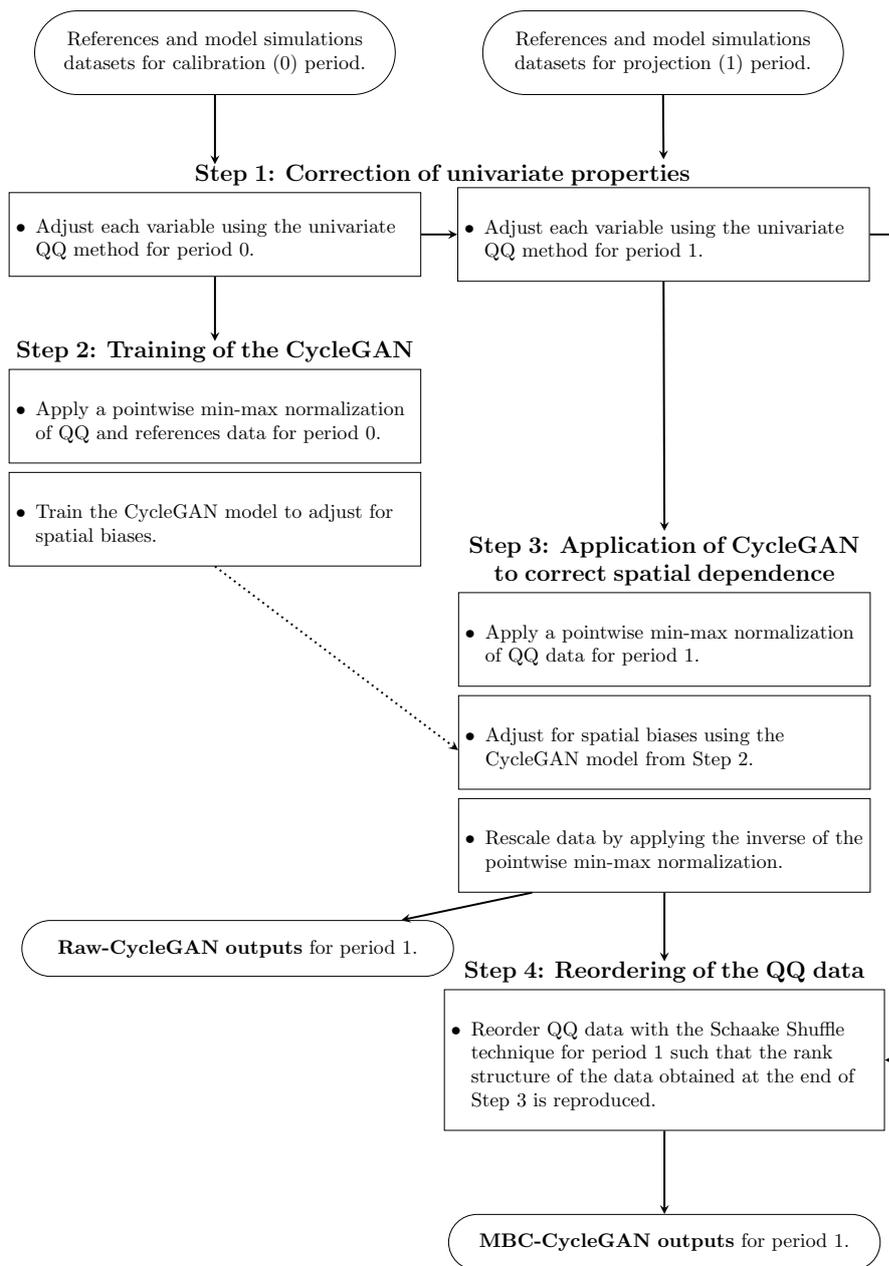


Fig. 3: Flowchart for the MBC-CycleGAN method to adjust climate simulations for the projection period.

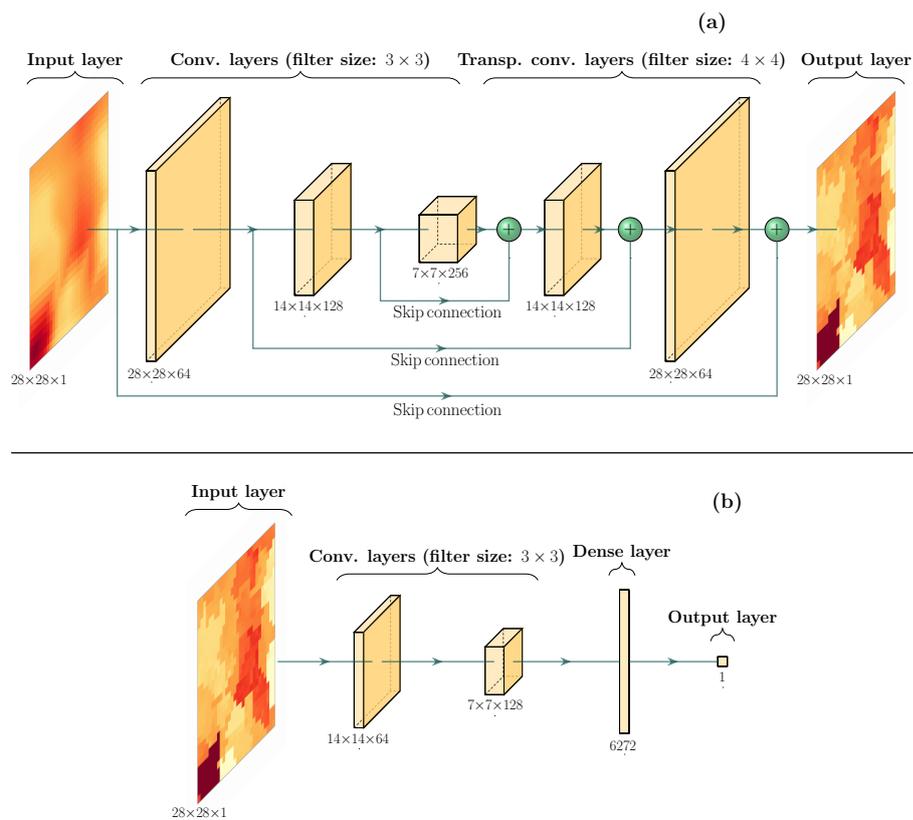


Fig. 4: Scheme of the convolutional neural networks for the **(a)** generators and **(b)** discriminators used in this study within the MBC-CycleGAN procedure. For each convolutional and transpose convolutional layers, the number of filters used is indicated by the third coordinate of their output size.

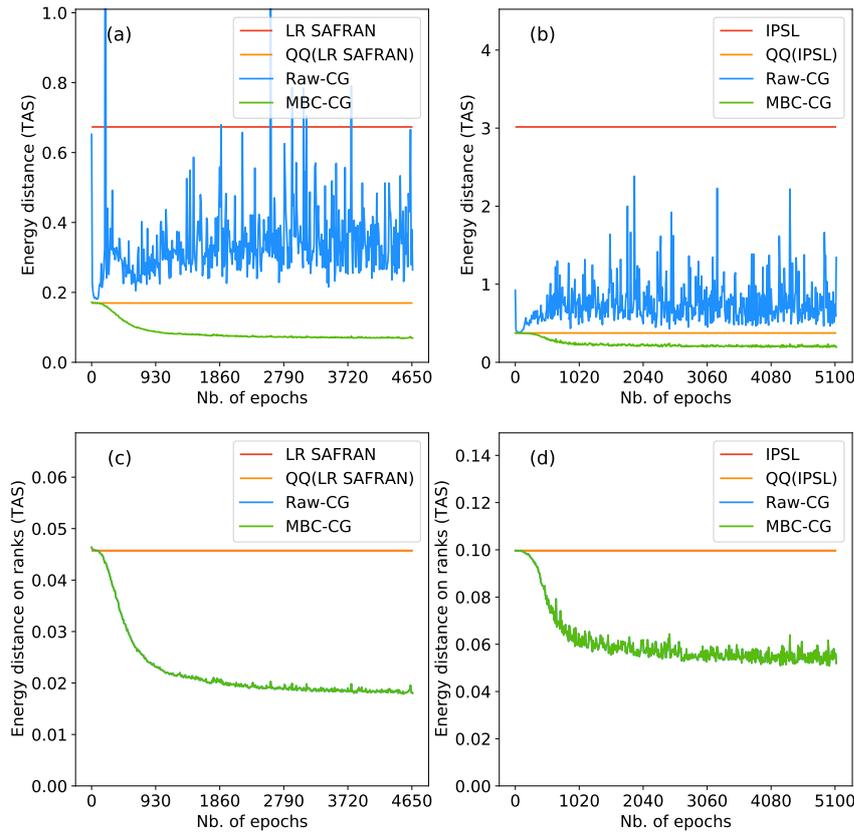


Fig. 5: Values of the energy distances with respect to SAFRAN reference for temperature computed on **(a-b)** physical values and **(c-d)** ranks during the training of MBC-CycleGAN. Results are shown for the different datasets involved in **(a, c)** the Perfect Prognosis approach and **(b, d)** the 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).

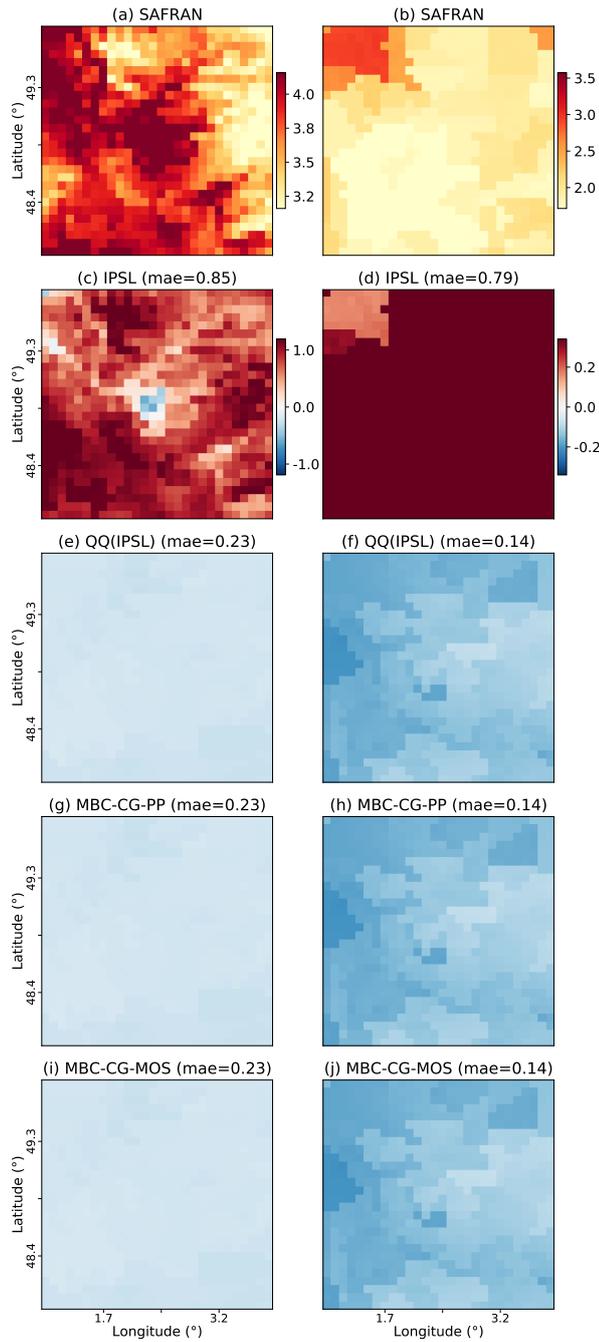


Fig. 6: Mean differences for (c, e, g, i) temperature and relative mean differences for (d, f, h, j) precipitation computed at each grid cell between SAFRAN reference and the different datasets (plain IPSL, QQ, MBC-CycleGAN-PP and MBC-CycleGAN-MOS outputs) during winter over the projection period. Note 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.

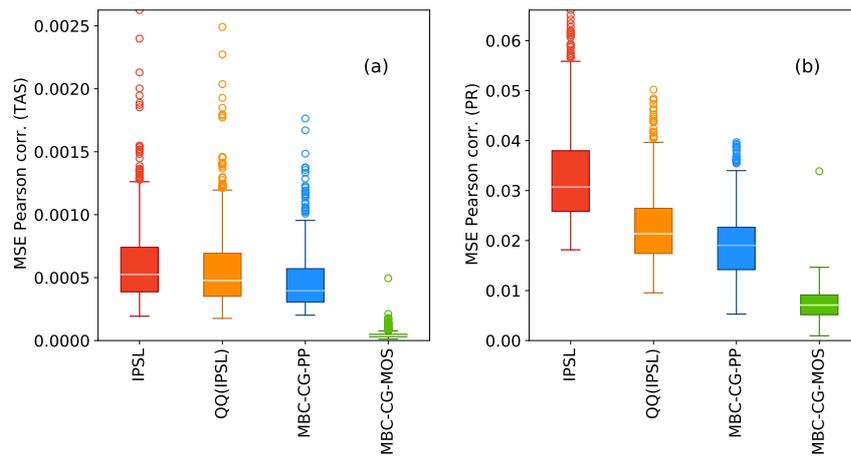


Fig. 7: Boxplots of mean squared errors of Pearson spatial correlations computed at each grid cell for **(a)** temperature and **(b)** precipitation over the projection period. Results are shown for plain IPSL, QQ, MBC-CycleGAN-PP and MBC-CycleGAN-MOS outputs.

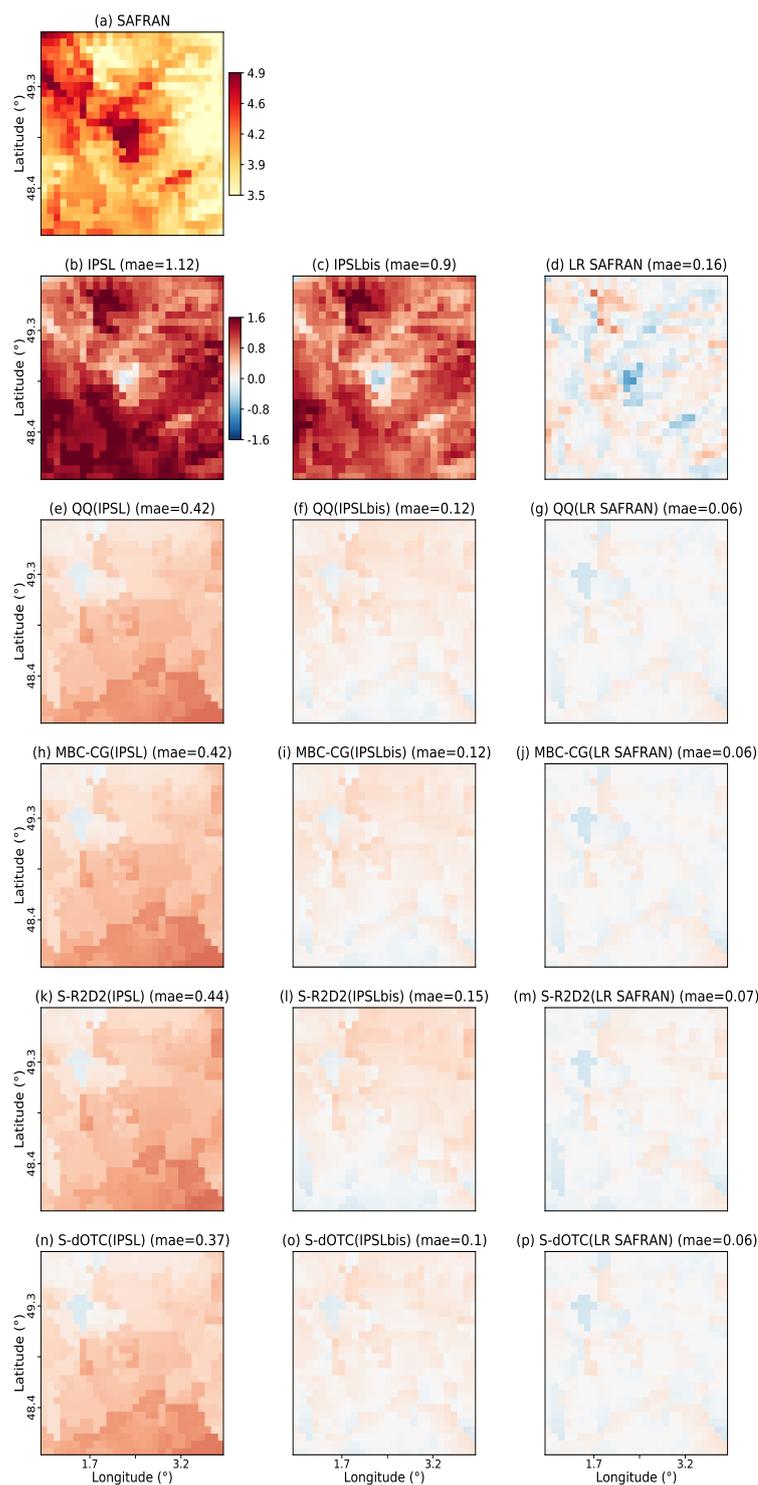


Fig. 8: Mean differences for temperature with SAFRAN reference for BC methods using as inputs (b, e, h, k, n) IPSL, (c, f, i, l, o) IPSLbis and (d, g, j, m, p) LR SAFRAN data. Results are shown during winter over the projection period for IPSL, IPSLbis, LR SAFRAN, QQ, MBC-CycleGAN, Spatial-R²D² and Spatial-dOTC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

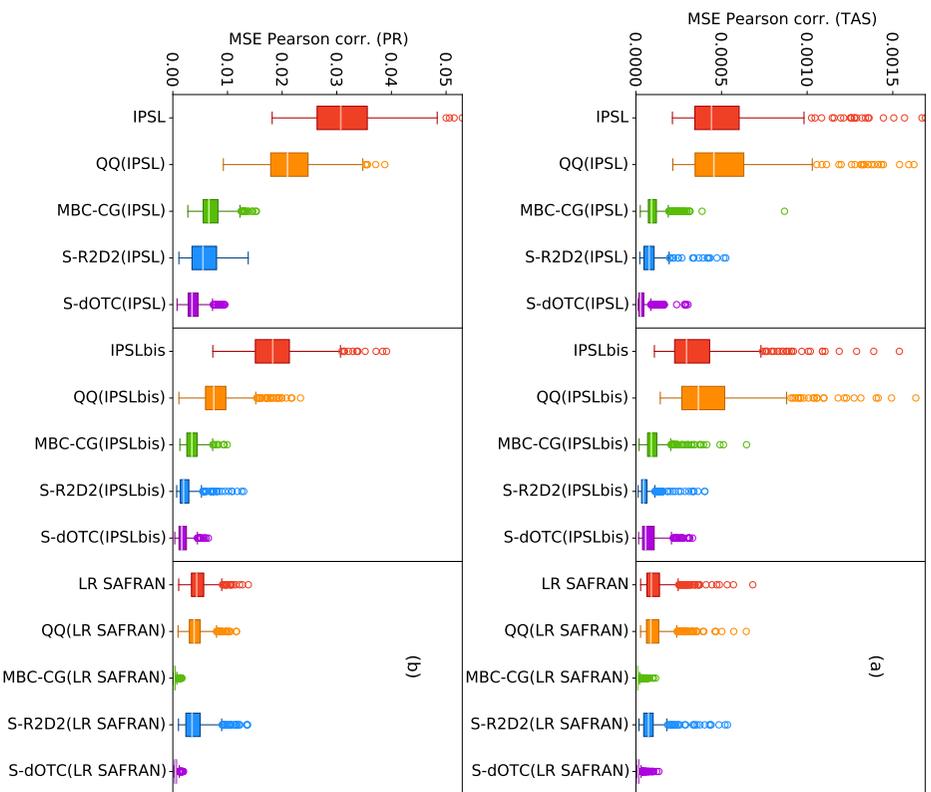


Fig. 9: Boxplots of mean squared errors of Pearson spatial correlations computed at each grid cell for (a) temperature and (b) precipitation over the projection period. Results are shown for IPSL, IPSLbis, LR SAFRAN, QQ, MBC-CycleGAN, Spatial-R²D² and Spatial-dOTC datasets.

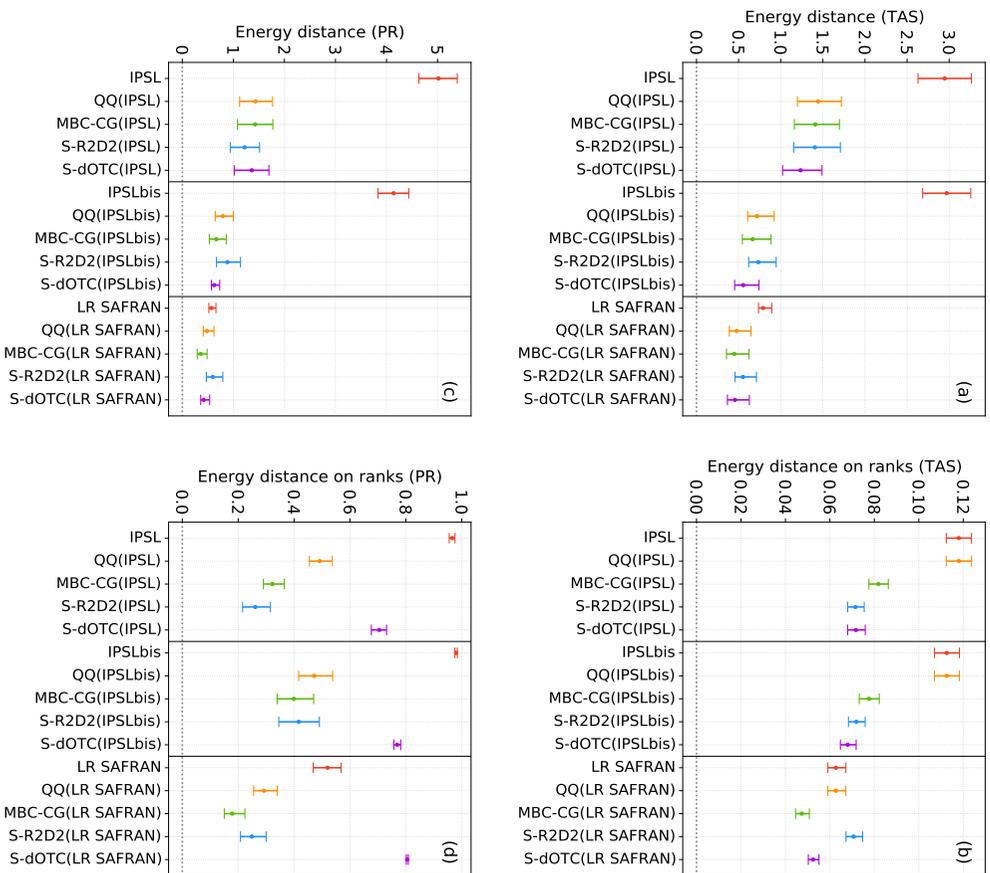


Fig. 10: Values of the estimated energy distances with respect to the reference SAFRAN for temperature (**a-b**) and precipitation (**c-d**) computed on physical values (**a, c**) and ranks (**b, d**) during the projection period. Results are presented for IPSL, IPSLbis, LR SAFRAN, QQ, MBC-CycleGAN, 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.

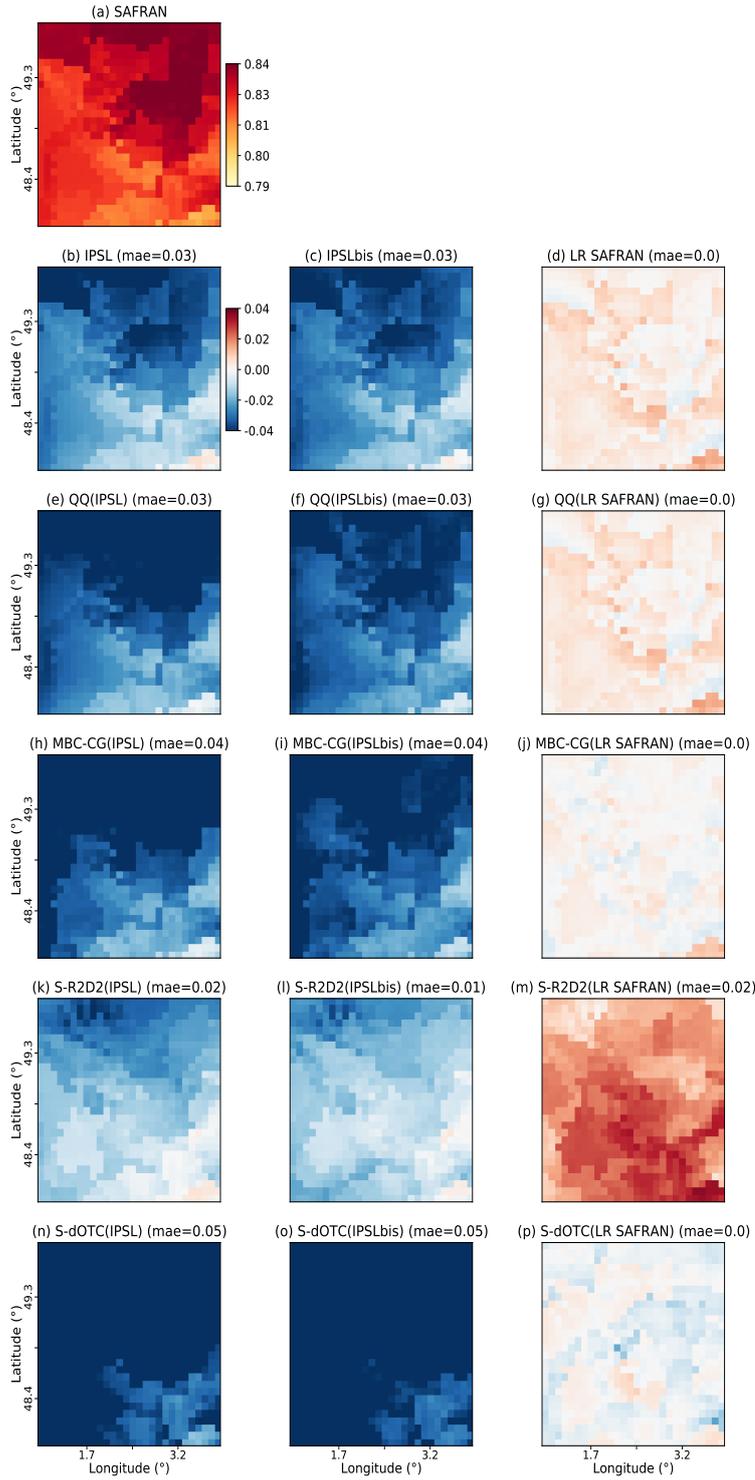


Fig. 11: Differences of order 1 Pearson autocorrelation for temperature with SAFRAN reference for BC methods using as inputs **(b, e, h, k, n)** IPSL, **(c, f, i, l, o)** IPSLbis and **(d, g, j, m, p)** LR SAFRAN data. Results are shown during winter over the projection period for IPSL, IPSLbis, LR SAFRAN, 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

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. 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.

Physical Variable	LR SAFRAN	QQ	MBC-CG	S-R ² D 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

1541 A Details on the MBC-CycleGAN method

1542 Let consider the correction of a random variable, denoted \mathbf{X} (e.g., biased climate simulations
 1543 outputs) with respect to a reference random variable, denoted \mathbf{Y} . In our study, \mathbf{X} and \mathbf{Y}
 1544 live in dimension $28 \times 28 = 784$ dimensions. We denote \mathbf{X}^0 and \mathbf{X}^1 the random variables to
 1545 correct from climate simulations during the calibration and projection period, respectively.
 1546 Similarly, \mathbf{Y}^0 is considered as the random variable of references for the calibration period.
 1547 The goal of any BC methods is to infer future unobserved data \mathbf{Y}^1 from the reference
 1548 variable \mathbf{Y}^0 during calibration, and the variables from model simulations for calibration
 1549 (\mathbf{X}^0) and projection (\mathbf{X}^1) periods.

1550 In practice, BC methods are applied to correct samples $(\mathbf{x}_1^0, \dots, \mathbf{x}_n^0)$ and $(\mathbf{x}_1^1, \dots, \mathbf{x}_n^1)$
 1551 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 ran-
 1552 dom 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
 1553 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 Section 3,
 1554 the CycleGAN approach within the MBC-CycleGAN methodology is applied between 1d-
 1555 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,
 1556 as well as two discriminators $D_{\mathbf{Q}\mathbf{Q}}$ and $D_{\mathbf{Y}}$. The different steps constituting the MBC-
 1557 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 $\{\mathbf{q}\mathbf{q}_i^0\}_{i=1}^{m/2}$ 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 $\{\widetilde{\mathbf{y}}_i^{0,fake}\}_{i=1}^{m/2}$:
 $\forall i \in \llbracket 1, \dots, m/2 \rrbracket, \widetilde{\mathbf{y}}_i^{0,fake} = G_{\mathbf{Q}\mathbf{Q}\rightarrow\mathbf{Y}}(\widetilde{\mathbf{q}\mathbf{q}}_i^0)$.
 - 8: Generate "fake" samples $\{\widetilde{\mathbf{q}\mathbf{q}}_i^{0,fake}\}_{i=1}^{m/2}$:
 $\forall i \in \llbracket 1, \dots, m/2 \rrbracket, \widetilde{\mathbf{q}\mathbf{q}}_i^{0,fake} = G_{\mathbf{Y}\rightarrow\mathbf{Q}\mathbf{Q}}(\widetilde{\mathbf{y}}_i^0)$.
 - 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:
 $\forall i \in \llbracket 1, \dots, n \rrbracket, \widetilde{\mathbf{s}}_i^0 = G_{\mathbf{Q}\mathbf{Q}\rightarrow\mathbf{Y}}(\widetilde{\mathbf{q}\mathbf{q}}_i^0)$.
 - 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:
 $\forall i \in \llbracket 1, \dots, p \rrbracket, \widetilde{\mathbf{s}}_i^1 = G_{\mathbf{Q}\mathbf{Q}\rightarrow\mathbf{Y}}(\widetilde{\mathbf{q}\mathbf{q}}_i^1)$.
 - 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 $(\mathbf{x}_1^1, \dots, \mathbf{x}_p^1)$.
-

1558 B Details on the simple architecture of neural networks used in 1559 MBC-CycleGAN

1560 The simple neural network architectures used for the discriminators and generators con-
1561 stituting the MBC-CycleGAN method in this study are described with more details in this
1562 appendix.

1563 B.1 Architecture of the generators

1564 As explained in subsection 3.3.2, skip connections are used in the architecture of the gen-
1565 erators to ease the training process. Skip connections permit to provide information to a
1566 given layer that comes not only from the direct previous layer, but also from other upstream
1567 convolution layers in the architecture. Skipping over layers permits to avoid vanishing gra-
1568 dients issues, which is a problem that can make the network hard to train. All layers except
1569 the first one have leaky rectified linear unit (leaky-ReLU) activation functions defined as:

$$1570 y = \begin{cases} x & \text{if } x \geq 0, \\ \alpha x & \text{otherwise,} \end{cases} \quad \text{with } \alpha = 0.2. \text{ Dropout regularization, that refers to ignoring neurons}$$

1571 chosen at random during training, is used after the second and third 2D convolutional lay-
1572 ers to prevent from overfitting. The probability used for dropout is 0.4. A summary of the
1573 simple neural network architecture used for the generators is described below in Table B1.

Table B1: 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.	$28 \times 28 \times 1$

1574 B.2 Architecture of the discriminators

1575 A summary of the simple neural network architecture used for the discriminators is described
1576 below in Table B2.

Table B2: 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 Leaky ReLU + Dropout	$3 \times 3 \times 64$ n.a.	(2, 2) n.a.	$14 \times 14 \times 64$ $14 \times 14 \times 64$
3	Conv2D Leaky ReLU + Dropout	$3 \times 3 \times 128$ n.a.	(2, 2) n.a.	$7 \times 7 \times 128$ $7 \times 7 \times 128$
4	Flatten	n.a.	n.a.	6272
5	Dense + sigmoid	1	n.a.	1

1577 C Methodology for the generation of IPSLbis

1578 For the generation of IPSLbis data, a two-step procedure is developed to construct, from
 1579 IPSL data, climate data that present marginal and spatial changes that are in line with
 1580 those from references between the calibration and projection periods. In order to stay with
 1581 comparable changes as those from LR SAFRAN, LR SAFRAN changes are reproduced. We
 1582 recall that, for the calibration period, IPSL and IPSLbis data are strictly identical. The
 1583 two-step procedure is only used to produce alternative climate data for the projection period.

1584 C.1 Marginal changes with CDF-t

1585 The first step of the procedure consists in producing time series for the projection period of
 1586 IPSLbis by taking into account marginal changes of LR SAFRAN with the 1d-BC named
 1587 CDFt (Vrac et al 2012). Initially, CDF-t is a version of univariate quantile mapping method
 1588 designed to correct at each individual grid cell marginal properties of climate simulations
 1589 outputs during the calibration and the projection period according to the data from the refer-
 1590 ence observed during calibration. CDF-t, by defining a specific transfer function, has been
 1591 conceived to take into account the potential simulated changes of univariate distributions
 1592 from the calibration to the projection period in order to produce the adjusted data such
 1593 that the marginal changes are in line with those from the simulations. While, traditionally,
 1594 this quantile-mapping approach is used to find, in a bias correction context, a mathematical
 1595 transformation allowing to go from simulations to references, we here applied CDF-t
 1596 to go from “large scale” references (LR SAFRAN) to simulations for future periods. By
 1597 proceeding this way, the produced time series are projected distributions in the domain of
 1598 IPSL simulations that have been obtained while taking into account the potential evolution
 1599 of CDFs of the LR SAFRAN dataset between the calibration and projection periods. By
 1600 concatenating times series from IPSL for the calibration period and those obtained from the
 1601 CDF-t method for the projection period, new climate times series are obtained, presenting
 1602 marginal distributions changes in line with those from references.

1603 C.2 Spatial changes with a matrix-recorrelation technique

1604 The second step consists in deriving a spatial dependence structure for the projection pe-
 1605 riod such that spatial changes of LR SAFRAN are reproduced. To do so, we take advantage
 1606 of a matrix-recorrelation technique used for the MBC method presented in Bárdossy and
 1607 Pegram (2012) to impose to climate data a specific spatial dependence structure for the pro-
 1608 jection period. Our methodology is summarized in Table C1. It consists in first projecting

1609 individually each variable of both IPSL simulations and LR SAFRAN during calibration
 1610 and projection periods to the univariate normal distribution with a Gaussian quantile map-
 1611 ping method. This ‘‘Gaussianization’’ step is particularly suited for variables with mixed
 1612 distributions such as precipitation (composed of wet and dry events). Computing Pearson
 1613 correlation matrices on such Gaussianized data instead of raw data permits to better describe
 1614 its dependence structure. Thus, Pearson correlation matrices of the different Gaussianized
 1615 data are computed. They are respectively denoted as $C_{I,C}$, $C_{I,P}$, $C_{I,C}^{(bis)}$, $C_{I,P}^{(bis)}$, $C_{S,C}$, $C_{S,P}$
 1616 for IPSL during calibration, IPSL during projection, IPSLbis during calibration, IPSLbis
 1617 during projection, LR SAFRAN during calibration and LR SAFRAN during projection.
 1618 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
 1619 construction, $C_{I,C}$ is the same as $C_{I,C}^{(bis)}$ and that $C_{I,P}^{(bis)}$ is unknown. Assessing the changes
 1620 of LR SAFRAN spatial correlations between calibration and projection periods is now re-
 1621 quired to derive the spatial dependence structure of IPSLbis for the projection period. A
 1622 simple approach to determine $r_{I,P}^{(bis)}$, the correlation of the Gaussianized data of IPSLbis for
 1623 projection, would be to compute it based on the difference of correlations from Gaussianized
 1624 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
 1625 can lead to obtain correlation values that are out of range, i.e. being greater than 1 or less
 1626 than -1, which is not appropriate.

1627 From Bárdossy and Pegram (2012), given $r_{I,C}$, $r_{S,C}$, $r_{S,P}$, one can derive r_{IbP} using
 1628 Fisher-Z transformation (Fisher 1915) as following:

$$r_{I,P}^{(bis)} = \frac{\frac{(1+r_{S,P})}{(1+r_{S,C})}(1+r_{I,C}) - \frac{(1-r_{S,P})}{(1-r_{S,C})}(1-r_{I,C})}{\frac{(1+r_{S,P})}{(1+r_{S,C})}(1+r_{I,C}) + \frac{(1-r_{S,P})}{(1-r_{S,C})}(1-r_{I,C})}. \quad (C1)$$

1629 Fisher-Z transformation permits to transform a bounded random variable to another
 1630 random variable that can be assumed to be Normal, and for which additive correction
 1631 can be performed (see Mehrotra and Sharma (2019) for the derivation of Eq. C1). By
 1632 deriving this way all the new correlation coefficients, the potential changes in correlations in
 1633 the Gaussianized LR SAFRAN data are preserved and the Pearson correlation matrix for
 1634 Gaussianized IPSLbis during the projection period is obtained.

1635 Now that the Pearson correlation matrix, $C_{I,P}^{(bis)}$, is computed, a combination of ‘‘decor-
 1636 relation’’ and ‘‘re-correlation’’ steps using decompositions of correlation matrices through sin-
 1637 gular value decomposition (SVD, Beltrami 1873; Jordan 1874a,b; Stewart 1993) is applied
 1638 on the Gaussianized data of IPSL during projection period, forcing its Pearson correlation
 1639 matrix to be exactly the same as the Pearson correlation matrix, $C_{I,P}^{(bis)}$. The new depen-
 1640 dence structure for IPSLbis is obtained. Finally, a reordering of time series from CDF-t
 1641 outputs according to this new dependence structure is performed using the Schaake Shuffle
 1642 method to obtain IPSLbis data for the projection period.

Table C1: 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. C1)	$C_{S,C} C_{S,P}$
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		$I_P^{(bis)}$	

1643 D Spatial correlation changes analysis

1644 We present a spatial changes analysis to provide a better picture of the properties of the
1645 climate data in terms of changes between the calibration and projection periods. As a
1646 reminder, IPSLbis data are generated using the two-step procedure described in Appendix C
1647 such that its marginal and dependence changes are in line with those from LR SAFRAN (and
1648 therefore SAFRAN) for the projection period. Fig. S3 displays scatterplots of differences
1649 between Spearman spatial correlations of temperature and precipitation evaluated for all
1650 pairwise combinations of sites, computed for the calibration (1979-2005) and the projection
1651 (2006-2016) period, respectively. Scatterplots compares differences of Spearman correlation
1652 with respect to those from LR SAFRAN. It permits one to visually verify if changes in the
1653 spatial dependence structure are in line to those from references at large-scale. Using rank
1654 correlation here permits to measure in isolation the spatial dependence between two sites
1655 rid of their marginal properties. Figures for the analysis of marginal changes - in particular,
1656 mean and standard deviation changes - are also displayed in Figs. S4 and S5 for information
1657 purposes only. As expected, changes in spatial correlations from SAFRAN references are
1658 (partially) in agreement with those from LR SAFRAN for both temperature (Fig. S3a) and
1659 precipitation (Fig. S3d). Concerning changes in the IPSL simulations, simulated changes
1660 of spatial correlations for temperature (Fig. S3b) are globally in line with those from LR
1661 SAFRAN, highlighting the ability of the climate model to provide appropriate temperature
1662 changes in spatial structure between the calibration and the projection periods. However,
1663 conclusions are quite different for precipitation, for which simulated changes are not in
1664 agreement at all with those from the reference at large scale (Fig. S3e). Hence, IPSL model
1665 presents discrepancy of changes for precipitation with respect to LR SAFRAN (and thus,
1666 SAFRAN references), that could potentially affect the quality of the correction depending
1667 on how MBC-CycleGAN accounts for these changes in its correction procedure. Concerning
1668 the results for IPSLbis, changes for both temperature (Fig. S3c) and precipitation (Fig. S3f)
1669 are similar to those from LR SAFRAN, confirming that the two-step methodology used to
1670 impose to IPSL specific changes of spatial correlations is appropriate here.

Figures

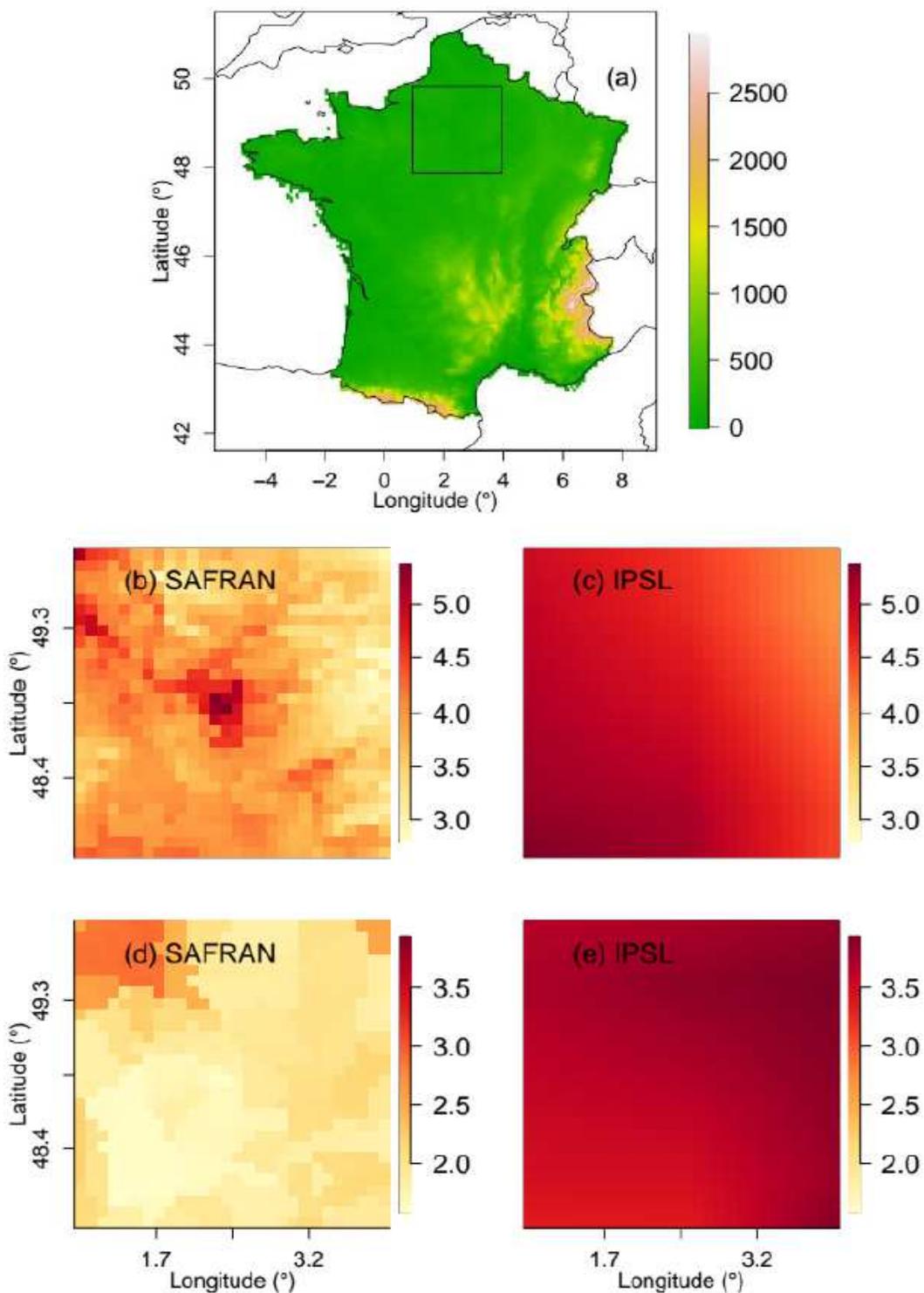
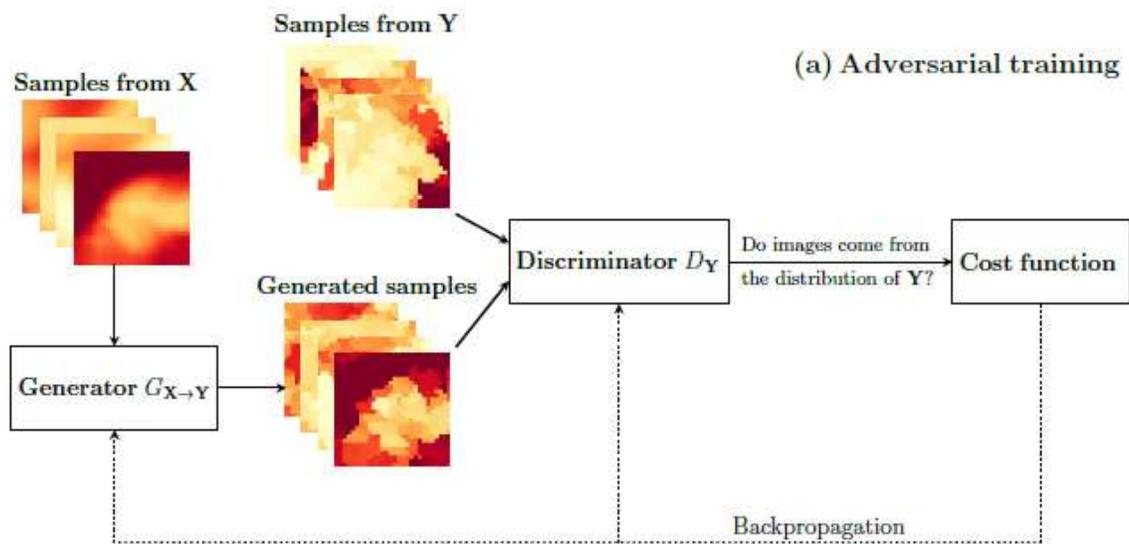
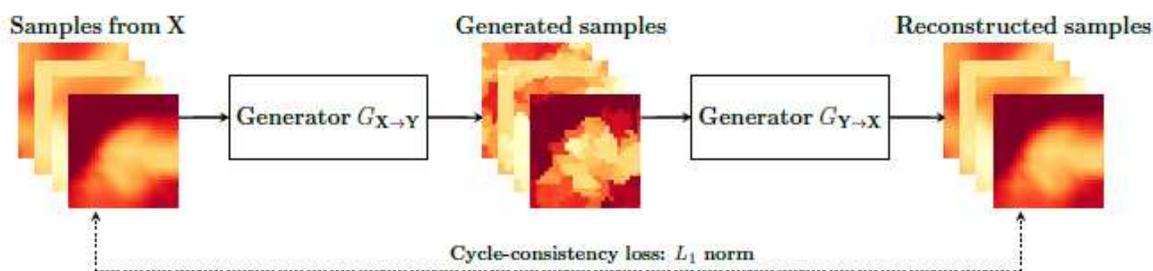


Figure 1

(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.



(b) Cycle-Consistency



(c) Identity

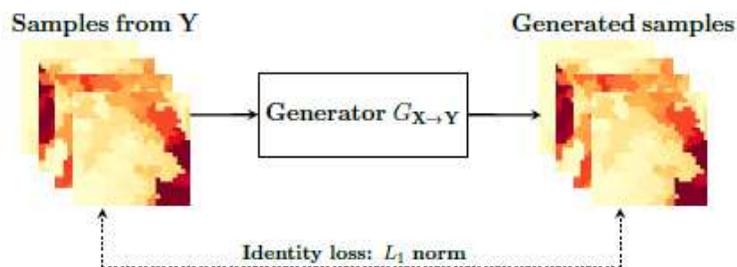


Figure 2

(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 $G_X \circ Y$ applied to images from Y to resemble to initial inputs from Y (and vice versa for $G_Y \circ X$). In our study, samples from X and Y are replaced by QQ outputs and references, respectively.

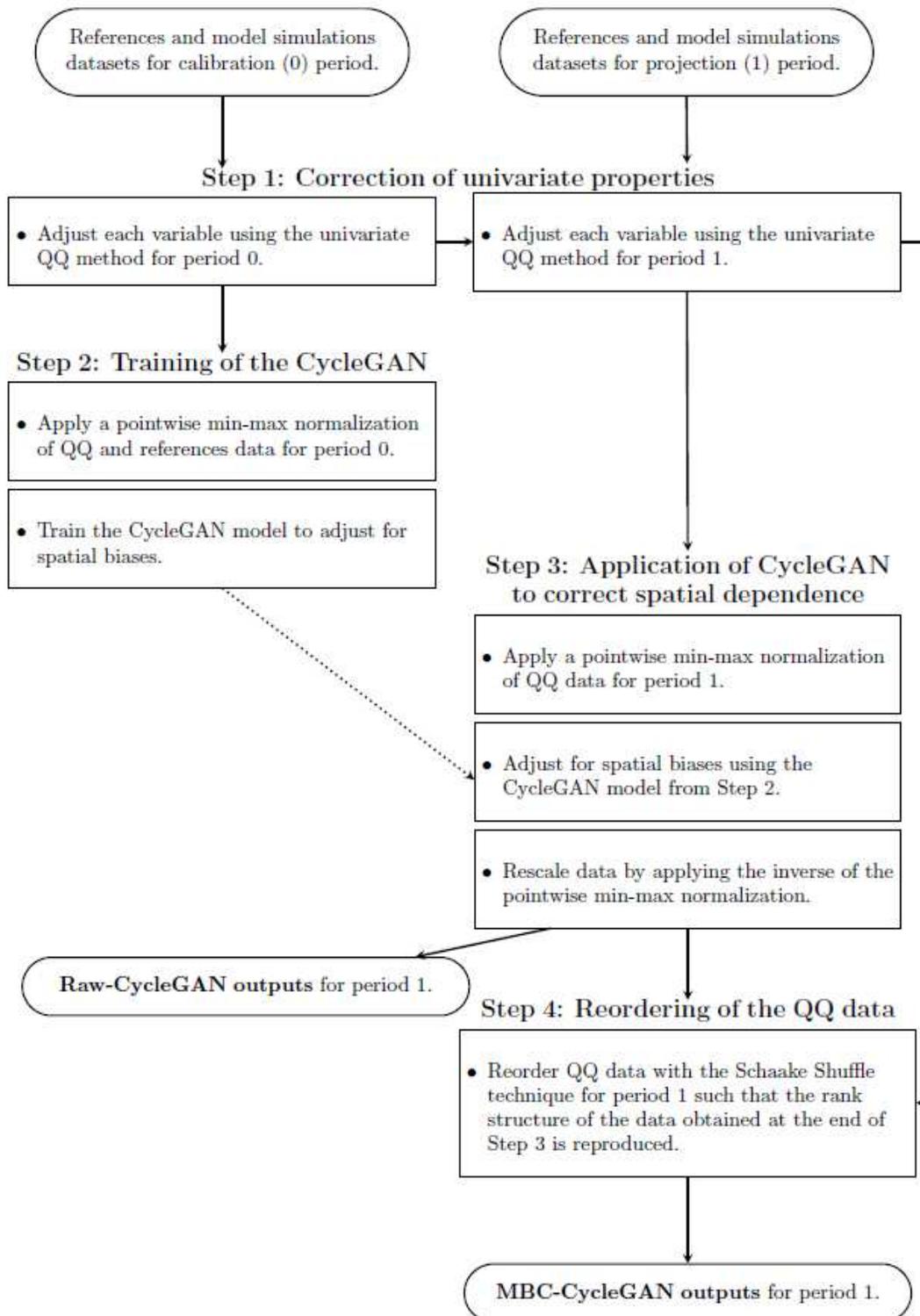


Figure 3

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

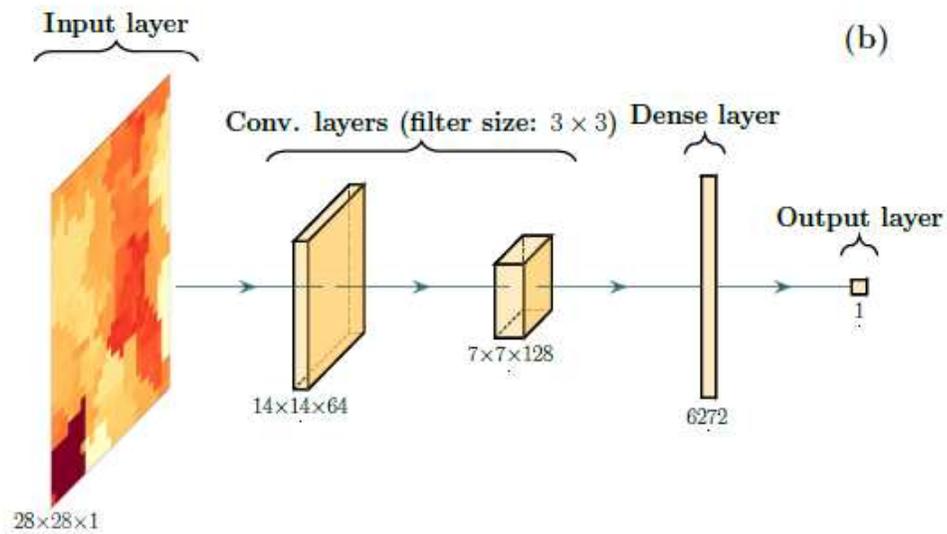
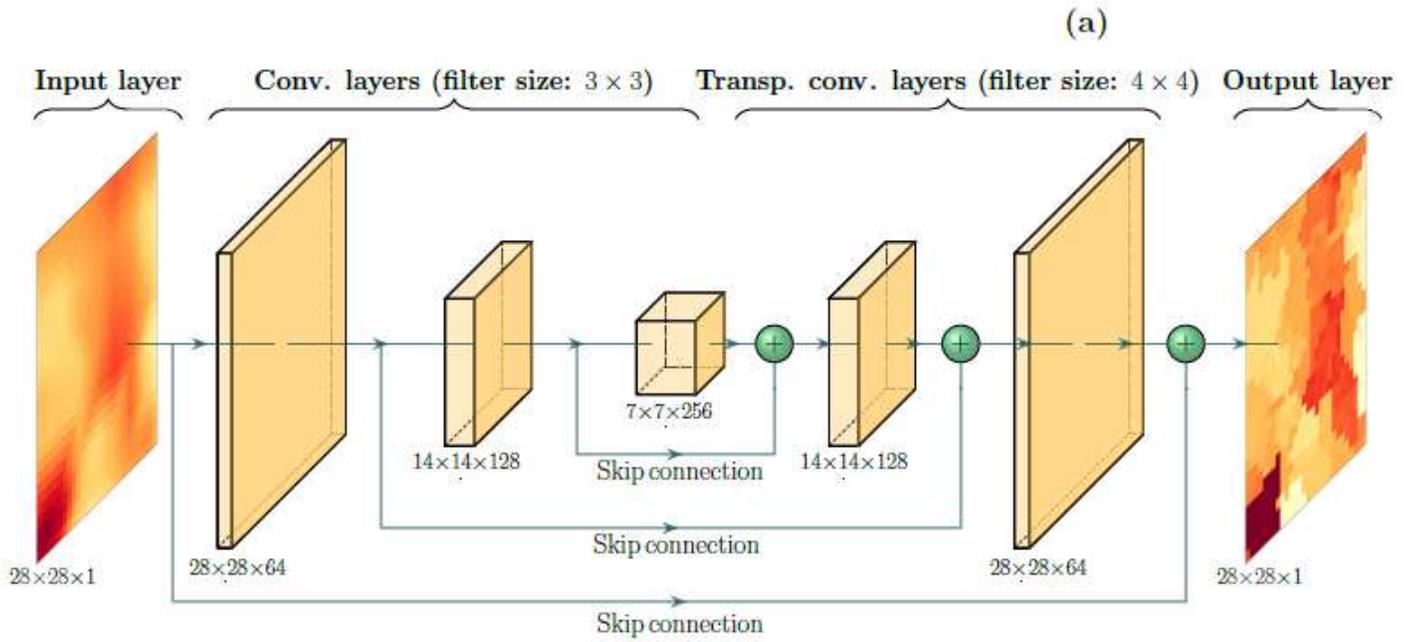


Figure 4

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.

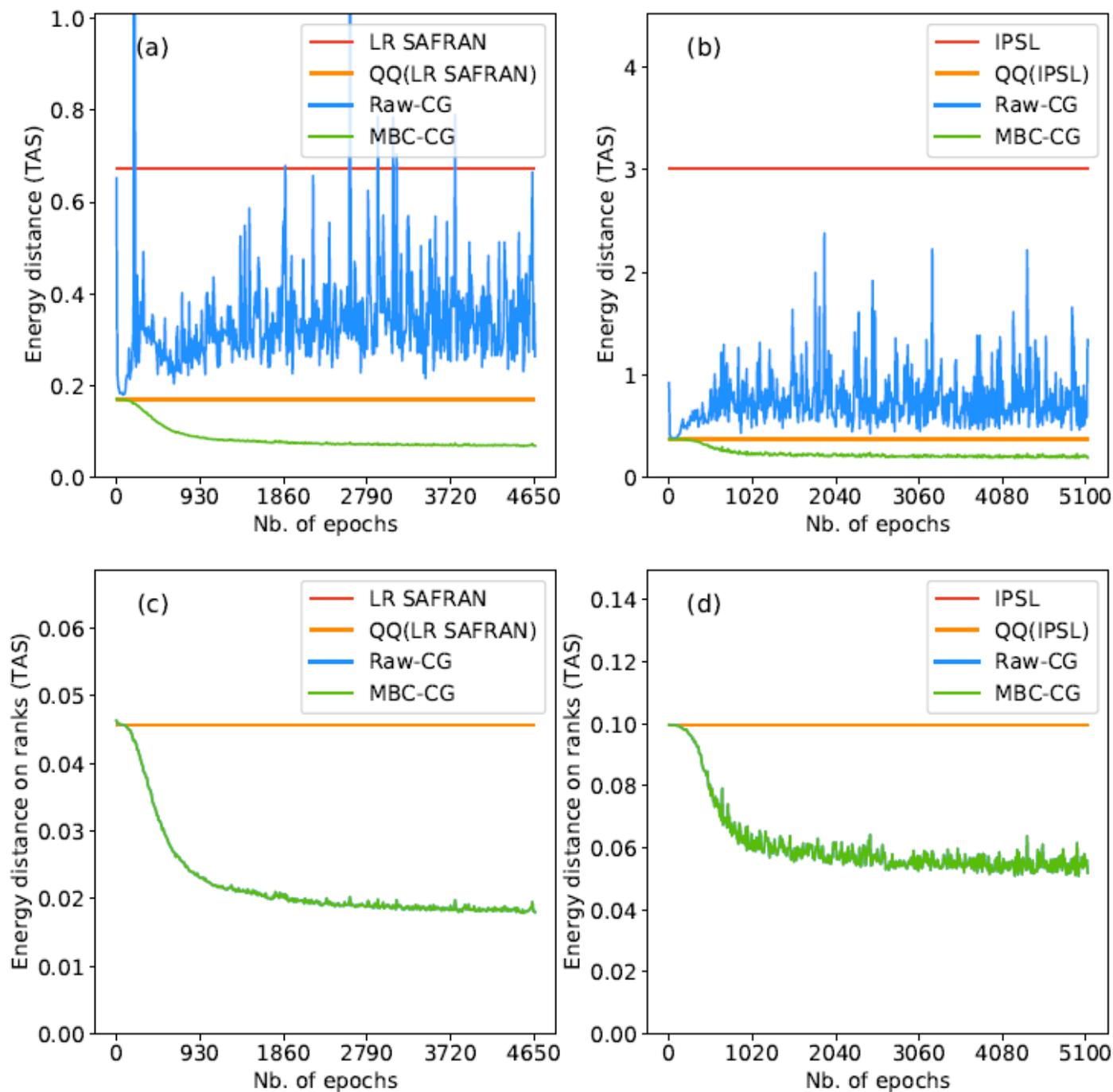


Figure 5

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).

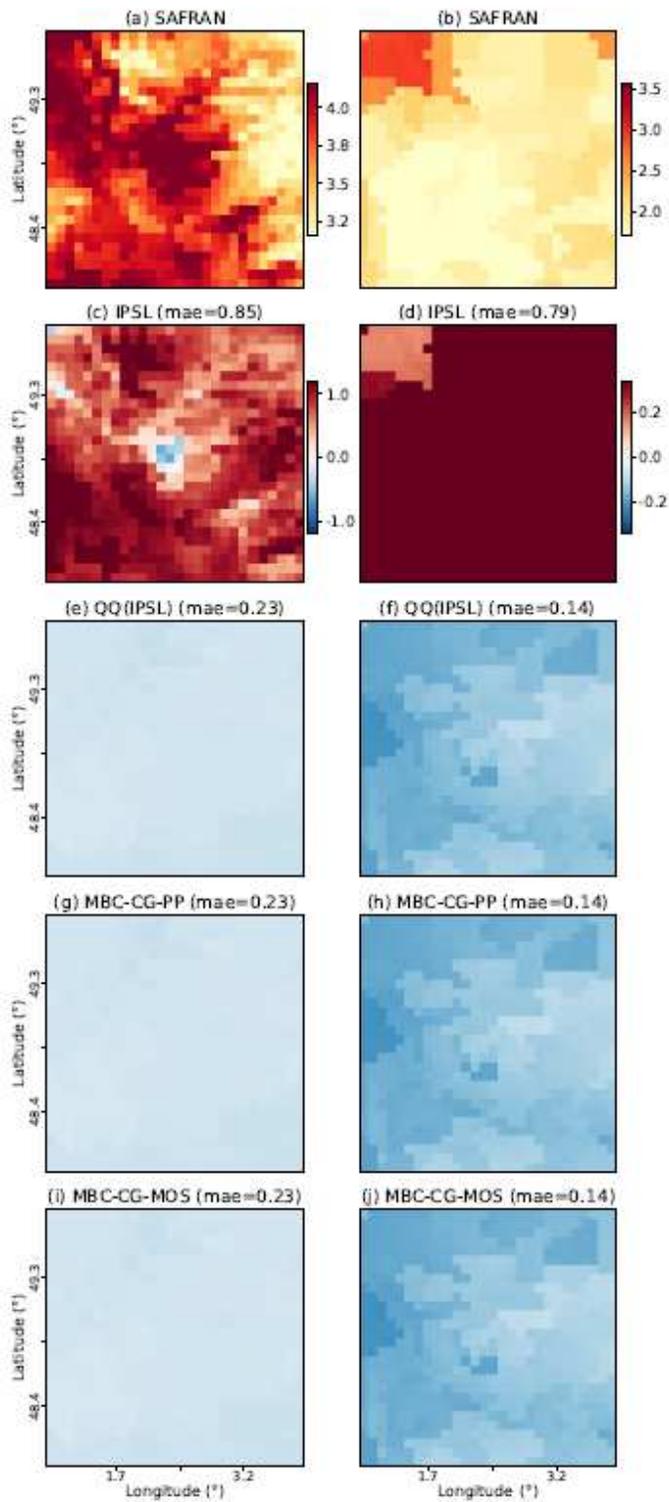


Figure 6

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.

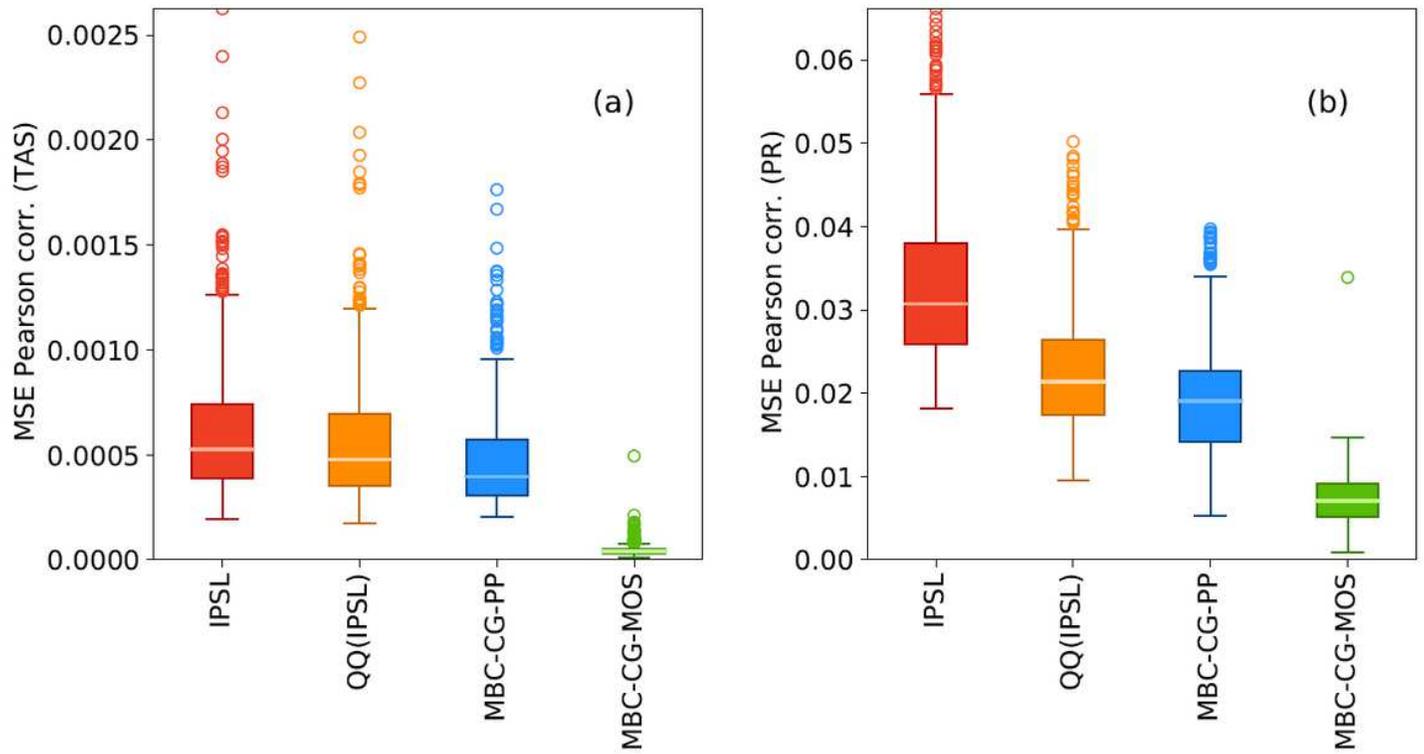


Figure 7

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.

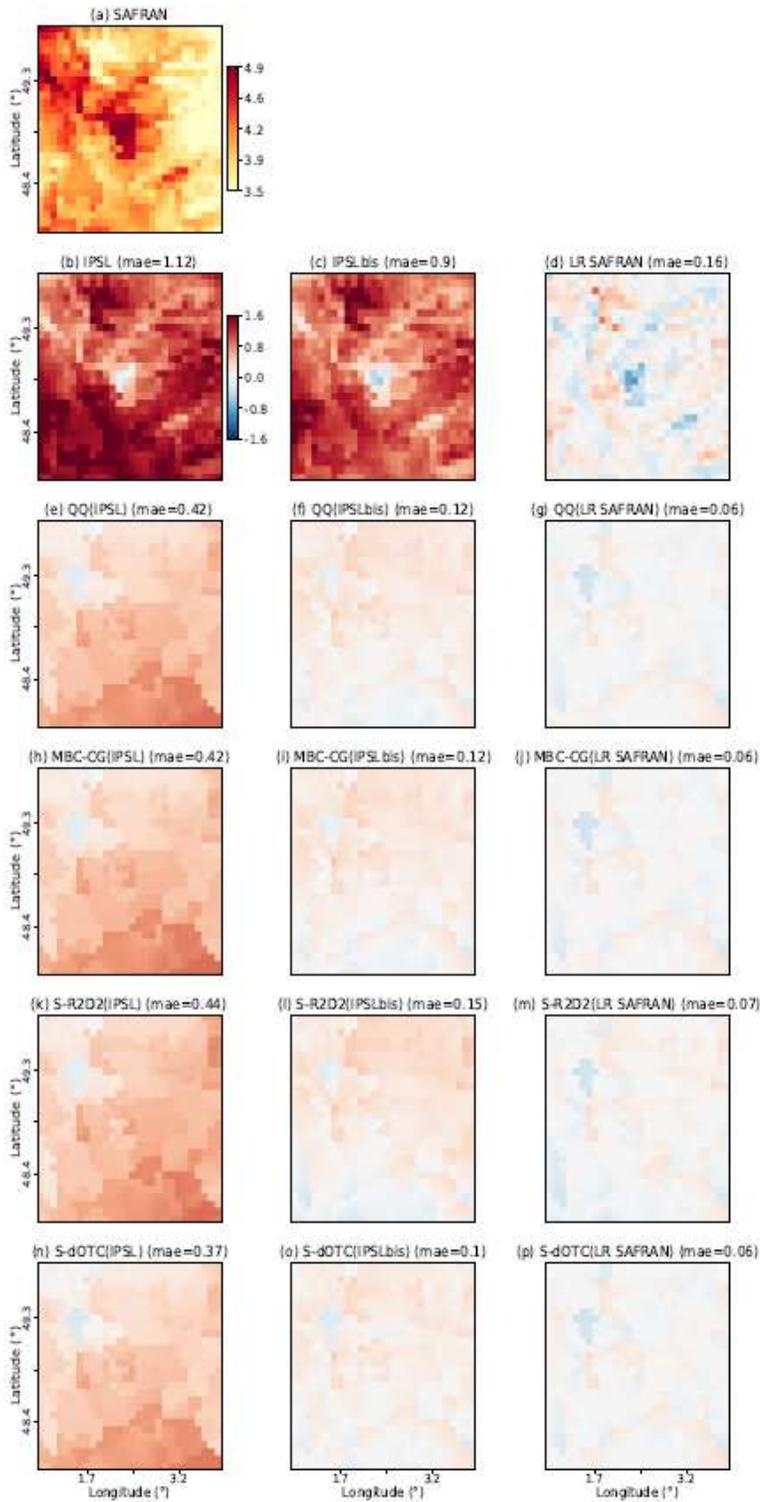


Figure 8

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).

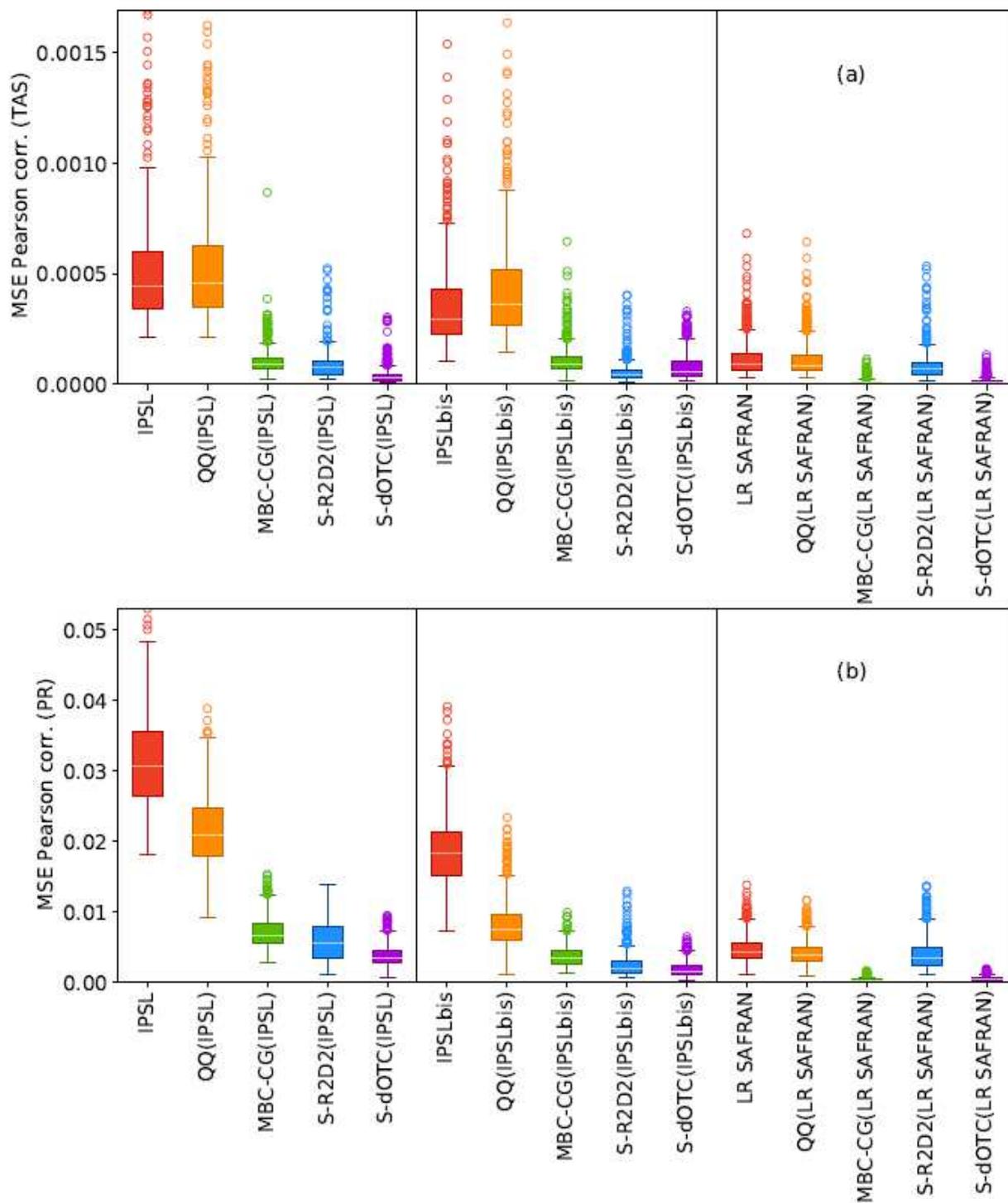


Figure 9

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.

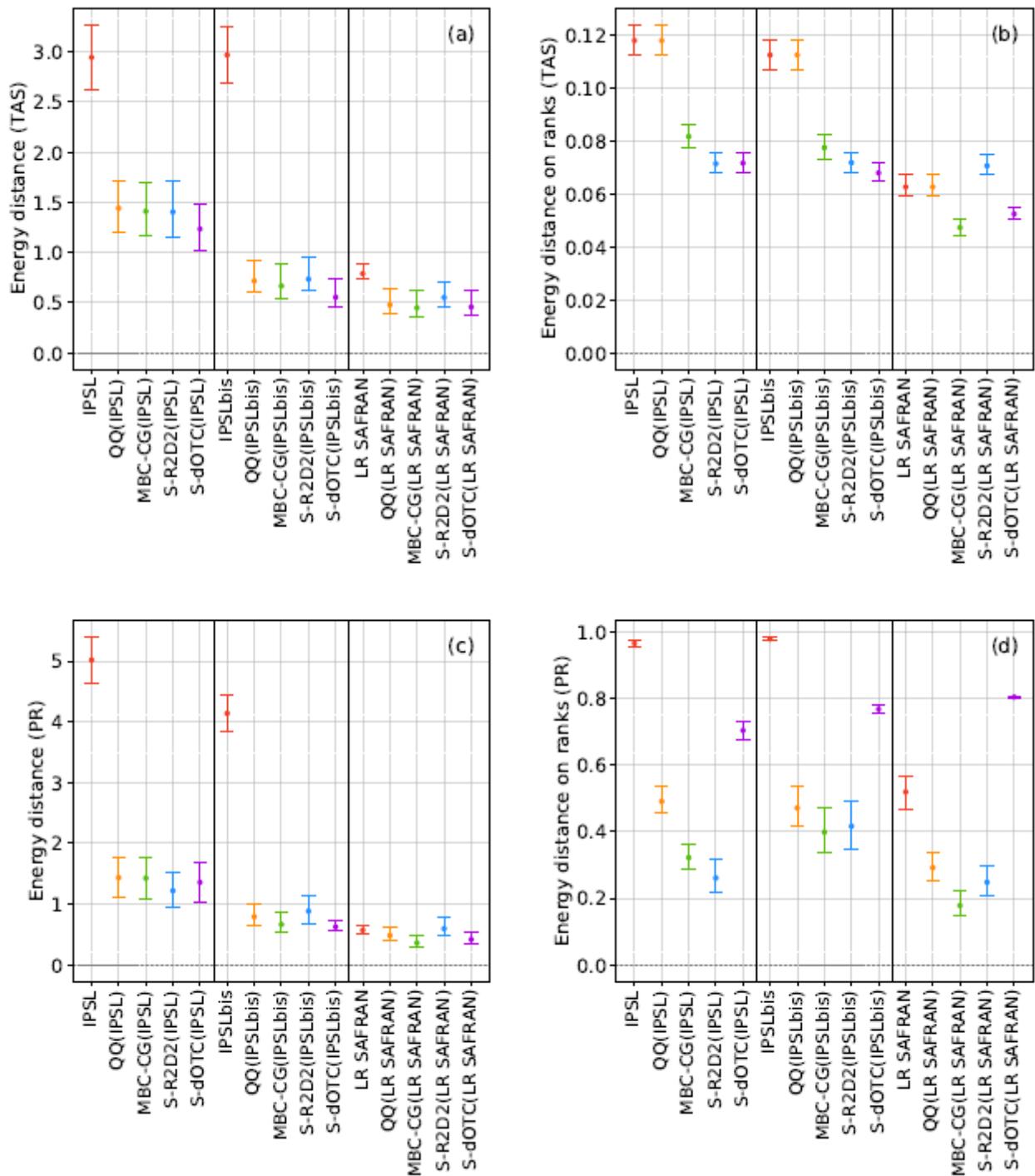


Figure 10

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.

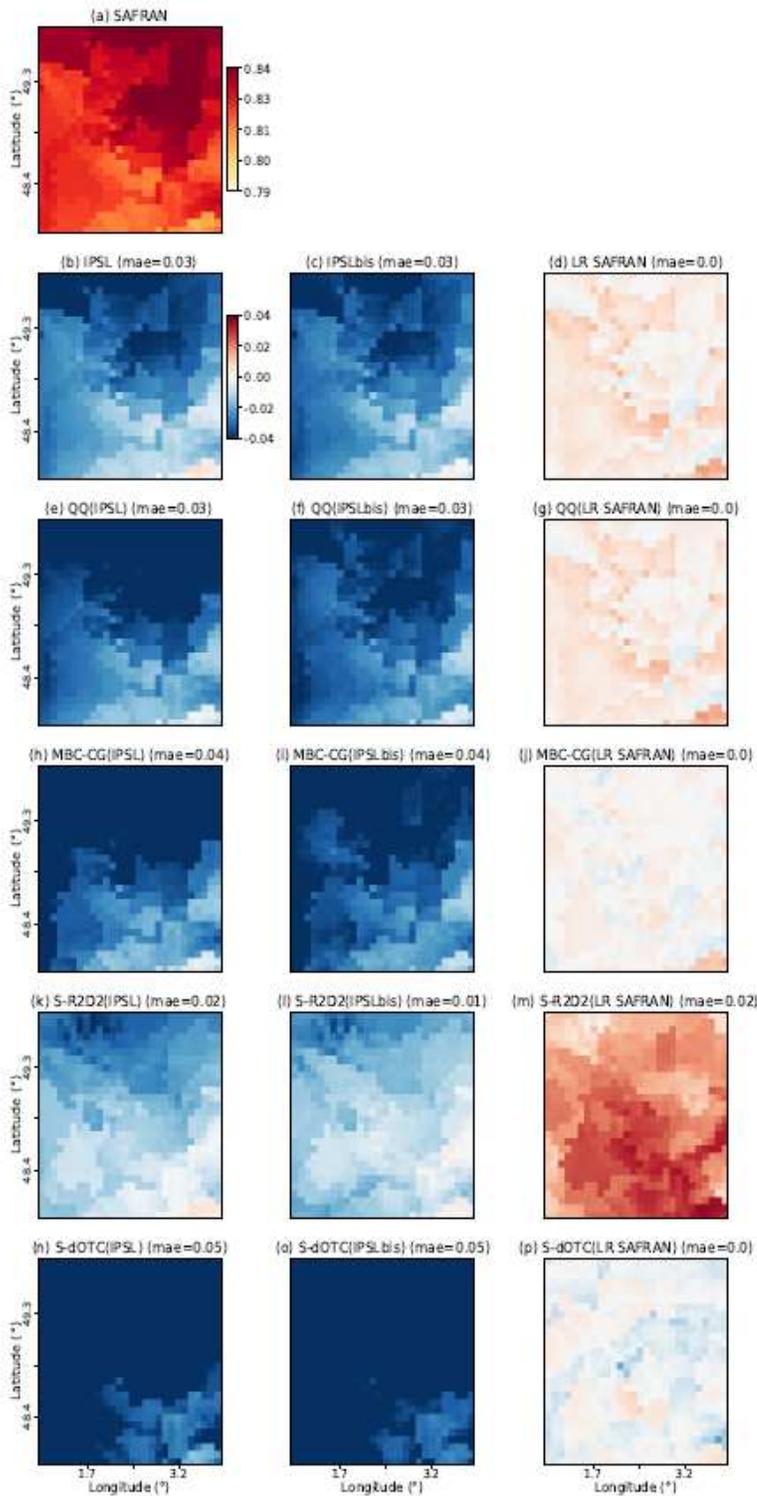


Figure 11

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).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Animation1.mp4](#)
- [Animation2.mp4](#)
- [FrancoisClimateDynamics2021Supplementary.pdf](#)