AGU PUBLICATIONS

JGR

Journal of Geophysical Research: Atmospheres

RESEARCH ARTICLE

10.1002/2014JD022236

Key Points:
- Statistical downscaling of precipitation
- Temporal transferability
- Precipitations changes over France

Correspondence to:
G. Dayon,
dayon@cerfacs.fr

Citation:
Dayon, G., J. Boé, and E. Martin (2015), Transferability in the future climate of a statistical downscaling method for precipitation in France, *J. Geophys. Res. Atmos.*, *120*, doi:10.1002/2014JD022236.

Received 27 JUN 2014
Accepted 30 DEC 2014
Accepted article online 7 JAN 2015

Transferability in the future climate of a statistical downscaling method for precipitation in France

G. Dayon[1], J. Boé[1], and E. Martin[2]

[1]Climate Modelling and Global Change Team, CERFACS/CNRS, Toulouse, France, [2]GAME/CNRM, Météo-France, CNRS, Toulouse, France

Abstract A statistical downscaling approach for precipitation in France based on the analog method and its evaluation for different combinations of predictors is described, with focus on the transferability of the method to the future climate. First, the realism of downscaled present-day precipitation climatology and interannual variability for different combinations of predictors from four reanalyses is assessed. Satisfactory results are obtained, but elaborated predictors do not lead to major and consistent across-reanalyses improvements. The downscaling method is then evaluated on its capacity to capture precipitation trends in the last decades. As uncertainties in downscaled trends due to the choice of the reanalysis are large and observed trends are weak, this analysis does not lead to strong conclusions on the applicability of the method to a changing climate. The temporal transferability is then assessed thanks to a perfect model framework. The statistical downscaling relationship is built using present-day predictors and precipitation simulated by 12 regional climate models. The entire projections are then downscaled, and future downscaled and simulated precipitation changes are compared. A good temporal transferability is obtained only with a specific combination of predictors. Finally, the regional climate models are downscaled, thanks to the relationship built with reanalyses and observations, for the best combination of predictors. Results are similar to the changes simulated by the models, which reinforces our confidence in the realism of the models and of the downscaling method. Uncertainties in precipitation change due to reanalyses are found to be limited compared to those due to regional simulations.

1. Introduction

The coarse resolution of current global climate models (GCMs; generally between 150 and 250 km, see Table 9.1 in *Stocker et al.* [2013]) does not allow for a realistic representation of local physiographic features, such as orography and coasts, which play a crucial role in shaping local climates. It is especially problematic for impact studies, as impact models generally require relevant meteorological forcings at kilometric resolutions. To bridge the resolution gap between GCMs and impact models, spatial downscaling techniques have been developed.

Dynamical downscaling (DD) is generally based on high-resolution regional climate models (RCMs) forced at their boundaries by GCMs. Relatively large ensembles of regional climate projections are now available (e.g., the ENSEMBLES European project [*van der Linden et al.*, 2009] and more recently CORDEX [*Jacob et al.*, 2013]). These data sets are still much less complete than equivalent multimodel ensembles of global climate simulations as the Coupled Model Intercomparison Project Phase 5 (CMIP5). It is therefore still difficult to fully address some scientific questions in impact studies using only RCMs, such as the estimation of the uncertainties associated with internal variability. Indeed, multiple members for a given RCM are seldom available. Another weakness of DD is that bias correction generally remains necessary before impact modeling, which raises important methodological issues [*Muerth et al.*, 2013].

Statistical downscaling (SD) is based on developing an empirical statistical relationship between large-scale climate variables and the local variables of interest for the impact study. Various SD methods have been developed during the last few decades (see review by *Maraun et al.* [2010], and references therein). For precipitation over France, the focus of this study, several SD methods have been developed for specific regions [*Obled*, 2002; *Timbal et al.*, 2003] and others have dealt with the whole country as *Radanovics et al.* [2013]. SD techniques have for main advantage to be computationally inexpensive. It is therefore possible to downscale a large ensemble of GCMs, which is generally necessary to deal correctly with the different uncertainties involved in impact studies (e.g., from climate models, scenarios, and internal variability).

©2015. American Geophysical Union. All Rights Reserved.

The SD relationship is built with present-day observations and generally applied to GCM projections, assuming implicitly the temporal transferability of the SD relationship to the future climate. This transferability assumption, common to all downscaling methods to some extent, is the main theoretical weakness of SD [*Wilby et al.*, 1998; *Schmith*, 2008]. To be valid, it requires that all the mechanisms that impact precipitation changes in the future climate are captured by the SD method. A first evaluation of SD methods generally consists in downscaling an atmospheric reanalysis to assess the ability of the method to capture correctly the observed present-day variability [*Timbal et al.*, 2003; *Radanovics et al.*, 2013]. The implicit rationale is that a method that is not able to capture current climate variability will be unable to correctly capture future changes, and therefore will not be transferable to the future climate.

A first step to better assess the temporal transferability of the SD method could be to evaluate its ability to reproduce observed low-frequency variations. If the observed trends are large, and especially if they are at least partially caused by anthropogenic forcing, it can be a good test of the ability of the SD method to capture contrasted climate states [*Maurer and Hidalgo*, 2008]. Limited magnitude or temporal inhomogeneities in observed trends in predictors from reanalyses may however be problematic in practice. Another approach is proposed by *Gutiérrez et al.* [2013]. They build their SD relationship including only the coldest observed years in the training period and then assess its transferability to the warmest observed years. The observed differences in temperature between the coldest and warmest years in the historical period are still however generally lower than the expected future warming. Additionally, the physical processes involved in the context of interannual variability and anthropogenic climate change may be different.

The consistency between statistical downscaling and dynamical downscaling or direct GCM results in the future climate is also an interesting clue on the transferability of the SD method to future climate conditions [*Spak et al.*, 2007; *Timbal et al.*, 2008; *Boé et al.*, 2009], but in case of divergence, it is impossible to know whether the error lies in the SD method or not.

As it is impossible to evaluate the capacity of the SD method to capture the future climate change signal based on observations, some authors have tried to evaluate the capacity of their method to reproduce future climate change using climate model results as pseudo-observations [*Vrac and Stein*, 2007; *Frias and Zorita*, 2006; *Beuchat et al.*, 2012]. In this framework, the statistical downscaling method is built using present-day model results instead of real observations. Then, the corresponding future climate simulation is downscaled and results from downscaling can be compared to the results directly simulated by the model. This approach belongs to the family of perfect model frameworks whereby a model is considered to be the truth and therefore its results perfect, in order to test some methodological hypotheses when the necessary observations do not exist (e.g., in the future climate). Although still rare in the SD studies, perfect model frameworks are more and more used in different fields of climate science. For example, in paleoclimatology, approaches similar to perfect model frameworks are commonly used to evaluate paleoclimate reconstruction methodologies through pseudoproxy experiments (see *Smerdon* [2012], for a review). Potential decadal climate predictability is generally assessed within a perfect model framework both for dynamical predictions, e.g., *Boer* [2004] and *Collins et al.* [2006] or statistical predictions [*Hawkins et al.*, 2011]. *Ribes et al.* [2013] use a perfect model framework in the context of detection and attribution of climate change.

In the SD context *Vrac and Stein* [2007] and *Beuchat et al.* [2012] use a GCM (RCM) to provide the low-resolution predictors (high-resolution predictands) considered as perfect pseudo-observations to develop the SD method in the present climate. The SD method is then applied to the GCM projection. Downscaled precipitation changes are finally compared to precipitation changes directly simulated by the RCM. Because of potential inconsistencies between the GCM and the RCM, one cannot necessarily expect the SD of the GCM and the RCM to provide exactly the same results, and therefore potential differences cannot be strictly attributed to the SD method. *Frias and Zorita* [2006] develop their SD method using results of a GCM as pseudo-observations for predictors and predictands and compare the future changes in downscaled precipitation to those directly simulated by the GCM. A potential limitation of this approach is the poor representation of precipitation in low-resolution GCMs.

A perfect model framework for SD simply allows testing whether the main mechanisms that control future precipitation changes in a given climate model are captured by the SD method. As the mechanisms leading to precipitation changes in a given model are not necessarily realistic, results of the perfect model

©2015. American Geophysical Union. All Rights Reserved.

framework do not necessarily apply to the real world. As in every study based on climate models, we have to rely on the hypothesis that despite some potential issues in individual climate models, collectively, they are able to capture correctly the most important features of future climate change. It is crucial to apply the perfect model framework to an ensemble of models rather than a single one in order to obtain robust results. Even if transferability to the future climate in the perfect model framework does not guarantee transferability in the real world, it is clear that between two SD methods with similar performances in the present climate when evaluated against observations, much less confidence should be given to a method that does not perform well in the perfect model framework.

A perfect model framework using multiple RCMs is developed in this study. Using RCMs allows the ability of the SD method in capturing high-resolution precipitation changes to be tested.

As the transferability of a SD method ultimately depends on the processes that it captures and therefore on the predictors, in this study, new predictors compared to previous work [e.g., *Boé and Terray*, 2007] are tested. The objective is to better capture the impact of atmospheric moisture and local thermodynamic processes (e.g., change in atmospheric stability) in precipitation changes. The predictors related to these processes might not be as well constrained in reanalyses as predictors such as sea level pressure more traditionally used in downscaling methods. Therefore, their representation in reanalyses might be more uncertain. As additionally, uncertainties in downscaling results associated with reanalyses have been seldom evaluated, even for classical predictors, our SD method will be tested using predictors from four atmospheric reanalyses.

After a description of the data sets and models used, the SD method developed in this study is introduced (section 2). Four reanalyses are then downscaled with this method using different combinations of predictors, and the results are evaluated against present-day observations, in terms of climatology, interannual variability, and trends (section 3). The temporal transferability of the method to the future climate is then assessed using a perfect model framework with a large ensemble of RCMs, and the best set of predictors is selected (section 4). The same RCMs are finally downscaled as for real case applications, and the different uncertainties are quantified in section 5.

2. Data and Downscaling Method

2.1. Data

This study is focused on the spatial statistical downscaling of daily precipitation over France. Precipitation comes from the near-surface analysis Safran [*Quintana-Seguí*, 2008; *Vidal et al.*, 2010]. Safran uses surface observations collected by Météo-France and is based on an optimal interpolation algorithm with first guess fields coming from an atmospheric reanalysis. Safran provides the near-surface atmospheric variables at the hourly time step on an 8 km grid over France necessary to force the Isba-Modcou hydrometeorological system [*Habets et al.*, 2008]. Based on *Quintana-Seguí* [2008] and *Vidal et al.* [2010], potential errors in Safran precipitation are assumed to be negligible compared to the other uncertainties involved in this work.

Low-resolution predictors necessary for SD come from four atmospheric reanalyses. The National Centers for Environmental Prediction - National Center for Atmospheric Research reanalysis (NCEP) [*Kalnay and Kanamitsu*, 1996] is a previous generation reanalysis that has widely been used in the climate community, including many SD studies. The atmospheric model has a horizontal resolution of about 200 km with 28 vertical levels. Data are available from 1948 to present on a 2.5 ° grid. In the Twentieth Century Reanalysis (20CR) [*Compo et al.*, 2011], only surface pressure is assimilated allowing the reanalysis to be conducted on a much longer period (from 1871 to 2010) compared to classical reanalyses that assimilate much more data (other surface variables, radiosondes, and satellite observations).

Two new generation reanalyses are also used in this study. The European Center for Medium-Range Weather Forecasts (ECMWF) reanalysis-Interim (ERA-Interim) [*Dee et al.*, 2011] is based on an atmospheric model with a horizontal resolution of about 79 km and 60 vertical levels. ERA-Interim covers the period from 1979 to present. Predictors used in this work have a horizontal resolution of 1.5 °. The Modern-Era Retrospective Analysis for Research and Applications (MERRA) [*Rienecker and Suarez*, 2011] uses the National Aeronautics and Space Administration (NASA) global data assimilation system. A major goal of MERRA was to achieve significant improvements in precipitation and water vapor climatology compared to earlier reanalyses [*Rienecker and Suarez*, 2011]. MERRA data are available on the native grid of the model with a resolution of $^1/_2\,° \times {}^2/_3\,°$. MERRA covers the same period as ERA-Interim. It will be interesting to see whether MERRA and

 ©2015. American Geophysical Union. All Rights Reserved.

Table 1. Regional Climate Simulations From the ENSEMBLES Project [*van der Linden et al.*, 2009] Used[a]

	BCM	CGCM3	CNRM	METO-Medium	METO-High	MPI
C4I					X	
CNRM			X			
DMI	X					X
ETHZ				X		
KNMI						X
Met.No	X			X		
METO				X	X	
MPI						X
OURANOS		X				

[a]The RCMs are given in the rows, and the GCMs that provide the boundary conditions are given in the columns.

ERAI lead to substantial improvements of SD results compared to a previous generation reanalysis as NCEP and how 20CR compares to the other systems that assimilate a much larger variety of observations.

Results from 12 regional climate simulations over Europe at a 25 km resolution from the ENSEMBLES project [*van der Linden et al.*, 2009] on the 1961–2050 period are used in this study. ENSEMBLES RCMs are forced on the historical period by observed greenhouse gases concentration and after 2000 by the special report on emissions scenarios (SRES) A1B scenario [*Nakicenovic et al.*, 2006]. Table 1 details the RCMs used in this study and the GCMs that provide their respective boundary conditions. Subsequent statistical tests are based on the hypothesis that the simulations are independent.

2.2. Analog Method

Statistical downscaling is based on the idea that the local climate variables depend on the large-scale climatic state and local features such as orography. Following this idea, an empirical statistical relationship between the high-resolution climate variable(s) of interest (or predictands, precipitation in our study) and the relevant low-resolution variables (or predictors) is developed with observations on a so-called training period. The statistical relationship can then be used to derive the high-resolution predictands on any period for which low-resolution predictors, for example, projected by a GCM, are available. A recent review of statistical downscaling methods for precipitation in the climate change context is given by *Maraun et al.* [2010].

The analog method [*Lorenz*, 1969; *Zorita et al.*, 1995] is used in this study. It has been widely used, for example, for weather forecasting [*Obled*, 2002; *Horton et al.*, 2012] but also in SD context [*Timbal and McAvaney*, 2001] or even for decadal prediction [*Hawkins et al.*, 2011]. On the training period, each observed predictor state at a time step t ($Po_t(t)$) is chronologically associated with a corresponding observed high-resolution predictand state ($Pa_t(t)$). Then, knowing the predictor state in a climate simulation to be downscaled at a given time step t_2 ($Po_s(t_2)$), the most similar observed predictor state in the training period is searched ($Po_t(t_3)$). Finally, the associated predictand state ($Pa_t(t_3)$) is selected as an estimation of the high-resolution predictand state corresponding to $Po_s(t_2)\ :\ Pa_s(t_2) \simeq Pa_t(t_3)$. The main methodological choices of the analog method consist in the selection of the predictors and the definition of a criterion to measure the similarity of predictor states. Its main limitation is its inability to produce a local state that has never been observed in the training period. As our impact study will end in the mid-21st century and because of the large variance of daily precipitation, it is assumed that this issue will have a limited impact.

The similarity criterion to define the analog is the Euclidian distance (as in *Ben Daoud et al.* [2011]) for all the predictors, except for sea level pressure for which the Teweles and Wobus score (TWS) is used [*Guilbaud and Obled*, 1998; *Obled*, 2002]. TWS measures the similarity in terms of horizontal pressure gradients, which are directly associated with the geostrophic wind strength. Preliminary tests showed that for the analog method tested in this study, the interannual variability of precipitation over France is generally better reproduced with TWS for sea level pressure (SLP) than with the Euclidian distance (not shown). The annual climatological mean over the 1979–2010 period is removed before calculating the distances.

A relatively large domain is used for the calculation of the similarity criterion for SLP. For the other predictors, the domain is smaller and centered on France (Figure 1). Tests with domains of different sizes have been

©2015. American Geophysical Union. All Rights Reserved.

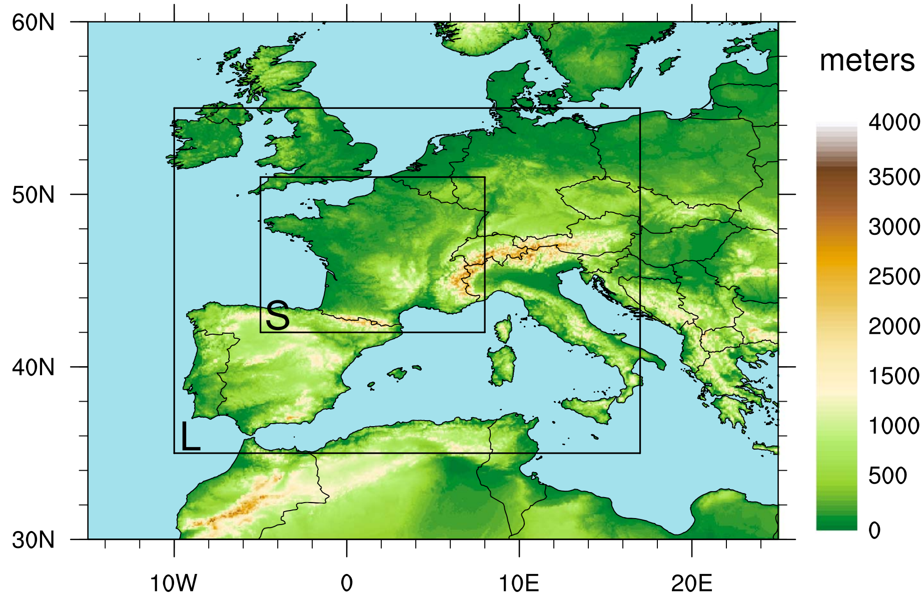

Figure 1. Analogy domains used for the atmospheric predictors and orography (in colors, meters). The domain (L) is used for sea level pressure and the domain (S) for the other predictors.

made for the choice of the analogy domains, by comparing the downscaling results obtained with the different domains, based on interannual correlations and seasonal biases (not shown). No systematic approach to explore all the possible domains has been followed as in *Radanovics et al.* [2013], and the chosen domains therefore cannot be said to be optimal. The search for the analog is restricted to the same season as the target date. A moving window of ±30 days is used in this study, as in *Ben Daoud et al.* [2011]. The most similar analog is then simply found by minimizing the sum of the standardized distances corresponding to each predictor.

2.3. Predictors

Different combinations (Table 2) based on seven predictors are tested: precipitation (PRCP), the sea level pressure (SLP), the 2 m air temperature (TAS), the lifted condensation level (LCL), the specific humidity at 850 hPa (HUS), the module of moisture flux at 850 hPa (QFX), and the totals total index (TTI).

The most basic combination of predictors is based on SLP, a predictor of large-scale circulation, and TAS used in this context as an indicator of specific humidity at saturation, as in *Boé et al.* [2009].

As humidity may be a limiting factor for precipitations [*Trenberth*, 1998], its changes may play an important role in precipitation changes. Globally, the change in specific humidity is expected to be roughly equal to the change in specific humidity at saturation and therefore almost proportional to temperature change

Table 2. Combination of Predictors Tested for the Statistical Downscaling Method in This Study[a]

	PRCP	SLP	TAS	LCL	QFX	TTI	HUS
C1	X						
C2		X	X				
C3		X	X	X			
C4		X	X		X		
C5		X	X		X	X	
C6		X	X		X	X	X

[a]PRCP: precipitation. SLP: sea level pressure. TAS: 2 m air temperature. LCL: lifted condensation level. QFX: the module of moisture flux at 850 hPa. TTI : totals total index. HUS : specific humidity at 850 hPa.

©2015. American Geophysical Union. All Rights Reserved.

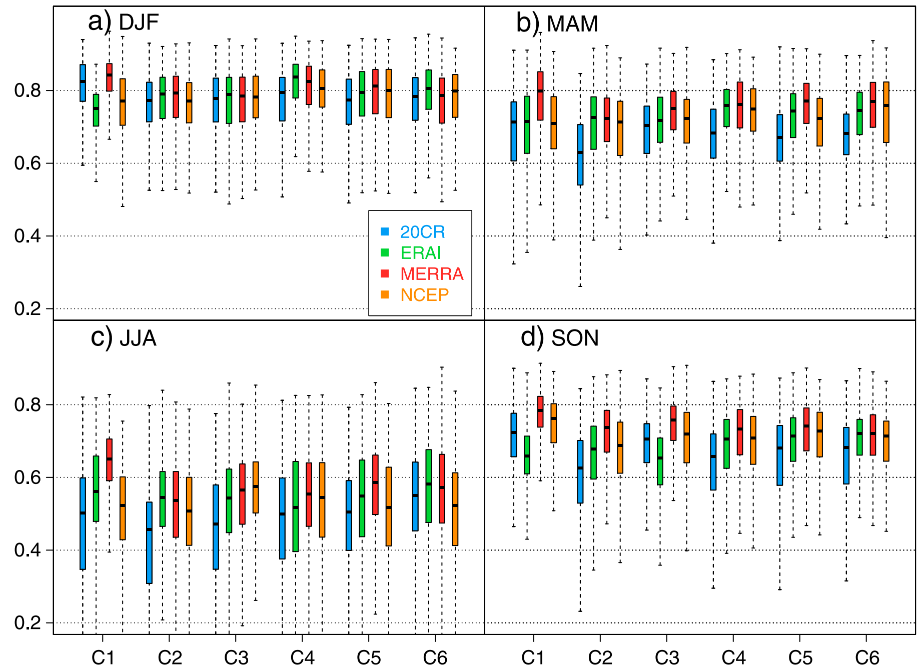

Figure 2. (a–d) Interannual correlation between seasonal downscaled precipitation and observed (Safran data set) precipitation for six combinations of predictors from four reanalyses. The box plot represents the spatial distribution of the biases over France. The lines show the 25th, 50th, and 75th quartiles. The whiskers extend to either 1.75 times the interquartile range or the maximum (minimum).

[*Trenberth et al.*, 2003], but it is not true regionally, and, in particular, over Europe and France, as the summer relative humidity decreases there in many models [*Sherwood et al.*, 2010]. Therefore, knowing the change in temperature is not sufficient to fully anticipate the change in specific humidity, hence the use of HUS as predictor.

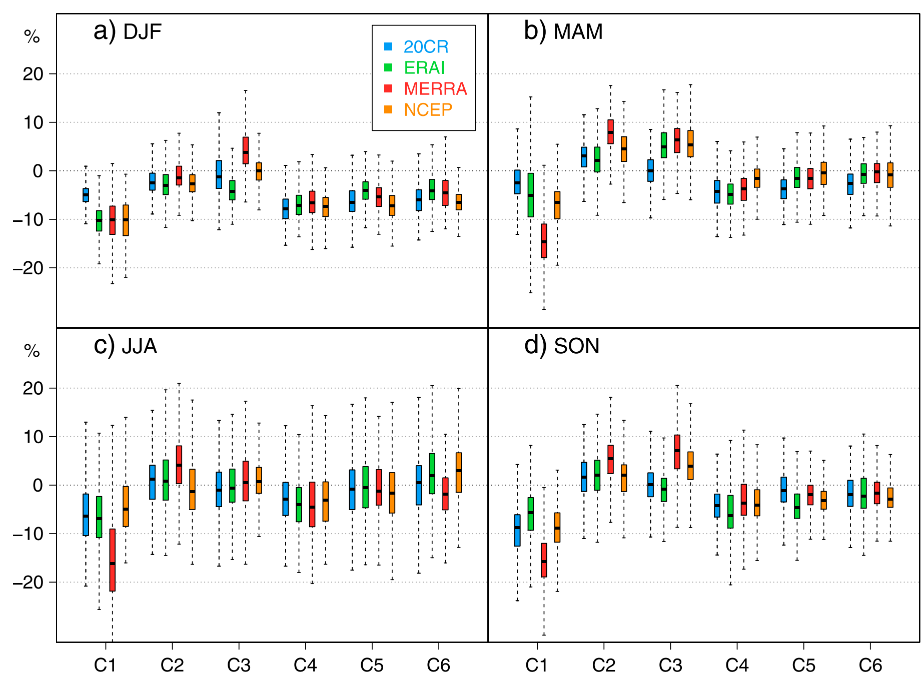

Figure 3. Same as Figure 2 for seasonal relative biases in downscaled precipitation.

©2015. American Geophysical Union. All Rights Reserved.

This decrease in relative humidity may have impacts on precipitation on its own, as condensation becomes more difficult to reach. For example, the decrease in relative humidity leads to a rise of the height to which an air parcel has to be lifted adiabatically to reach saturation (or LCL). The higher the LCL is, the more difficult it is to reach condensation and thus the more unlikely precipitations are. As a result, a strong link exists over France between precipitation and the LCL in summer [*Boé*, 2012]. The LCL is estimated here as being proportional to the surface dew point depression, following *Lawrence* [2005]. The link between LCL and precipitation is expected to hold in other areas, but its potential importance in the climate change context is expected to depend on the future evolution of relative humidity.

Even if the moisture flux only depends on large-scale circulation and atmospheric humidity, which are taken into account by other predictors, the moisture flux at 850 hPa is also included in some combinations of predictors.

A simple indicator of atmospheric stability, the TTI, is also tested. Atmospheric stability is indeed particularly important for precipitation, especially in summer over Europe, because of the more convective nature of precipitation [*Trenberth*, 1999]. The higher the TTI is, the greater the atmospheric instability is and as a result, the most likely thunderstorms are. *Boé* [2012] has shown a strong link between precipitations over France in summer and the TTI. Moreover, because of changes in the vertical temperature and humidity profiles with climate change, the TTI is expected to evolve in the future climate. The TTI is defined as follows: $TTI = (Ta_{850} - Ta_{500}) + (Td_{850} - Ta_{500})$.

Ta_{500} and Ta_{850} are respectively the air temperature at 500 hPa and 850 hPa. Td_{850} is the dew point temperature at 850 hPa.

Finally, precipitation (PRCP) is also tested as a potential obvious predictor for local-scale precipitation [*Maurer and Hidalgo*, 2008]). However, precipitation biases are generally particularly important in reanalyses [*Kalnay and Kanamitsu*, 1996; *Simmons et al.*, 2010] and climate models [*Boberg et al.*, 2009; *Kelley et al.*, 2012], which may be problematic in the statistical downscaling context.

3. Evaluation of the SD Method With Present-Day Observations

3.1. Climatology and Interannual Variability

A first step to evaluate the SD method is generally to downscale an atmospheric reanalysis and to compare the downscaled variable (precipitation in our case) to observations, in terms of temporal sequence (e.g., temporal correlations) and basic statistical properties (e.g., climatological means).

As reanalyses do not provide a perfect estimate of the atmospheric state, it is in fact not simply the SD method that is evaluated but a combination of the SD method and a reanalysis. In order to evaluate the impact of reanalyses in this context, four reanalyses (NCEP, 20CR, MERRA and ERAI) are downscaled on the training period (1971–2010) with a pool permutation technique as in *Timbal et al.* [2003]. Analogs are searched excluding the current year in order to avoid artificial skill due to temporal autocorrelation. Downscaled precipitations are then compared to Safran.

Consistently with previous studies [e.g., *Radanovics et al.*, 2013; *Lafaysse and Hingray*, 2014], the highest interannual correlations between downscaled and observed precipitations are obtained in winter (median correlation over France close to 0.8, Figure 2a) and the lowest in summer (median correlation close to 0.6, Figure 2c), independently of the predictors and reanalyses used. In France, precipitation is predominantly driven by large-scale flow in winter [*Plaut et al.*, 2001; *Sanchez-Gomez and Terray*, 2005; *Boé and Terray*, 2008] and by local convection in summer, the former being more easily captured by statistical downscaling than the latter. The spatial spread in correlations is also much larger in summer than in winter.

In terms of median over France, the highest correlations are always obtained with precipitation as predictor (C1) for MERRA (Figure 2). It is not the case for the other reanalyses. For example, C1 never leads to the highest correlations for ERAI, and only in autumn for NCEP. Precipitations in reanalyses are therefore not necessarily the best predictor for local precipitation. It is consistent with the fact that precipitation in reanalyses is generally poorly constrained by observations and strongly depends on model physics [*Rienecker and Suarez*, 2011].

For the discussion of the correlations obtained with the other combinations of predictors, C2 (SLP and TAS as predictors) is considered as the reference, as the other combinations are derived from it by addition of

©2015. American Geophysical Union. All Rights Reserved.

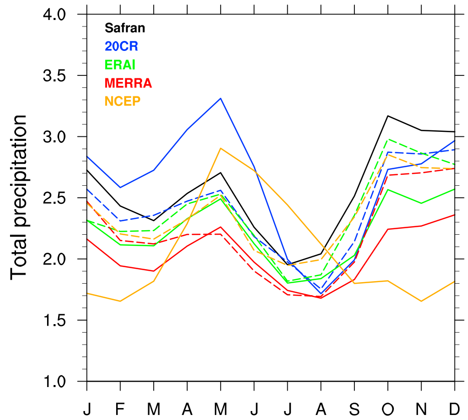

Figure 4. Seasonal cycle of precipitation over France on the 1979–2010 period (in mm/day) in observation (solid black line) and in reanalyses (solid-colored lines). Dash-colored lines are seasonal cycles obtained with downscaled precipitation from reanalyses with combination C1.

predictors. In winter, the addition of moisture flux (C4) leads to the largest increase in correlations relatively to C2 (Figure 2a), independently of the reanalysis. In summer, the addition of LCL as predictor (C3) leads to the largest improvements in correlations for ERAI, and NCEP to a lesser extent, while TTI (C5) leads to the largest increase in correlation for MERRA (Figure 2c). This result is consistent with the seasonal variations of the dominant processes leading to precipitation over France. An impact of LCL and/or TTI is also noted for ERAI and MERRA in spring and autumn (Figures 2b and 2d). The impact of TTI and LCL is much more limited for 20CR, but as 20CR only assimilates surface pressure, TTI and LCL are not expected to be well constrained in this reanalysis.

It is important to note that the differences in correlations associated with the different combinations of predictors are generally small. Moreover, the results obtained with the different reanalyses are not necessarily consistent. For example, in summer, C6 generally outperforms C5 for ERAI but not for MERRA (Figure 2c). The interreanalyses differences in correlations for a given combination of predictors are in the end often greater than the differences due to predictors for a given reanalysis. It makes it difficult to choose a specific combination of predictors in a robust way. With regard to reanalyses, overall, the best performances are often obtained with the new generation reanalyses MERRA and ERAI. At the exception of winter, when 20CR is on par with the other reanalyses, 20CR generally leads to smaller correlations, when local processes, not as well constrained in 20CR, are important for precipitation. Yet, given the breadth of data assimilated by the other reanalyses compared to 20CR, the relatively good performance obtained with 20CR is noteworthy, as this reanalysis has the advantage to be available on a much longer period.

Regarding climatological means (Figure 3), precipitation as predictor leads to important biases, especially with MERRA (−15% in summer and −10% in winter). It is interesting to note that the best performances correlation-wise were seen with MERRA and precipitation as predictor (Figure 2). Most reanalyses show large biases in precipitation over France, with MERRA and NCEP generally showing the largest biases (Figure 4). Note that in summer, when the bias in precipitation in MERRA is especially large, the bias in downscaled precipitation from MERRA is also especially large. Not surprisingly, the SD method with precipitation as predictor cannot fully overcome the biases in reanalysis precipitation. Biases in downscaled precipitation obtained with the other combinations of predictors are generally much smaller. Overall, taking into account the four seasons, the lower biases are obtained with the C5 (SLP, TAS, QFX, and TTI) and C6 (SLP, TAS, QFX, TTI, and HUS) combinations. With ERAI, for example, the median biases over France are always smaller than 5% for C6 and C5. The combinations C2 (SLP and TAS) and C3 (SLP, TAS, and LCL) lead to limited biases in summer and winter but to larger ones in intermediate seasons.

Given the fact that precipitation as predictor (C1) does not necessarily lead to higher correlations between downscaled and observed precipitation, and leads to serious climatological biases, this combination is not investigated further in the manuscript. The addition of TTI, LCL, and QFX to the reference combination based on SLP and TAS generally leads to improvements in terms of interannual correlation and biases. The improvements however remain rather limited, and it is therefore impossible to discard any combinations of predictors except C1 simply based on the evaluation of present-day interannual correlations and biases.

Many SD methods for precipitation over France have been developed, based on different approaches: e.g., weather typing [*Boé et al.*, 2009], analog methods [*Obled*, 2002; *Radanovics et al.*, 2013; *Chardon et al.*, 2014], or a combination of a weather generator and *k*-nearest neighbor approach [*Mezghani and Hingray*, 2009]. A large variety of predictors has been used in those SD methods, most of studies using large-scale circulation variables as main predictor (sea level pressure or geopotential height, possibly at different levels, or wind) and additional variables such as temperature [e.g., *Boé et al.*, 2009] and/or variables related to

©2015. American Geophysical Union. All Rights Reserved.

humidity [e.g., *Mezghani and Hingray*, 2009; *Radanovics et al.*, 2013]. The results of the different studies are not easy to compare, because some of them are focused on specific regions while others deal with the entire France, and the skill scores used to evaluate the methods vary (here we use interannual correlation while other studies as *Radanovics et al.* [2013] and *Chardon et al.* [2014] use continuous rank probability score). *Lafaysse and Hingray* [2014] provide an interesting intercomparison of the SD methods (or variants with different predictors) described in *Obled* [2002], *Boé et al.* [2009], and *Mezghani and Hingray* [2009] in the Durance basin. They obtain interannual correlations between 0.7 and 0.9 in winter and between 0.2 and 0.6 in summer depending on the methods and locations in summer, which is similar to what we obtain in the Durance region (not shown), even if our method is not optimized for this specific region. Our interpretation of the current literature and of the results described in this section is therefore that the skill of current SD techniques over France is rather similar when evaluated against present-day observations in terms of climatology and daily to interannual variability, despite the variety of methods and predictors used. Even if the statistical method, the choice of the domain, and the predictors, can lead to slight improvements, no major differences in performance are to be expected.

3.2. Trends

It has been shown that the SD method is able to capture between 25% and 65% of the interannual variance in precipitation depending on predictors and seasons. It does not imply that they are suitable to capture lower frequency variations in precipitation and most importantly the ones potentially caused by global warming [*Maraun et al.*, 2010], as the processes involved or their respective importance may be different. This is why SD methods are also commonly evaluated on their capacity to reproduce low-frequency-observed variations and in particular trends [*Maurer and Hidalgo*, 2008].

Seasonal trends are calculated over the training period (1971–2010) on observed and downscaled precipitation. The statistical significance of observed trends is tested with the nonparametric Mann-Kendall test [*Kendall and Gobbons*, 1990]. The potential autocorrelation of precipitation is taken into account in the test following *Hamed and Ramachandra Rao* [1998].

Trends in downscaled precipitation with the C6 combination of predictors for the four reanalyses are compared to observed trends. The across-reanalysis minimum and maximum at each point of downscaled trends are shown in Figure 5. Negative trends are observed over the Alps and south western France in winter, over the Alps in autumn, and in northern France in spring. Positive trends are observed in southeastern France and over central France in summer.

The spread in downscaled trends due to reanalyses is generally very large. The minimum-maximum range is for example greater than 1.2 mm/day (a value seldom seen for observed trends) in winter around the Mediterranean basin and in spring in southern France. A minimum-maximum range as large as 2.5 mm/day is noted over southeastern France in autumn. Note that the large spread in downscaled precipitation cannot be attributed to a single outlier reanalysis (not shown). The same analysis was done for the C2, C3, C4, and C5 combinations of predictors and similar results were obtained (not shown) except for C3. The large across-reanalysis spread seen in downscaled precipitation is likely due to temporal inhomogeneities and errors that may exist in reanalyses, for example, caused by the introduction of new instruments [*Dee et al.*, 2011; *Rienecker and Suarez*, 2011] or changes in the observations network [*Krueger et al.*, 2013]. Note that consistently with that idea, a large spread also exists in precipitation trends in reanalyses (not shown).

As hoped for, the observed trends generally fall within the minimum-maximum range of downscaled trends. It is the case here for the C6 combination and the other combinations tested (not shown), except C3 (PSL, TAS, and LCL). Indeed for C3, the observed trends are clearly outside the across-reanalysis range of downscaled precipitation trends throughout the country in summer (Figure 6 compared to Figure 5k).

Except for northwestern France and the Mediterranean coast, where observations are close to the maximum downscaled trends, trends obtained with C3 are more negative than observed ones, independently of the reanalysis. In particular, maximum downscaled trends are negative in central France (between −0.2 and −0.6 mm/day), whereas they are positive in Safran (between 0.2 and 0.6 mm/day). These spurious negative trends can be attributed to the use of LCL as predictor as they are not seen with C2 (SLP and TAS as predictors, not shown), which is similar to C3 except that LCL is not included as predictor. A quick visual inspection of LCL in reanalyses and of the variables used to compute it, suggests that it could be due to

©2015. American Geophysical Union. All Rights Reserved.

10.1002/2014JD022236

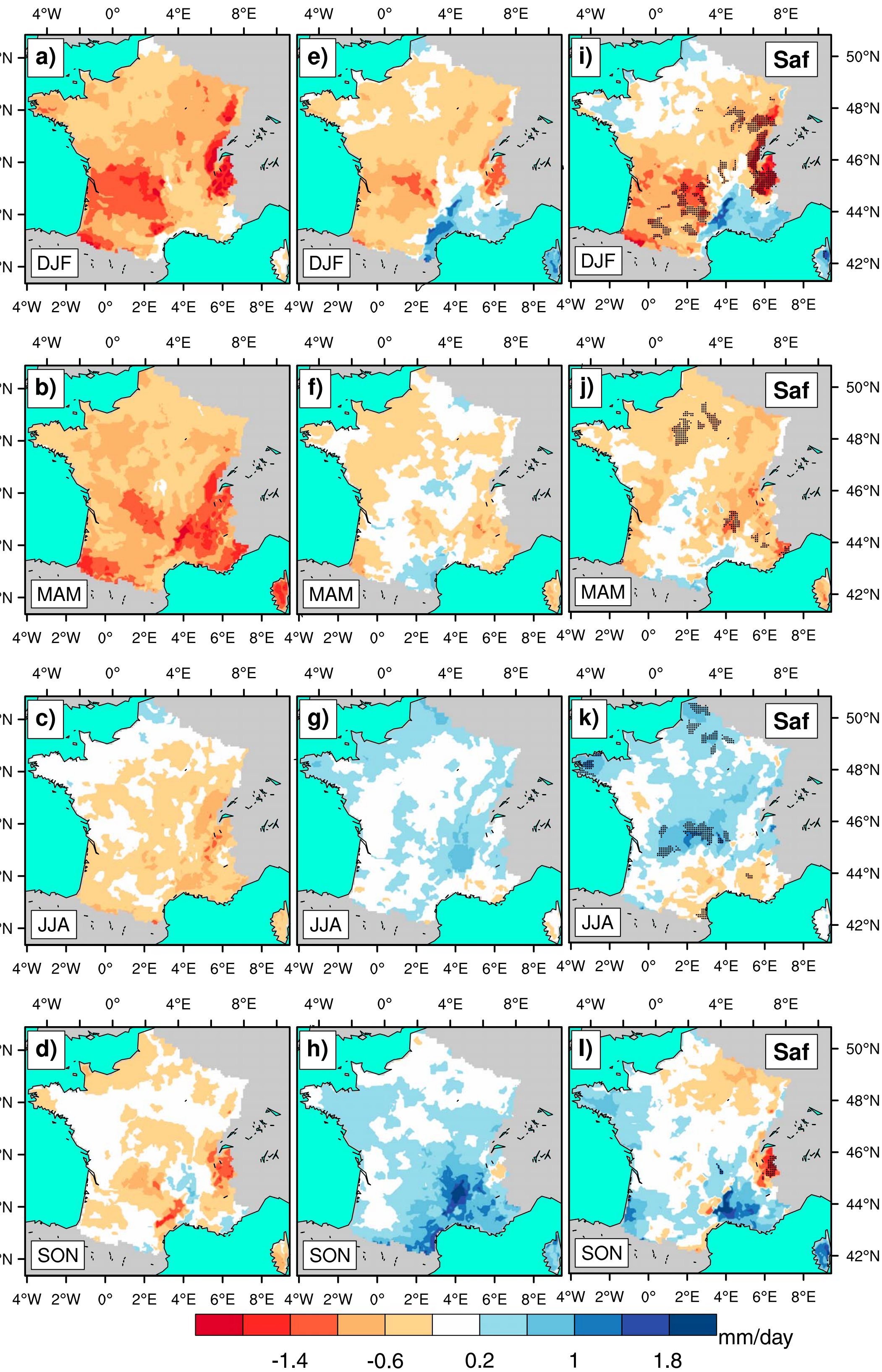

Figure 5. Linear trends in seasonal precipitation over the 1979–2010 period (in mm/day). Four reanalyses have been downscaled with the C6 combination of predictors. (a–d) Across-reanalysis minimum downscaled trends. (e–h) Across-reanalysis maximum downscaled trends. (i–l) Observed trends (Safran data set). Dots show where observed trends are significant with $p < 0.05$ following the Mann Kendall test.

©2015. American Geophysical Union. All Rights Reserved.

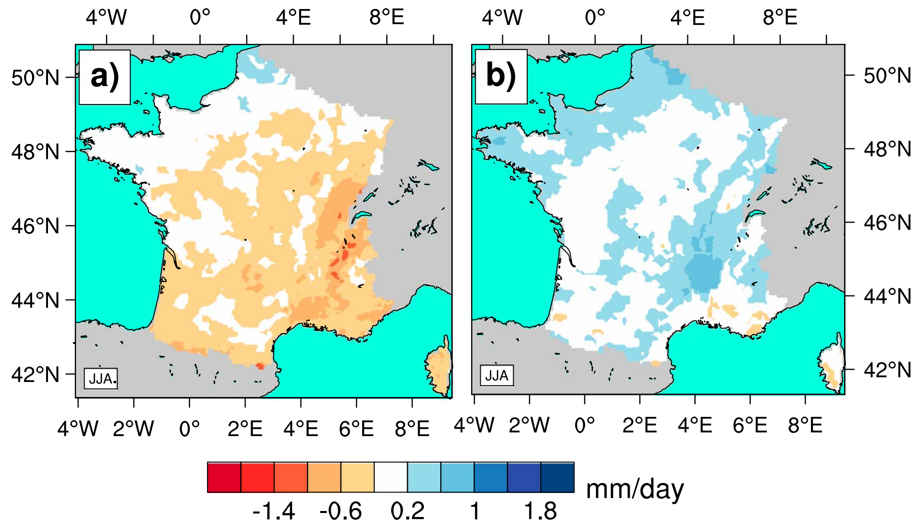

Figure 6. Linear trends in summer precipitation on the 1979–2010 period (in mm/day). Four reanalyses have been downscaled with the C3 combination of predictors. (a) Across-reanalysis minimum downscaled trends. (b) Across-reanalysis maximum downscaled trends.

temporal inhomogeneities in surface relative humidity (not shown). LCL and therefore the C3 combination will not be used subsequently. Note, however, that with a proper representation of LCL in a future reanalysis, LCL could become an interesting predictor for summer precipitation downscaling.

The ability of the SD method to capture observed trends is often seen as an interesting indicator of the potential transferability of the SD method under climate changes. However, in our application, the observed trends in precipitation are weak and seldom significant and a large spread in downscaled precipitation trends exists because of reanalyses. Given this large spread due to reanalyses, to go further, it would be necessary to be able to separate the error in downscaled trends that is due to the SD method from the error that is due to errors in the predictors from reanalyses, but it is not possible, as observations to evaluate most of the predictors do no exist. In the end, no strong conclusion can be reached. Except for C3, the others combinations of predictors tested cannot be discriminated and their temporal transferability has not been demonstrated.

In this section, the evaluation against present-day observations of the downscaling results obtained with different sets of predictors has led to the rejection of two sets: C1 that only uses precipitation as predictor because it leads to large climatological biases and C3 that includes LCL because it leads to too unrealistic trends in summer.

4. Evaluation of the SD Method in the Future Climate Through a Perfect Model Framework

In this section, a perfect model framework is introduced to assess the transferability to the future climate of the SD method. In this framework, as a first step, the SD relationship is developed within the model world. Predictors and precipitation come from a RCM simulation, on the same training period used in the previous sections (1979–2010). Then, the entire regional climate projection (1961–2050) is downscaled based on the SD relationship between RCM predictors and RCM precipitation. Finally, it is possible to compare changes in downscaled precipitation to the changes directly simulated by the RCM, at their native resolution of 25 km, to assess the capacity of the SD method to correctly reproduce future precipitation changes and therefore the temporal transferability of the SD relationship within the RCM world. In order for the results of our study to be robust and not dependent of the specificities of a single regional model, our perfect model framework is applied to an ensemble of 12 RCMs (see Table 1). The term "downscaling" is somewhat unsuitable in the perfect model framework as predictors and precipitation are in fact at the same resolution (25 km). Note also that in real case downscaling, with the statistical relationship built between reanalyses and Safran precipitation, the resolution of downscaled precipitation is 8 km and therefore finer than the one in the

 ©2015. American Geophysical Union. All Rights Reserved.

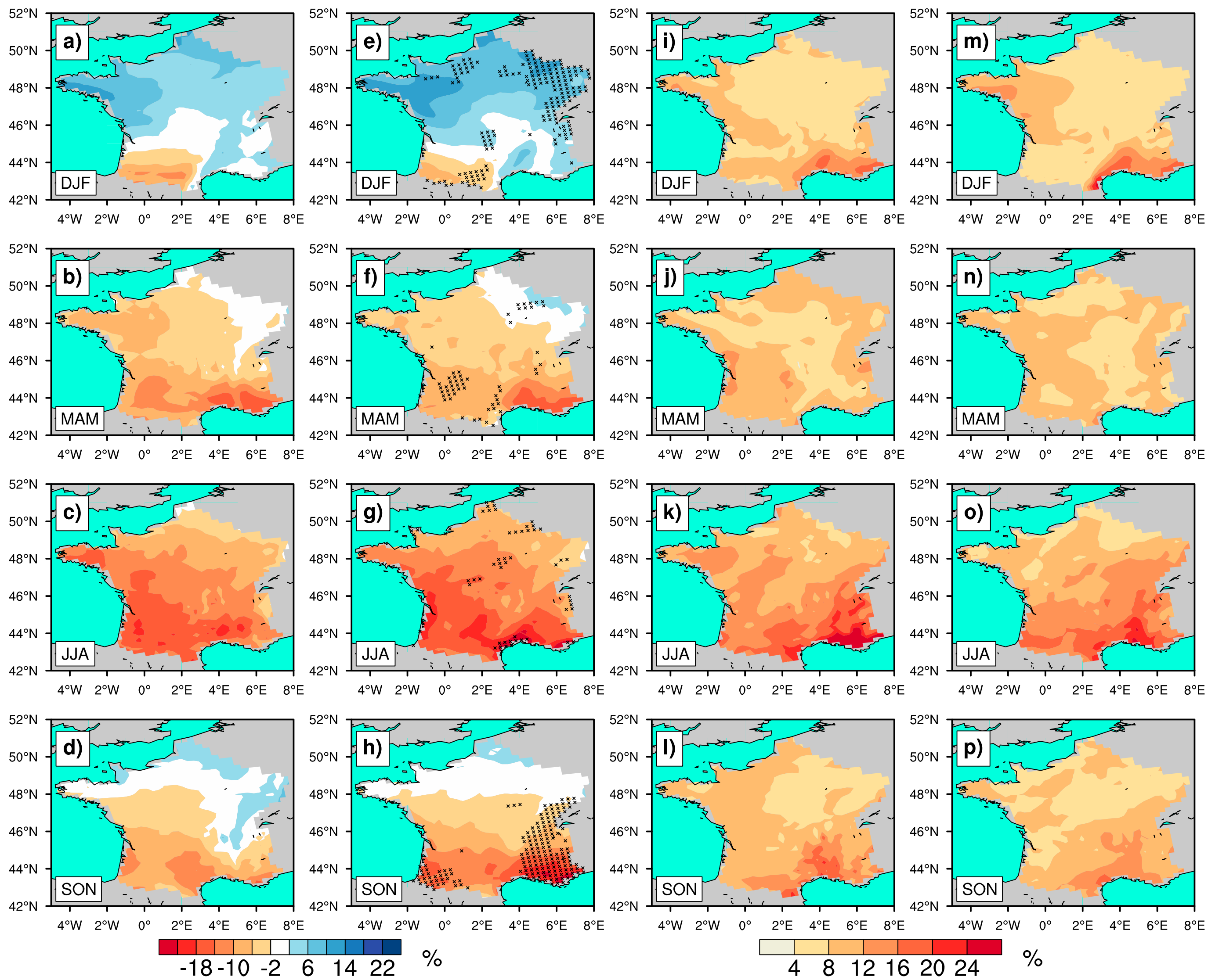

Figure 7. Seasonal relative precipitation changes (%) between the 1961–1990 and 2030–2050 periods. Ensemble means: (a–d) simulated, (e–h) obtained by statistical downscaling within the perfect models framework using the C6 combination of predictors. Ensemble standard deviation across models: (i–l) simulated, (m–p) obtained by statistical downscaling as in Figures 7e–7h. Dots show where the differences between simulated and downscaled changes following a paired *t* test are significant with $p < 0.05$.

perfect model framework. However, as the difference in resolution (25 km versus 8 km) is limited, we do not expect that it alters the relevance of the results obtained within the perfect model framework for real case downscaling. All the combinations of predictors that were not eliminated after the evaluation against present-day observations described in the previous sections (i.e., C2, C4, C5, and C6) have been tested in the perfect model framework. Results are only shown for the combination C6 (SLP, TAS, QFX, TTI, and HUS) in Figure 7, as it is the only combination that leads to changes in the perfect model framework sufficiently similar to those simulated by RCMs (not shown).

As an ensemble, the RCMs simulate an increase in precipitation of 10% in the northern half of the country in winter and drier conditions in southwestern France (Figure 7e). The intermodel standard deviation is especially large around the Mediterranean basin (around 15%, Figure 7m). The strongest changes are seen in summer with a large decrease in precipitation (up to 20%) over the entire France, with maximum values in the south (Figure 7g). The largest intermodel standard deviation is also noted in summer (Figure 7o) over the entire country and particularly in southeastern France.

©2015. American Geophysical Union. All Rights Reserved.

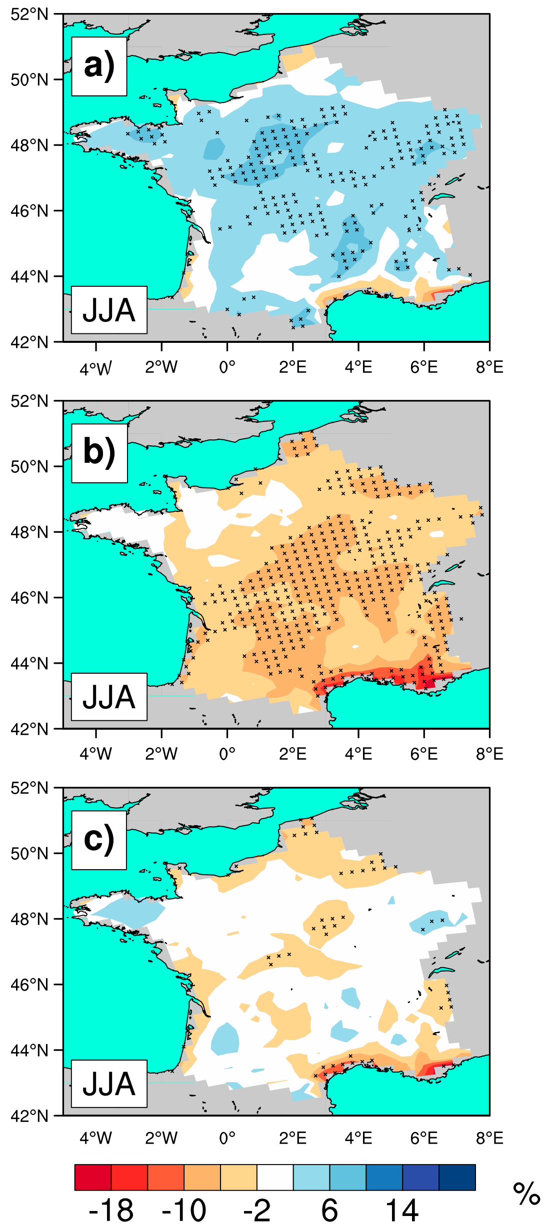

Figure 8. Ensemble mean differences (%) between relative precipitation changes in summer, as simulated and as obtained with the perfect model framework for three combinations of predictors. Dots show where the differences are significant following a paired *t* test at the 5% level. The combinations of predictors are (a) PSL-TAS-QFX-HUS; (b) PSL-TAS-QFX-TTI; and (c) PSL-TAS-QFX-TTI-HUS.

The ensemble mean changes given by statistical downscaling with C6 in the perfect model framework are displayed in Figures 7e–7h. The main features of simulated precipitation changes are well captured by the SD method. In all seasons, the pattern of simulated changes is very well reproduced by the SD method (spatial correlation: December-January-February: 0.94, March-April-May: 0.88, June-July-August: 0.78, and September-October-November: 0.83). The magnitude of changes is also generally well captured. Most notably, the strong decrease in summer precipitation is very similar in direct and downscaled model results. Changes in spring precipitation (Figures 7b and 7f) are also very well reproduced. The largest significant differences between downscaled and simulated precipitation changes are seen in autumn in the southeastern France. *Radanovics et al.* [2013] and *Chardon et al.* [2014] work suggests that the analogy domain used in our study is not optimal for this particular region, at least in the present-day climate context. Tests with analogy domains more adapted to this specific region did not lead to substantial improvements in downscaled precipitation changes (not shown). Some processes leading to precipitation changes in autumn in southeastern France may not be captured correctly by the predictors tested in our study.

Relatively large areas with significant differences between downscaled and simulated precipitation changes following a paired *t* test are also noted in winter in the east of the country, but these differences remain small relatively to the ensemble spread in precipitations changes (the differences are not significant with a standard *t* test), and in the end, have therefore little importance for practical applications.

The intermodel spread in downscaled precipitation changes (Figures 7m–7p) is very similar to the one in simulated precipitation changes (Figures 7i–7l), indicating that the SD method captures the complete range of RCM behaviors and leads to a reliable estimation of RCM simulation uncertainties, which is also important for practical applications.

To better understand which predictors are especially important to capture future precipitation change, tests have been done by removing successively individual predictors from C6. The TTI and the specific humidity at 850 hPa have been found to be important in that context. Without TTI as predictor in C6, the future decrease in precipitation simulated by the RCMs is clearly underestimated (Figure 8a). In all the RCMs analyzed in this study, the atmospheric stability in summer increases (i.e., the TTI decreases), which is consistent with a decrease in convective precipitation. The decrease in TTI is partly related to the decrease in the lapse rate.

©2015. American Geophysical Union. All Rights Reserved.

Without the specific humidity at 850 hPa as predictor in C6, the decrease in downscaled summer precipitation is greater than the decrease simulated by the RCMs (Figure 8b). Specific humidity can be a limiting factor for precipitation [*Trenberth*, 1999], and it is therefore not surprising to find such an impact. Indeed, an increase in specific humidity is robustly simulated by climate models and is related to the large increase in specific humidity at saturation with temperature (as given by the Clausius-Clapeyron relation). Even if locally, the change in specific humidity may be smaller than the increase in atmospheric water holding capacity (i.e., the relative humidity may decrease), as it is the case over France in many RCMs [e.g., *Boé and Terray*, 2014], specific humidity still increases, with a potential impact on precipitation.

Very interestingly, TTI and HUS as predictors were not associated with large differences in downscaling results during summer in the present climate (for example, compare the results of C6, C5, and C4 in Figure 2). These variables however lead to large differences in terms of future climate change signals, as shown with the perfect model framework. This is an illustration of the limitation of classical approaches for the evaluation of downscaling methods based only on present-day observations. It is also interesting to note that the lowest interannual present-day correlations between downscaled precipitation and observations were found in summer. Despite that, Figure 7 show that summer changes are especially well captured. It is another illustration that present-day performances are not necessarily a very good indicator of performances in the climate change context.

Given data availability at the beginning of the study, ENSEMBLES regional climate models driven by the SRESA1B scenario have been used for the analyses described in this section. The CORDEX regional climate simulations based on the new Radiative Concentration Pathway (RCPs) scenario [*Meinshausen et al.*, 2011] are now available. There is no specific reason to expect that the conclusions of this section would be much different with the new RCM projections.

5. Real Case Statistical Downscaling of Future Climate Projections

5.1. Mean Precipitation Changes

The evaluation of the different sets of predictors, against present-day observations (section 3) and within the perfect model framework (section 4), has led to the selection of a single combination of predictors, C6 (SLP, TAS, QFX, TTI, and HUS). In this section, the real case application of the SD method with C6 to downscale the ENSEMBLE RCMs is presented. The SD relationship is built with the four reanalyses as predictors and Safran for precipitation, as in section 3 and on the same training period. Then the 12 RCMs described in Table 1 are downscaled on the 1961–2050 period with C6.

It is interesting to compare real case downscaled precipitation changes to downscaled precipitation changes in the perfect model framework (section 4) and to directly simulated changes. Differences between real case and perfect model downscaled precipitation can be attributed to differences in the links on the learning period between predictors and precipitation in the models and in the observations, as future predictors are identical in the two cases. In the perfect model framework, the present-day links between predictors and precipitation come from the RCMs, and therefore might be biased.

The consistency between real case and perfect model precipitation changes would therefore greatly reinforce our confidence in the realism of simulated present-day links between predictors and precipitation, more generally in the RCMs and therefore in the relevance of the perfect model framework to the real world, and in the end, in the results of real case statistical downscaling.

Future precipitation changes between the 2030–2050 and 1961–1990 periods are plotted in Figure 9 and have to be compared to the changes depicted in Figure 7. Results obtained with ERAI as well the maximum and minimum at each point among the results obtained for the four reanalyses are shown.

Note that precipitation changes obtained with ERAI (Figures 9i–9l) are in good agreement with the across-reanalyses and across-RCMs ensemble mean of downscaled changes (not shown). The dipolar structure of winter precipitation changes in France is very consistent between downscaled and direct results (see Figures 7 and 9). In summer, the magnitude of real case downscaled changes is also in good agreement with simulated and perfect model changes. The greatest differences between simulated and real case downscaled precipitation changes are seen in the southeast of the country in autumn. It was expected since the SD method is not able to reproduce correctly simulated changes within the perfect model framework in this region in autumn (Figures 7d and 7h).

©2015. American Geophysical Union. All Rights Reserved.

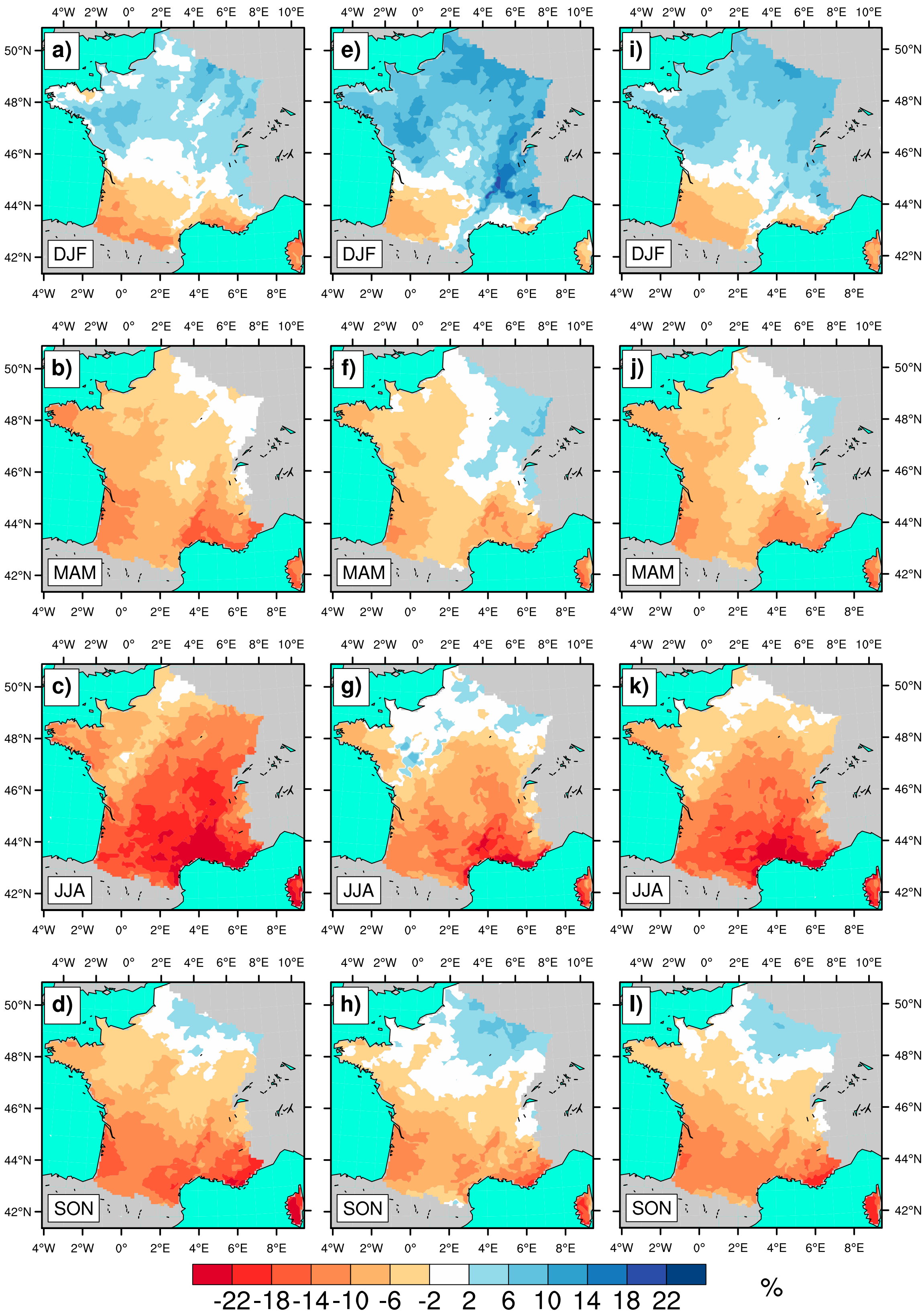

Figure 9. Seasonal relative changes in downscaled precipitation (%) between the 1961–1990 and 2030–2050 periods. The SD method with the C6 combination of predictors is built for the four reanalyses, and the 12 RCMs from the ENSEMBLES project are downscaled. The multi-RCM average is then computed. (a–d) Across-reanalysis minimum. (e–h) Across-reanalysis maximum. (i–l) Results obtained for ERAI.

©2015. American Geophysical Union. All Rights Reserved.

As real case downscaled changes are very similar to the changes in the perfect model framework, it can be deduced that the present-day links between predictors and precipitation are consistent in RCMs and observations. This result reinforces our confidence in the realism of RCM simulations and in the relevance of the perfect model results to the real world. Moreover, the difference between the resolution of predictors and precipitation on the learning period between real case and perfect model downscaling does not lead to important differences in future precipitation changes. The high resolution of predictors in the perfect model approach compared to the typical resolution of reanalyses did not bias the results of the analysis.

The range of future downscaled precipitation changes associated with the choice of the reanalysis is limited but not negligible (Figures 9a to 9h). Even if the spatial pattern is robust to the choice of the reanalysis, differences in magnitudes greater than 5% are often seen (e.g., in summer or in winter in northern France). As explained in section 3, uncertainties in reanalyses are a serious issue for the present-day evaluation of SD methods, especially as far as trends are concerned. Even if it is not negligible, the impact of reanalyses in the uncertainties of future precipitation changes appears here less crucial. Indeed, the spread associated with reanalyses has to be thought of with regard to the other uncertainties involved, and, in particular, the intermodel spread in precipitation changes, which is very large (Figures 7i–7p).

5.2. Uncertainties

A decomposition of variance [*Déqué et al.*, 2007, 2011] is now used to quantify more precisely the contribution of reanalyses to the total uncertainty and its relative importance against uncertainties from RCM simulations and stochastic uncertainties. As the local precipitation state is not entirely controlled by the large-scale predictors, for the same predictors state, several corresponding equally probable local precipitation states may exist. Following *Lafaysse et al.* [2011], this uncertainty is called here stochastic. *Lafaysse et al.* [2011] have shown its importance in the French Alps for the present-day evaluation of SD methods. Here the stochastic uncertainty is assessed by including a random step to the downscaling algorithm described in section 2. Instead of selecting the best analog day (in terms of predictors similarity), the 10 best analogs are preselected, and then, each day, five analogs are randomly chosen among those 10, to obtain five members, judged as equally probable and to correspond to the same predictors state.

Note that what is called uncertainties due to RCM simulations in this study is somewhat complex because of the experimental design of ENSEMBLES regional projections, used here as an ensemble of opportunity. As only one member is generally available for each RCM/GCM combination, it is not possible to separate cleanly the uncertainties due to model formulation and due to internal variability (which is itself both related to the internal variability of the RCM and to the internal variability of the GCM used for boundary conditions). Therefore, uncertainties from RCM simulations include three types of uncertainties: uncertainties from RCMs, GCM boundary conditions, and internal variability. Moreover, as the same RCM may be forced by several GCMs and the same GCM may force several RCMs (Table 1) the estimation of RCM simulation uncertainties may be somewhat biased, as all the configurations are not totally independent.

Three sources of uncertainties in the mid-21st century downscaled precipitation changes are evaluated. Let P_{mrs} be the projected relative changes in precipitation on a grid cell obtained after the statistical downscaling of the RCM simulation m, for the reanalysis r, and the member corresponding to stochastic uncertainties s:

1. $m = 1 - 12$: according to the RCM simulation;
2. $r = 1 - 4$: according to the reanalysis; and
3. $s = 1 - 5$: according to the member.

We therefore have $12 \times 4 \times 5$, i.e., 240 precipitation projections over France. Each contribution to the total variance is then calculated as in *Déqué et al.* [2007]. The total variance of P can be decomposed in seven contributions:

$$Var(P) = M + R + S + MR + RS + MR + MRS \quad (1)$$

©2015. American Geophysical Union. All Rights Reserved.

10.1002/2014JD022236

Table 3. Percentage of Variance in Relative Precipitation Changes Explained by the RCM Simulations (*M*), the Reanalyses (*R*), and the Stochastic Variations (*S*), and the Multifactor Terms (*MR*, *MS*, *RS*, and *MRS*)

	M	R	S	MR	MS	RS	MRS
DJF	70.3	4.4	0.5	3.8	4.9	0.5	15.6
MAM	74.8	1.6	0.4	4.1	4.9	0.4	13.9
JJA	62.0	5.4	0.5	5.9	6.6	0.5	19.1
SON	62.3	7.7	0.7	4.5	5.7	0.5	18.7

A dot (.) subscript represents the mean with respect to the corresponding indices.

$$M = \frac{1}{12}\sum_{m=1}^{12}(P_{m..} - P_{...})^2$$

$$MR = \frac{1}{48}\sum_{m=1}^{12}\sum_{r=1}^{4}(P_{mr.} - P_{m..} - P_{.r.} + P...)^2$$

$$MRS = \frac{1}{240}\sum_{m=1}^{12}\sum_{r=1}^{4}\sum_{s=1}^{5}(P_{mrs} - P_{mr.} - P_{.rs} - P_{m.s} + Pm.. + P.r. + P..s - P...)^2$$

M is the contribution of RCM simulations to the total variance, *R* the contribution of the reanalyses, and *S* of stochastic uncertainties. The other terms represent the interactions. For example, *MR* is the interaction effect between RCM simulations and reanalyses. All the possible combinations of RMC simulations, reanalyses, and members are available in our case, such as the estimation of each contribution can be directly calculated. The RCM simulations (a combination of an RCM and a GCM for boundary conditions) are considered as independent here. The mean contributions over France of each source of uncertainty to the total variance of relative precipitation change between 2030–2050 and 1961–1990 are given in Table 3.

RCM simulations are the largest source of uncertainties (up to 75%). As shown previously in Figure 9, reanalyses are a nonnegligible source of uncertainties (contribution between 5% and 10%). The part of variance associated with stochastic uncertainties only is, independently of the season, very small (around 0.5%). The second main contributor to the total variance is the full interaction term (*MRS*). The other interaction terms that include RCM simulations are close to 5%. The interaction terms associated with reanalyses and stochastic uncertainties are very small, close to 0.5%, as the stochastic term alone.

It can be concluded that the RCM simulation uncertainties, that include structural uncertainties from the RCMs, from GCMs through boundary forcing and from internal climate variability play the most important role in future downscaled precipitation change. Note that the estimation of stochastic and reanalyses uncertainties are expected to depend on the SD method and on the predictors. For example, a similar decomposition of variance for the C2 combination of predictors (SLP and TAS) leads to a smaller contribution of reanalyses (between 1% and 5%, not shown), presumably because there are fewer predictors in C2 and they are probably less uncertain than some predictors in C6 like TTI. Even with C6, the reanalyses are not a crucial source of uncertainty for future precipitation change.

6. Conclusion

In this study, a statistical downscaling methodology for precipitation over France based on the analog method has been presented. First, several combinations of predictors have been tested in the present climate with four reanalyses (NCEP, 20CR, ERAI, and MERRA). For most combinations of predictors, the SD method leads to a good reproduction of the climatology and interannual variability of precipitation on the training period. It is difficult to select a particular combination of predictors based on this evaluation, as the differences due to the choice of the reanalyses are often greater than the differences associated with the choice of predictors. The evaluation of the SD method and predictors on their ability to capture observed present-day trends has not led to strong conclusions. The observed trends in precipitation are indeed small and seldom significant, and a large spread in downscaled trends due to reanalyses exists. Still, it has been

©2015. American Geophysical Union. All Rights Reserved.

possible to reject a predictor (LCL), because it led to too unrealistic trends, probably because of temporal inhomogeneities in reanalyses. As for most predictors, observations with a correct spatiotemporal coverage do not exist, one has to rely on reanalyses for predictors to build the SD method. When evaluating the SD method against observations as we did, it is in fact not only the SD method that is evaluated but a combination of the SD method and a given reanalysis. The substantial differences in downscaling results associated with reanalyses in that context suggests that the role of reanalyses should not be underestimated when evaluating the SD method.

The transferability to the future climate of the SD method has been assessed using a perfect model framework. In the perfect model framework, the downscaling method is developed with predictors and precipitation from a present-day RCM simulation. The RCM is then statistically downscaled, and downscaled future precipitation changes can be compared to precipitation changes simulated by the RCM. The downscaling method is able to reproduce the ensemble mean change in precipitation simulated by 12 RCMs from the ENSEMBLES project but only with a specific combination of predictors. This result gives a better confidence on the transferability of the downscaling method with these predictors to the future climate. The successful combination of predictors includes the sea level pressure, the temperature at 2 m, the moisture flux at 850 hPa, the TTI, and the specific humidity at 850 hPa. Tests have shown that the TTI and the specific humidity were important to correctly capture the simulated decrease in summer precipitation over France. It highlights the importance of changes in atmospheric moisture and stability in that context. Regarding TTI changes, there is a competition between the increase in moisture (leading to an increase in TTI) and the decrease in the vertical temperature gradient (leading to a decrease in TTI). The decrease in the vertical temperature gradient dominates, as the TTI decreases in all the RCMs studied, which is associated with a decrease in summer precipitation over France. Regarding changes in specific humidity, because of the Clausius-Clapeyron relation, the specific humidity at saturation strongly increases with temperature and so does the specific humidity, although to a lesser extent. As humidity may be a limiting factor of precipitation under some conditions, the increase in humidity somewhat mitigates the decrease in summer precipitation due to other factors.

The same RCMs have also been downscaled as for real case application, thanks to the observed statistical relationship between predictors and precipitation. The results of real case downscaling are very close to those of perfect model downscaling and very similar to directly simulated precipitation changes. This result highlights the realism of the relationship between predictors and precipitation in the RCMs, enhancing our confidence in RCM results and in the relevance of the perfect model framework to the real world.

Finally, the respective importance of different sources of uncertainties in future downscaled precipitation change has been studied with a decomposition of variance. Uncertainties from RCM simulations, i.e., uncertainties from RCMs, GCM boundary conditions, and from their respective internal variability are found to be the dominant source of uncertainties. The importance of reanalyses in that context is much smaller. Consequently, for an impact study, it does not seem crucial to develop several variants of the SD method based on different reanalyses for the calibration on the learning period and then downscale climate projections with all these variants to sample the uncertainties associated with the reanalyses.

From a more general perspective, our study illustrates some important methodological issues for SD. First, even if uncertainties in reanalyses are not crucial for future climate change, they are important in the context of the development and evaluation of SD methods in the present climate. In our case, the reanalyses do not necessarily simply agree on the best combination of predictors. The optimization of the SD method with a single reanalysis is therefore likely to be reanalysis dependent and might lack robustness.

Our results also illustrate the issues that exist in evaluating a statistical downscaling method intended to be applied in the climate change context using only present-day observations. The skill in the present climate is not necessarily a good indicator of the skill in the future climate. In our case, combinations of predictors leading to very similar and satisfactory results in terms of present-day climatology, interannual variability, and trends are associated with important differences in future precipitation changes. Our results show the interest of a perfect model framework with multiple models to better evaluate the transferability through time of a statistical downscaling method, even if the transferability to the future climate in the perfect model framework is not an absolute guarantee of the transferability in the real world.

©2015. American Geophysical Union. All Rights Reserved.

Our results also suggest that one must remain cautious when trying to estimate the uncertainties associated with different downscaling methods. Different SD methods are often considered as equally realistic in the future climate simply because they lead to similarly good results in the present climate. Differences in future downscaled changes are consequently interpreted as uncertainties. We have seen that equally realistic methods with regard to present-day evaluation can lead to different results for future changes in the perfect model framework. To our opinion, for SD methods with equally good results in the present climate, those that do not show transferability to the future climate in the perfect model framework should be given much less weight or even discarded for impact studies. Using only the SD methods that demonstrate their ability in a perfect model framework in the climate change context would likely seriously reduce what is generally seen as uncertainties due to SD but are probably, to a large extent, errors. As a matter of fact, interestingly, our final SD method applied in real case configuration, i.e., with the SD relationship built with observations and reanalyses and tested for temporal transferability in the perfect model framework, leads in the end to results very similar to dynamical downscaling. It is possible only if the present-day links between predictors and precipitation in the RCMs (used in the perfect model framework) and in the real world (used in real case downscaling) are consistent. It highlights the robustness of downscaling results to the choice of the downscaling framework in this study.

The assessment of the downscaling method presented in this work has been focused on precipitation, because it is the most important variable for our future hydrological application. Our ultimate objective is indeed to study the impacts of climate change on the continental hydrological cycle in France using a physically based hydrometeorological model, based on statistically downscaled CMIP5 models. This will be the object of a future study.

Acknowledgments
NCEP reanalysis data have been provided by the National Oceanic and Atmospheric Administration (NOAA), Office of Oceanic and Atmospheric Research (OAR) and the Earth System Research Laboratory (ESRL). European Center for Medium-Range Weather Forecasts (ECMWF) ERA-Interim data used in this study have been obtained from the ECMWF data server. MERRA data used in this study have been provided by the Global Modelling and Assimilation Office (GMAO) at NASA Goddard Space Flight Center through the NASA Goddard Earth Sciences Data and Information Services Center (GES DISC) online archive. Support for the Twentieth Century Reanalysis Project data set is provided by the U.S. Department of Energy, Office of Science Innovative and Novel Computational Impact on Theory and Experiment (DOE INCITE) program, and Office of Biological and Environmental Research (BER), and by the National Oceanic and Atmospheric Administration Climate Program Office. The authors would like to thank Météo France (Direction de la climatologie and CNRM/GAME) for providing the SAFRAN data. The ENSEMBLES data used in this work were funded by the EU FP6 Integrated Project ENSEMBLES (contract 505539) whose support is also gratefully acknowledged. This work was supported by the French National Research Agency (ANR) in the framework of its JCJC program (ECHO, decision No ANR 2011 JS56 014 01).

References

Ben Daoud, A., E. Sauquet, M. Lang, G. Bontron, and C. Obled (2011), Precipitation forecasting through an analog sorting technique: A comparative study, *Adv. Geosci.*, *29*, 103–107, doi:10.5194/adgeo-29-103-2011.

Beuchat, X., B. Schaefli, M. Soutter, and A. Mermoud (2012), A robust framework for probabilistic precipitations downscaling from an ensemble of climate predictions applied to Switzerland, *J. Geophys. Res.*, *117*, D03115, doi:10.1029/2011JD016449.

Boberg, F., P. Berg, P. Thejll, W. J. Gutowski, and J. H. Christensen (2009), Improved confidence in climate change projections of precipitation further evaluated using daily statistics from ENSEMBLES models, *Clim. Dyn.*, *35*(7–8), 1509–1520, doi:10.1007/s00382-009-0683-8.

Boé, J. (2012), Modulation of soil moisture-precipitation interactions over France by large scale circulation, *Clim. Dyn.*, *40*(3–4), 875–892, doi:10.1007/s00382-012-1380-6.

Boé, J., and L. Terray (2007), A weather-type approach to analyzing winter precipitation in France: Twentieth-century trends and the role of anthropogenic forcing, *J. Clim.*, *21*(13), 3118–3133, doi:10.1175/2007JCLI1796.1.

Boé, J., and L. Terray (2008), Uncertainties in summer evapotranspiration changes over Europe and implications for regional climate change, *Geophys. Res. Lett.*, *35*, L05702, doi:10.1029/2007GL032417.

Boé, J., L. Terray, E. Martin, and F. Habets (2009), Projected changes in components of the hydrological cycle in French river basins during the 21st century, *Water Resour. Res.*, *45*, W08426, doi:10.1029/2008WR007437.

Boer, G. J. (2004), Long time-scale potential predictability in an ensemble of coupled climate models, *Clim. Dyn.*, *23*(1), 29–44, doi:10.1007/s00382-004-0419-8.

Chardon, J., B. Hingray, A.-C. Favre, P. Autin, J. Gailhard, I. Zin, and C. Obled (2014), Spatial similarity and transferability of analog dates for precipitation downscaling over France, *J. Clim.*, *27*(13), 5056–5074, doi:10.1175/JCLI-D-13-00464.1.

Collins, M., M. Botzet, and A. F. Carril (2006), Interannual to decadal climate predictability in the North Atlantic: A multimodel-ensemble study, *J. Clim.*, *19*, 1195–1203.

Compo, G. P., et al. (2011), The twentieth century reanalysis project, *Q. J. R. Meteorol. Soc.*, *137*(654), 1–28, doi:10.1002/qj.776.

Dee, D. P., et al. (2011), The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, *Q. J. R. Meteorol. Soc.*, *137*(656), 553–597, doi:10.1002/qj.828.

Déqué, M., D. P. Rowell, D. Lüthi, F. Giorgi, J. H. Christensen, B. Rockel, D. Jacob, E. Kjellström, M. Castro, and B. Hurk (2007), An intercomparison of regional climate simulations for Europe: Assessing uncertainties in model projections, *Clim. Change*, *81*, 53–70, doi:10.1007/s10584-006-9228-x.

Déqué, M., S. Somot, E. Sanchez-Gomez, C. M. Goodess, D. Jacob, G. Lenderink, and O. B. Christensen (2011), The spread amongst ENSEMBLES regional scenarios: Regional climate models, driving general circulation models and interannual variability, *Clim. Dyn.*, *38*, 951–964, doi:10.1007/s00382-011-1053-x.

Frias, M. D., and E. Zorita (2006), Testing statistical downscaling methods in simulated climates, *Geophys. Res. Lett.*, *33*, L19807, doi:10.1029/2006GL027453.

Guilbaud, S., and C. Obled (1998), Prévision quantitative des précipitations journalières par une technique de recherche de journées antérieures analogues: Optimisation du critère d'analogie, *Géosci. Surf.*, *327*, 181–188.

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

Habets, F., et al. (2008), The SAFRAN-ISBA-MODCOU hydrometeorological model applied over France, *J. Geophys. Res.*, *113*, D06113, doi:10.1029/2007JD008548.

Hamed, K. H., and A. Ramachandra Rao (1998), A modified Mann-Kendall trend test for autocorrelated data, *J. Hydrol.*, *204*(1–4), 182–196, doi:10.1016/S0022-1694(97)00125-X.

Hawkins, E., J. Robson, R. Sutton, D. Smith, and N. Keenlyside (2011), Evaluating the potential for statistical decadal predictions of sea surface temperatures with a perfect model approach, *Clim. Dyn.*, *37*(11–12), 2495–2509, doi:10.1007/s00382-011-1023-3.

©2015. American Geophysical Union. All Rights Reserved.

Horton, P., M. Jaboyedoff, R. Metzger, C. Obled, and R. Marty (2012), Spatial relationship between the atmospheric circulation and the precipitation measured in the western Swiss Alps by means of the analogue method, *Nat. Hazard Earth Syst. Sci.*, *12*(3), 777–784, doi:10.5194/nhess-12-777-2012.

Jacob, D., et al. (2013), EURO-CORDEX: New high-resolution climate change projections for European impact research, *Reg. Environ. Change*, *14*(2), 563–578, doi:10.1007/s10113-013-0499-2.

Kalnay, E., and M. Kanamitsu (1996), The NCEP/NCAR 40-year reanalysis project, *Bull. Am. Meteorol. Soc.*, *77*(3), 437–471.

Kelley, C., M. Ting, R. Seager, and Y. Kushnir (2012), Mediterranean precipitation climatology, seasonal cycle, and trend as simulated by CMIP5, *Geophys. Res. Lett.*, *39*, L21703, doi:10.1029/2012GL053416.

Kendall, M. G., and J. D. Gobbons (1990), *Rank Correlation Methods*, 5th ed., 272 pp., vol. 3, Oxford Univ. Press, New York.

Krueger, O., F. Schenk, F. Feser, and R. Weisse (2013), Inconsistencies between long-term trends in storminess derived from the 20CR reanalysis and observations, *J. Clim.*, *26*(3), 868–874, doi:10.1175/JCLI-D-12-00309.1.

Lafaysse, M., and B. Hingray (2014), Internal variability and model uncertainty components in future hydrometeorological projections: The Alpine Durance basin, *Water Resour. Res.*, *50*, 3317–3341, doi:10.1002/2013WR014897.

Lafaysse, M., B. Hingray, P. Etchevers, E. Martin, and C. Obled (2011), Influence of spatial discretization, underground water storage and glacier melt on a physically-based hydrological model of the Upper Durance River basin, *J. Hydrol.*, *403*(1–2), 116–129, doi:10.1016/j.jhydrol.2011.03.046.

Lawrence, M. G. (2005), The relationship between relative humidity and the dewpoint temperature in moist air: A simple conversion and applications, *Bull. Am. Meteorol. Soc.*, *86*(2), 225–233, doi:10.1175/BAMS-86-2-225.

Lorenz, E. N. (1969), Atmospheric predictability as revealed by naturally occurring analogues, *J. Atmos. Sci.*, *26*, 636–646.

Maraun, D., et al. (2010), Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user, *Rev. Geophys.*, *48*, RG3003, doi:10.1029/2009RG000314.

Maurer, E. P., and H. G. Hidalgo (2008), Utility of daily vs. monthly large-scale climate data: An intercomparison of two statistical downscaling methods, *Hydrol. Earth Syst. Sci.*, *12*(2), 551–563.

Meinshausen, M., et al. (2011), The RCP greenhouse gas concentrations and their extensions from 1765 to 2300, *Clim. Change*, *109*(1–2), 213–241, doi:10.1007/s10584-011-0156-z.

Mezghani, A., and B. Hingray (2009), A combined downscaling-disaggregation weather generator for stochastic generation of multisite hourly weather variables over complex terrain: Development and multi-scale validation for the Upper Rhone River basin, *J. Hydrol.*, *377*(3–4), 245–260, doi:10.1016/j.jhydrol.2009.08.033.

Muerth, M. J., B. Gauvin St-Denis, S. Ricard, J. A. Velázquez, J. Schmid, M. Minville, D. Caya, D. Chaumont, R. Ludwig, and R. Turcotte (2013), On the need for bias correction in regional climate scenarios to assess climate change impacts on river runoff, *Hydrol. Earth Syst. Sci.*, *17*(3), 1189–1204, doi:10.5194/hess-17-1189-2013.

Nakicenovic, N., P. Koip, K. Riahi, M. Kainuma, and T. Hanaoka (2006), Assessment of emissions scenarios revisited, *Environ. Econ. Policy Stud.*, *7*, 137–173, doi:10.1007/s10018-005-0112-0.

Obled, C. (2002), Quantitative precipitation forecasts: A statistical adaptation of model outputs through an analogues sorting approach, *Atmos. Res.*, *63*(3–4), 303–324.

Plaut, G., E. Schuepbach, and M. Doctor (2001), Heavy precipitation events over a few Alpine sub-regions and the links with large-scale circulation, 1971–1995, *Clim. Res.*, *17*, 285–302.

Quintana-Seguí, P. (2008), Analysis of near-surface atmospheric variables: Validation of the SAFRAN analysis over France, *J. Appl. Meteorol. Climatol.*, *47*(1), 92–107, doi:10.1175/2007JAMC1636.1.

Radanovics, S., J.-P. Vidal, E. Sauquet, A. Ben Daoud, and G. Bontron (2013), Optimising predictor domains for spatially coherent precipitation downscaling, *Hydrol. Earth Syst. Sci.*, *17*(10), 4189–4208, doi:10.5194/hess-17-4189-2013.

Ribes, A., S. Planton, and L. Terray (2013), Application of regularised optimal fingerprinting to attribution. Part I: Method, properties and idealised analysis, *Clim. Dyn.*, *41*(11–12), 2817–2836, doi:10.1007/s00382-013-1735-7.

Rienecker, M. M., and M. J. Suarez (2011), MERRA: NASA's modern-era retrospective analysis for research and applications, *J. Clim.*, *24*(14), 3624–3648, doi:10.1175/JCLI-D-11-00015.1.

Sanchez-Gomez, E., and L. Terray (2005), Large-scale atmospheric dynamics and local intense precipitation episodes, *Geophys. Res. Lett.*, *32*, L24711, doi:10.1029/2005GL023990.

Schmith, T. (2008), Stationarity of regression relationships: Application to empirical downscaling, *J. Clim.*, *21*(17), 4529–4537, doi:10.1175/2008JCLI1910.1.

Sherwood, S. C., W. Ingram, Y. Tsushima, M. Satoh, M. Roberts, P. L. Vidale, and P. A. O'Gorman (2010), Relative humidity changes in a warmer climate, *J. Geophys. Res.*, *115*, D09104, doi:10.1029/2009JD012585.

Simmons, A. J., K. M. Willett, P. D. Jones, P. W. Thorne, and D. P. Dee (2010), Low-frequency variations in surface atmospheric humidity, temperature, and precipitation: Inferences from reanalyses and monthly gridded observational data sets, *J. Geophys. Res.*, *115*, D01110, doi:10.1029/2009JD012442.

Smerdon, J. E. (2012), Climate models as a test bed for climate reconstruction methods: Pseudoproxy experiments, *Wiley Interdiscip. Rev. Clim. Change*, *3*(1), 63–77, doi:10.1002/wcc.149.

Spak, S., T. Holloway, B. Lynn, and R. Goldberg (2007), A comparison of statistical and dynamical downscaling for surface temperature in North America, *J. Geophys. Res.*, *112*, D08101, doi:10.1029/2005JD006712.

Stocker, T., D. Qin, G.-K. Plattner, M. Tignor, S. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. M. Midgley (2013), *Climante Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge Univ. Press, Cambridge, U. K., and New York.

Timbal, B., and B. J. McAvaney (2001), An analogue-based method to downscale surface air temperature: Application for Australia, *Clim. Dyn.*, *17*(12), 947–963.

Timbal, B., A. Dufour, and B. Mcavaney (2003), An estimate of future climate change for western France using a statistical downscaling technique, *Clim. Dyn.*, *20*(7–8), 807–823, doi:10.1007/s00382-002-0298-9.

Timbal, B., P. Hope, and S. Charles (2008), Evaluating the consistency between statistically downscaled and global dynamical model climate change projections, *J. Clim.*, *21*(22), 6052–6059, doi:10.1175/2008JCLI2379.1.

Trenberth, K. E. (1998), Atmospheric moisture residence times and cycling: Implications for rainfall rates and climate change, *Clim. Change*, *39*, 667–694.

Trenberth, K. E. (1999), Atmospheric moisture recycling: Role of advection and local evaporation, *J. Clim.*, *12*(5), 1368–1381, doi:10.1175/1520-0442(1999)012<1368:AMRROA>2.0.CO;2.

Trenberth, K. E., A. Dai, R. M. Rasmussen, and D. B. Parsons (2003), The changing character of precipitation, *Bull. Am. Meteorol. Soc.*, *84*(9), 1205–1217, doi:10.1175/BAMS-84-9-1205.

©2015. American Geophysical Union. All Rights Reserved.

van der Linden, P., J. F. B. Mitchell, and P. Gilbert (2009), *ENSEMBLES: Climate Change and Its Impacts: Summary of Research and Results From the ENSEMBLES Project*, 160 pp., Met Off. Hadley Centre, Exeter, U. K.

Vidal, J.-P., E. Martin, L. Franchistéguy, M. Baillon, and J.-M. Soubeyroux (2010), A 50-year high-resolution atmospheric reanalysis over France with the Safran system, *Int. J. Climatol.*, *30*(11), 1627–1644, doi:10.1002/joc.2003.

Vrac, M., and M. L. Stein (2007), A general method for validating statistical downscaling methods under future climate change, *Geophys. Res. Lett.*, *34*, L18701, doi:10.1029/2007GL030295.

Wilby, R. L., H. Hassan, and K. Hanaki (1998), Statistical downscaling of hydrometeorological variables using general circulation model output, *J. Hydrol.*, *205*(1–2), 1–19.

Zorita, E., J. P. Hughes, D. P. Lettemaier, and H. V. Storch (1995), Stochastic characterization of regional circulation patterns for climate model diagnosis and estimation of local precipitation, *J. Clim.*, *8*, 1023–1042.

©2015. American Geophysical Union. All Rights Reserved.