Toulouse School of Economics - Université Toulouse III Paul Sabatier
M1 Economics & Statistics
Magistère d'Economiste Statisticien

A study of the performance of DSCLIM, a statistical downscaling tool for impact studies

Marie Vialaret

2016-07-29

Internship Report
From the 2nd May to the 29th July 2016 at CERFACS

Climate Modelling and Global Change
TR-CMGC-16-212

CERFACS
Centre Européen de Recherche
et de Formation Avancée en Calcul Scientifique
42, avenue G. Coriolis
31057 Toulouse Cedex, France

Acknowledgments

I would like to express my sincere thanks to my supervisor at CERFACS, Mr. Christian Pagé, who enabled me to find this internship and helped me throughout it though letting me a lot of freedom.
I would also thank Mr. Gildas Dayon for the time he took to present me his downscaling method and results, for his help and answers during my whole internship.
Finally, I would like to thank all the members of the GLOBC team for their welcome and kindness throughout my stay at CERFACS.

Abstract

In the framework of climate change, being able to produce accurate climate scenarios for impact studies is essential for decisions makers, firms and citizens, especially at a fine scale. Therefore a large share of climate change studies use the information contained in global climate scenarios to derive higher spatial resolution scenarios (downscaling). The objective of my internship is to assess the performance of a downscaling tool developed at CERFACS, DSCLIM, to simulate long-term rainfall in the European region.

First of all I analyzed the two steps of this downscaling method in a perfect model framework. I will show that the regression phase gives good results if one disregards colinearity problems. The ability of DSCLIM to fit the distribution of observed precipitations is good whereas the projection of extreme rainfall is not reproduced, as it is anticipated with this type of method.

Then I have been interested in the performance of DSCLIM with real data and compared analog precipitations obtained with ERAI reanalysis and CMIP5 model data to SAFRAN observations, which gave similar results as in the perfect model framework.

The two big problems highlighted by this report are the colinearity of explanatory variables in the regression phase and the low quality of fit for extreme events. Therefore I will be eventually interested in assessing the performance of an other statistical downscaling method developed by G. Dayon during his PhD. This method makes less strong hypothesis than DSCLIM, in particular on regression and weather types. The temporal and spatial correlations are better with this method but it does not reproduce distributions well.

Warning

Several results have changed compared to the intermediate internship report (2016-06-15). The analysis of the regression remains unchanged but some modifications have been done in the DSCLIM method concerning the selection of analog days. Indeed, the final analog day is not uniformly drawn from the initial eleven analog days as before but the stochastic resampling method described in section 2.2.3. Thus some results, tables and figures presented in the analysis of the downscaled precipitations have been modified in this new report.

Contents

List of Tables

List of Figures

1 Introduction

1.1 Presentation of the firm

The CERFACS (Centre Européen de la Recherche et de la Formation Avancée en Calcul Scientifique) is a research center created in 1987 and specialized in modelisation and numeric simulation. This professional partnership maintains a strong cooperation with its shareholders (CNES, EDF, Airbus Group, Total, Safran, Météo France and Onera) and has developed partnerships with CNRS, IRIT, CEA and INRA.

Its main mission is to develop scientific and technical research to improve advanced calculus methods and to solve scientific problems that require a significant calculation power. The center comprises nine research axes: High Performance Computing, Data assimilation and optimization, Numerical methods and linear algebra, Combustion, Aerodynamics, Uncertainties, Climate, Coupling and interfaces, Environment. There are five teams composed of physicians, applied mathematicians, numeric analysts and software engineers.

During my internship I joined the team Climate Modeling and Global Change (GLOBC) that participates in the development of coupled models between oceans, atmosphere and other components of the climate system. The team's strategy is based on a dual approach combining theoritical modeling studies and the development of high-level softwares required to address various issues involved in climate science, industrial applications and impact studies.

1.2 General problematics and objectives of the internship

The main changes that have been observed from the beginning of the last century in the climate are the warming of the lower atmosphere and the increase of the sea level. The main reason for these changes has been shown to be anthropic, the concentration of greenhouse gases due to human activities rising at unprecedented levels. The awareness of this situation by the scientific community arose in the end of the 80s and lead to the setting up of the IPCC (Intergovernmental Panel for Climate Change) and of international conventions for the limitation of greenhouse gas emissions. In this framework being able to set up scenarios for climate change becomes essential for decision makers, firms and citizens. Thus reliable projections are required to produce impact studies and to deal with global climate change as accurately as possible (adaptation of agriculture, allocation of investments, frequency of extreme events...). In particular one needs to simulate climate at a sufficiently fine scale since it can importantly vary at local scales because of ground use, topography, seasonal phenomena, sea-ground contrast, etc. Therefore an essential step for most of the climate change studies is to use the information contained in global climate scenarios to derive higher spatial resolution scenarios: this procedure is called downscaling. Throughout my internship, I am working with a statistical downscaling method developed at CERFACS and called DSCLIM (DonwScaled CLIMate scenarios). This model is currently used in an inter-comparison experience between various downscaling methods in the framework of the European action COST VALUE.

Objectives of the internship Before applying the DSCLIM method to make climate projections over the whole European region, as it is planned in the long term, one has to assess its performance in a perfect model framework. Indeed, if the method does not work in such a framework then it has very little chances to be efficient with real data. From now I will present the DSCLIM method and then the statistical analysis of the downscaled precipitations over metropolitan France. A first analysis will be done in a perfect model framework, a second with real observations and eventually I will apply the diagnostic tools I have developed to an other downscaling method (described in [8]). Both results of the previous report (the internship summary submitted on 15th June 2016) and new results are presented.

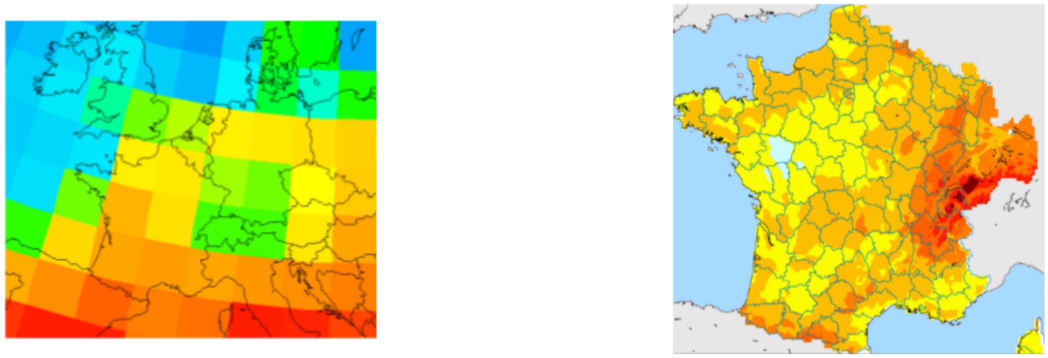


Figure 1: Precipitation modeling at global (left) and local scales (right) [9]

2 The DSCLIM method

2.1 Global context : statistical downscaling

Downscaling methods are traditionally classified in two categories: dynamical and statistical downscaling. Whereas the former rests upon solving dynamic atmospheric equations[1], the latter is based on the idea that local climate is conditioned by two factors: large scale climate and some physiographic local properties. Following this idea, one builds a statistical model that establishes a link between the large-scale state of climate and its local state (learning phase). This model is then used to derive local scale variables (precipitations, temperatures...) from large scale simulations obtained with low-resolution climate models. Statistical downscaling rests upon three major hypotheses:

- **Relevance of predictors:** predictors must have a strong link with regional climate and be simulated in a realistic way by the global climatic models ; if it is not the case, the local climatic projections will be biased.
- **Stationarity hypothesis:** the model built today remains valid in the future climate ie for a climate that is disrupted by anthropic forcing. This hypothesis represents the main weakness of statistical downscaling: it can not be formally validated. Moreover, today's secondary phenomena may become more and more important in the establishment of the future link between global and local events. One can somehow drop this hypothesis but it is impossible to remove it completely.
- **Signal of climate change:** we try to establish a statistical link between some predictors and a predictand in the framework of climate change ; as a consequence, if predictors do not react to the climate change, then building this relationship is useless. In our case, climate change is reflected by the changes of occurrence of weather types and by the changes in the distribution of days around the centroid of a given weather type.

Despite some drawbacks (principally the stationarity hypothesis), statistical methods are often less costly than dynamical ones and allow to grasp uncertainties better. Statistical downscaling uses four main types of methods which are weather type classification, regression, analogs, weather generators. DSCLIM combines the first three methods, which will be developed in section 2.2.

[1] Actually dynamical downscaling is not that simple, one has also to deal with several physical parametrizations, among which surface characteristics, clouds' microphysics... Precisions can be found in [1]

2.2 Downscaling method

2.2.1 General framework

Weather regimes and weather types The large-scale atmospheric circulation is characterized by fluctuations around more or less stationary states called weather regimes. One counts four weather regimes in western Europe: NAO+, NAO-, Scandinavian Blocking and Atlantic Ridge. On the other hand, weather types are generally used to characterize finer scale phenomena. It has been shown[2] that weather types are quite useful to characterize links between global atmospheric circulation in the European and North-Atlantic zone and regional climate variables. They are usually obtained using classification algorithms as I will explain in the description of the DSCLIM method (downscaling phase). For an illustration of weather types, see figure 15.

Seasonal links between Mean Sea-Level Pressure and precipitations Different downscaling models are defined for the four seasons. Indeed the links between large scale phenomena (mainly the mean sea level pressure (MPSL)) and local scale precipitations varies with seasons. They are defined as follows: SON-Autumn (September, October, November), DJF-Winter (December, January, February), MAM-Spring (March, April, May) and JJA-Summer (June, July, August). Let me now roughly describe the links between MPSL and precipitations. DFJ shows a depression centered over northern-western Europe, while intermediate seasons (MAM and SON) show a weaker minimum of MPSL stretched from western to eastern northern Europe. Finally, JJA shows a high pressure centered over eastern Mediterranean sea and encompassing all southern Europe. Thus in summer the cold air mass is concentrated in northern Europe and the atmospheric circulation is slower in southern Europe and France. Conversely, in winter, the atmospheric circulation is accelerated because of the low pressure. Therefore, precipitations in summer are less linked to APSL (Anomalies of Sea Level Pressure) than in winter. The precipitations associated to the weather types I described above are globally as follows: in autumn, it s quite wet for almost all of France, especially over the mountains, the Atlantic coast and northern France. Winter is drier than autumn for regions exposed to the Mistral. In spring, the Mediterranean coast and the plains in northwestern France are drier, and in summer the latter regions are even drier.

2.2.2 Learning phase

The data Two datasets are used: a set of observations (precipitations and temperatures) for France (SAFRAN) or Europe (EURO-4M) for the period 1989-2010 and reanalyses[3] for the sea level pressure (PSL). The former has a 5-10 km resolution while the latter has a 50-250 km resolution. A pre-processing is executed on these data. A first step consists in subtracting the seasonal cycle of the PSL data in order to get rid of seasonal trends. Then one performs a climatology[4] of the PSL, subtracts the average PSL of each day to the daily PSL and smooths the pressure anomalies (APSL) curve.

Classification phase This phase includes two steps. First of all one performs a principal component analysis (PCA) on the APSL and the squared root of precipitation level ; the first ten principal components are kept for each variable (ten components are generally sufficient to explain most of the variance of the variables). The squared root of precipitation level is used instead of precipitations in order to deal with a variable which is more Gaussian. Thus we are left with ten empirical orthogonal functions (EOF) for both variables and each day. The second step of this classification phase consists in the application of a k-means algorithm[5] in order to classify the days in ten weather

[2] See [3] and [4].

[3] The reanalyze is a meteorologic method that aims to objectively reanalyze global surface and altitude data over long periods (several decades) for data assimilation in numeric projection climate models. One can use reanalyses from several sources: NCEP, ERAI, MERA, etc. The grid for the reanalyses varies from 50 to 250 km resolution and is for most of them about 150 km.

[4] ... a definir !!

[5] For further informations about the algorithm see [14] and [15].

types. At the initialization of the algorithm, ten centroids are randomly chosen, that are re-calculated at each step of the algorithm. At the end of this iterative process[6] one is left with ten new centroids defining the weather types. Then one can compute the distance between each day and each centroid (using the euclidean distance) and normalize it.

Regression phase This is the key part of the learning phase. One aims to build a link between weather types (large scale APSL) and local precipitations. To do so, a regression is performed for each grid point and each season:

$$\sqrt{precip_{kij}} = \beta_0 + \beta_1 X_{1kij} + ... + \beta_{10} X_{10kij} + \epsilon_{kij} \quad (1)$$

where k is a given day, i a given grid point and j a given season. The explained variable is the squared root of precipitations and the explanatory variables are the distances between each day and the ten weather-type centroids. A GLS method is used in order to deal with the homoskedasticity of residuals. Finally, one reconstructs precipitations using the coefficient estimates we just obtained and gets fitted values of precipitations.

2.2.3 Downscaling phase

This second step aims to produce future projections of precipitations at the spatial resolution of the regression. One begins again with a deseasonalisation of the data and the computation of EOFs for the APSL. Each day is thus characterized by its EOFs. Then the Euclidean distance between each day and each weather-type centroid we previously found is computed, each day is associated with a weather type using this distance and precipitations are computed for each day using the regression estimates derived in the learning phase. But following these steps one only gets precipitation data at the spatial resolution of the regression, which is coarser than the full resolution of the local precipitation dataset. That's why a second step is included to perform the downscaling further, using the analog method.

The analog method The idea that lies beneath the analog method is that the same causes lead to the same effects, e.g. the same large-scale APSL should lead to the same spatial precipitation distribution. Therefore for a given day d_{ref} one chooses the closest day (using a criteria of minimum Euclidean distance) d_1 in the learning period and associates the precipitations observed in d_1 to d_{ref}. Actually DSCLIM generalizes the method and selects the eleven closest days to d_{ref}, denoted d_1, ..., d_{11}, in the same weather type as d_{ref}. Moreover, the analog days have to be close to d_{ref} in terms of spatial correlation. Once the analog days are selected, one of them is randomly (uniformly) drawn as the final analog day. The observed precipitations of this day are then associated with d_{ref}. Note that the closest day to d_{ref} in terms of Euclidean distance is not necessarily better than the other ten days. That is why it is not a problem to draw randomly one day among the eleven days to be the final analog day, if the sample size is sufficiently large. Moreover, this stochastic step increases the variance in the downscaled precipitations, which is a clear advantage as mostly any statistical downscaling method reduces the variance of the downscaled variable (sometimes up to 50%). Nevertheless, depending on the size of the sample, it may happen that eleven good analog days are selected, but it may also happen that some very bad days are selected. As a consequence one can draw randomly a day which is "among the closest days", though quite different from the day of interest. Therefore, a method of stochastic resampling has been developed in order to avoid such situations. One looks at the spatial correlation between the analog days and the reconstructed precipitations and chooses the five days for which this correlation is the highest. The median correlation is computed using these five days. Then one selects the days that have a spatial correlation with the reconstructed precipitations which is close to this median[7]

[6] The algorithm converges quite quickly with DSCLIM and is usually stopped after roughly 30 iterations.

[7] One takes the days for which this correlation do not exceed 2σ of the correlation distribution. Usually this number is not higher than 10-11 days.

; time series are drawn for one of these days[8] and the final selection of the analog day is done using a maximum variance criteria.

3 Assessment of the method in a perfect model framework

3.1 The perfect model

Method The main difficulty encountered when testing the quality of a statistical downscaling model lies in the impossibility to test whether its simulated future observations will match the climate variables observed in the future or not. Indeed these climate projections can only be compared to the past data (containing the learning period). If the learning period does not correspond to the complete past period, one can evaluate the goodness of fit between observations and climate projections in the past period and thus partially evaluate the quality of the model, but the difficulty remains for the future, especially in the context of climate change. The idea of the perfect model is to produce local-scale projections for one particular climate variable (here the precipitation level) with a regional climate model and to consider them as pseudo-observations (as if one could observe past, present and future climate). Then, one applies the statistical downscaling method to large-scale climate variables obtained using the same climate model (and computed for past and future as well) and gets downscaled local precipitations over this long period. This provides a set of "observations" and a set of downscaled climate projections for a large period that includes both past and future dates, and one can compare these two sets. This method makes it possible to evaluate the internal consistency of the downscaling model for a given climate model.

Data Three datasets are used for the perfect model analysis:

- a set of pseudo-observations,
- a set of regression parameters and fitted values of precipitations obtained with them,
- a set of analog dates associated to each day we are interested in (ie each day of the period 1961-2050).

These datasets are available for each climate model and downscaled precipitations can be recovered using them. We use the reanalyses from twelve climate models[9] (C4I, CNRM, DMI-BCM, DMI-ECHAM5, IACETH, KNMI, METNO-BCM, METNO-Had, METO-HadQ0, METO-HadQ16, MPI and OURANOS) on metropolitan France. We have therefore pseudo-observations and analog days for the period starting on 1961-01-01 and ending on 2050-12-31.

Periods of interest In this perfect model analysis I will compare the results of two different periods : a past period (1961-1990) and a future period (2030-2050).

3.2 Analysis of the regressions

3.2.1 Regression coefficients

DSCLIM performs one regression for each grid point and each season ie 516*4 = 2064 regressions for each model, each of them being constructed using ten explanatory variables corresponding to the distances between the APSL

[8] One day is randomly drawn among the ones that were previously selected. This random selection is a way to increase the variance, the loss of variance being an important matter of downscaling. If one always takes the closest day to the median, chances are high to take several time the same day, which contributes to the loss of variance. Moreover, the days selected are not very different from each other so that it is not a big matter to randomly select one of them even if it is not the best one.

[9] The models are not exactly build in the same way e.g. they do not construct the variables in the same manner or the physical dynamics are not exactly the same ; numerical methods can be slightly different too. As a consequence the perfect model analysis performed with a model m is only valid in m: the analysis is carried out with several different models so that the results can be generalized (in particular the transferability hypothesis has more weight id it holds with several models than with only one in a perfect model framework).

of a given day and the APSLs associated with the ten weather types. I use a Student t-test to establish the significance of the coefficient estimates obtained by the GLS method, with the null hypothesis: H_0: "$\beta_k = 0$" where $k \in \{1, ..., 10\}$. H_0 is rejected in most of the cases and a wide majority of estimates are significant at the 99% level (p-value < 0.01). Non-significant estimates (p-value > 0.1) are observed in some regressions but they are never the same[10]. Moreover, as the test is done on a large number of coefficient estimates (20640 coefficients * 12 models) it seems natural that a few of them should be non-significant (actually less than 1%) so one can conclude that there is a good significance of the coefficient estimates in these regressions.

3.2.2 Residuals' behavior

- **Mean and variance:** The null hypothesis H_0: "$\mu_{residuals} = 0$" (for each model) can not be rejected by a Student's t-test whatever the regression and significance level considered. Moreover, the variance of residuals is not too wide. These two elements lead to the conclusion that residuals are globally null. This result is illustrated with figures 16 and 2.

- **Autocorrelation:** Considering all models together (ensemble mean), the residual's autocorrelation happens to be relatively small (cf table 1). Indeed, its minimum is 0.06 and its maximum 0.57, the average autocorrelation is around 0.2, and 75% of its values do not exceed 0.3 whatever the season. No season is clearly better or worst than the others, except for the maximum value of autocorrelation for which summer is worse than the other three seasons. When one considers the models separately, summer happens to be the worst season in terms of autocorrelation of residuals in several models (C4I, CNRM, METO-HadQ16) but the results are more puzzling for the other models (cf figure 17).

- **Normality:** A Shapiro-Wilk test shows that residuals are normally distributed whatever the regression and the model considered (figure 16 provides an illustration of this result).

- **Spatial distribution of residuals:** No pattern in the spatial distribution of residuals and of their standard error can be observed, meaning that the regression performs quite well on the whole country and that no region causes any specific problem (cf figure 2).

3.2.3 Other criteria for the quality of the regression

- **The R^2:** The coefficient of determination is quite variable depending on the regression we consider (cf table 1), with a minimum value of 0.14, a maximum of 0.71 and a mean around 0.43 (ensemble means), which is not too bad. One can observe on figure 17 that the R^2 is on average better for winter than for other seasons and worse for summer. For the intermediate seasons, spring is a little bit better than autumn for several models but it is not clear for all of them.

- **The adjusted R^2:** But remember that the R^2 can be artificially increased by adding non-relevant explanatory variables. The computation of the adjusted R^2 (denoted $\bar{R}^2$) shows a decrease of the determination coefficient, that is to say that the share of the variance of precipitations that is explained by the model is actually smaller than the one given by the classical R^2 (cf table 1 and figure 17).

- **The Variance Inflation Factor:** I finally use the VIF to get an indicator of the colinearity between the explanatory variables. One knows that a VIF greater than 10 indicates an important colinearity between variables. Table 1 and figure 17 (the red dashed line has equation $y = 10$) show that the VIFS are huge whatever the model.

[10] The second coefficient for spring is non-significant three times but it is the only one in this case. Some coefficients are non-significant once or (rarely) twice, and most of them are never non-significant.

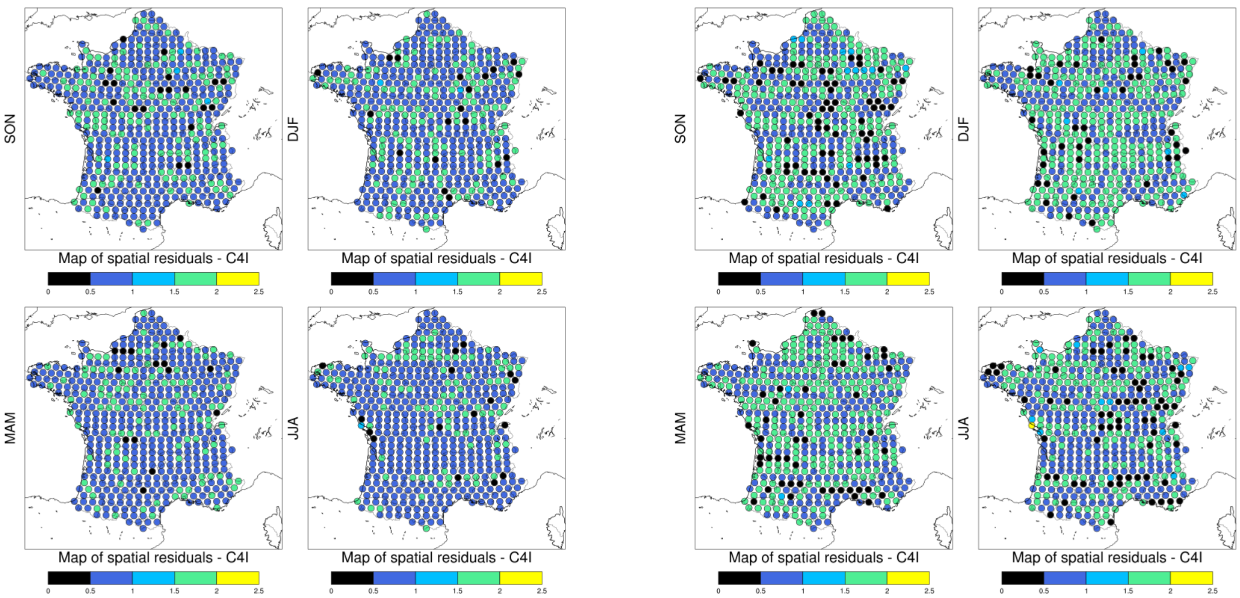


Figure 2: Spatial distribution of residuals (left) and their standard errors (right) obtained using regression estimates for each grid point and each season for model C4I

	AUTOCORRELATION				**VIF**			
	SON	**DJF**	**MAM**	**JJA**	**SON**	**DJF**	**MAM**	**JJA**
Mean	0.23	0.21	0.23	0.24	7380	14380	31949	4292
Min	0.07	0.06	0.08	0.08	5	26	9	39
1st quartile	0.18	0.17	0.20	0.18	117	199	109	446
3rd quartile	0.26	0.26	0.26	0.29	2570	4408	1329	2795
Max	0.48	0.48	0.45	0.57	178876	401707	1371710	77480
	R^2				**ADJUSTED** R^2			
	SON	**DJF**	**MAM**	**JJA**	**SON**	**DJF**	**MAM**	**JJA**
Mean	0.42	0.47	0.44	0.38	0.29	0.38	0.31	0.24
Min	0.16	0.14	0.22	0.17	-0.02	-0.05	0.06	-0.01
1st quartile	0.37	0.42	0.40	0.33	0.24	0.29	0.27	0.19
3rd quartile	0.46	0.53	0.48	0.43	0.35	0.43	0.37	0.31
Max	0.64	0.71	0.61	0.57	0.57	0.64	0.53	0.48

Table 1: Summary statistics on the autocorrelation of regressions' residuals, the VIFs, the R^2s and the adjusted R^2s of the regressions for all models by season

3.2.4 Conclusions of the regression analysis

The GLS regression has a good quality in terms of residual's behavior and significance of estimates, but shows an important problem of colinearity of the explanatory variables. Therefore, the coefficient estimates are biased and thus not the best ones that could have been expected for this first step of precipitation downscaling. Moreover, even if we deal with a large sample, one can not state that the estimates are consistent because of this important colinearity. One easily guesses that this inconsistency and inefficiency of estimates at the first step of the downscaling procedure will bring about imprecisions in the second step too. Indeed, if the precipitations are not well forecast by the regression, DSCLIM might choose wrong analog days for these precipitations (or at least not the best ones).

3.3 Analysis of the downscaled precipitations

I will now carry out a deeper analysis of the final downscaled precipitations. As I wrote before, I use now the stochastic resampling method to select the final analog date from which I will get the downscaled precipitations, while I used the uniform selection method in the internship summary submitted in June 2016. I will provide some pieces of comparison between the two selection methods for the analog days ; moreover, the figures provided in the intermediate report are available in the appendices.

3.3.1 Time-series analysis

Time series I plot time series of inter-annual precipitations (average precipitation by season for each year) for pseudo-observations and analogs on both past and future periods (cf figures 18 and 19). The results are variable depending on the model one chooses to work with, but one can at least remark one particular point: most of the time the analog precipitations underestimate the observed precipitations, more precisely they do not fit the peaks well. This is particularly observable for models KNMI, METNO-BCM, MPI in the past period and for DMI-BCM, METNO-BCM and MPI in the future period for both uniform selection and stochastic resampling methods. Moreover, the choice of the model appears to be crucial at a first glance on these time series. Indeed, the amplitude of both observed and downscaled precipitations may double up from one model to another (see DMI-ECHAM5 and OURANOS, one can observe gaps up to almost 5 mm/day for a given season). Thus the ability of DSCLIM to fit the observed precipitations is limited for both periods, even if the past period is slightly better than the future period. These results do not vary a lot when one switches from the uniform selection of the analog date to the stochastic resampling method (for a comparison between the two methods, see the figures provided here and figures 29, 30 and 27). Nevertheless one can see that the underestimation is less important when using the stochastic resampling method.

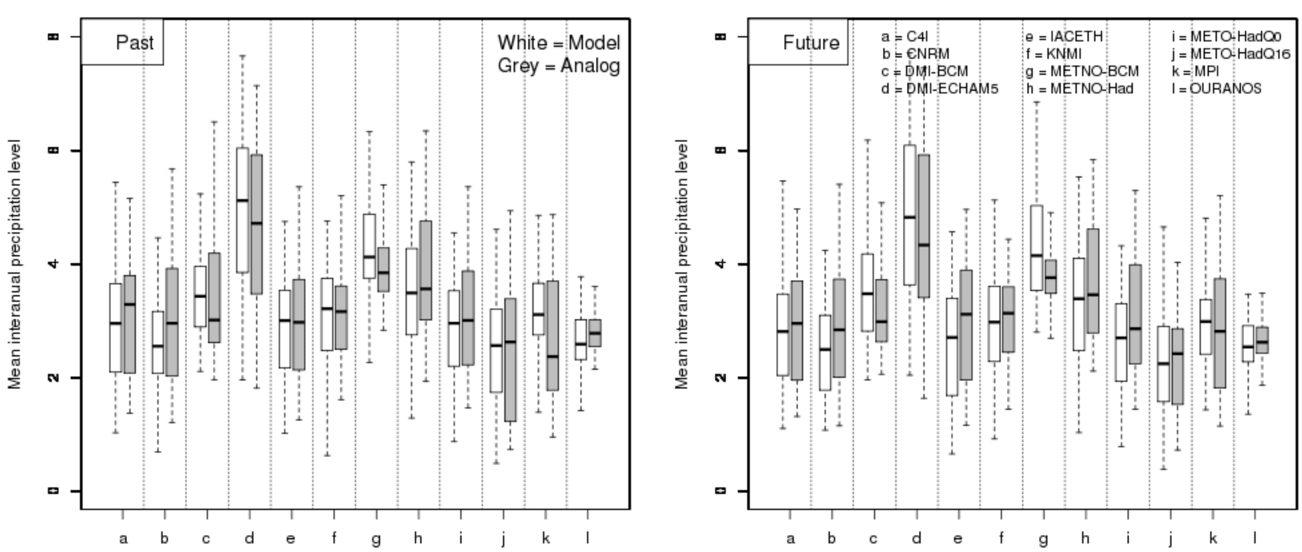


Figure 3: For each model and for both past and future periods, boxplots of the values of the observed and analog inter-annual precipitations. Nota: outliers are not represented.

Temporal correlations The results are puzzling here too depending on the model. While some models perform quite well[11] (C4I, IACETH, METNO-Had, METO-HadQ0, METO-HadQ16, OURANOS, except sometimes for JJA), some show a majority of negative average inter-annual correlation for some seasons (CNRM,DMI-BCM,DMI-ECHAM5, KNMI, METNO-BCM, MPI depending on the season). Less negative correlations are observed in DJF compared to JJA and intermediate seasons, this result being more evident with the uniform selection of the analog date. Past and future periods do not present a huge difference in their average level for most of the models, but the future period shows a wider dispersion. If we look at the cartography of temporal correlation (by season and for both periods), again there is no particular region showing a better correlation than the others. CNRM, DMI-BCM, DMI-ECHAM5, KNMI, METNO-BCM and MPI show a particularly bad (negative or very small) correlation for a large share of grid points, while C4I, IACETH and OURANOS are particularly good. One can not clearly say that one period is better than the other (cf figure 20 for all models and figure 4 for a geographical illustration with model C4I). The results are not better for the stochastic resampling method than for the uniform selection method as you can see with figure 32 and moreover the areas that showed a good fit between analogs and pseudo-observations with the uniform selection are not necessarily the same with the stochastic resampling (see figure 37).

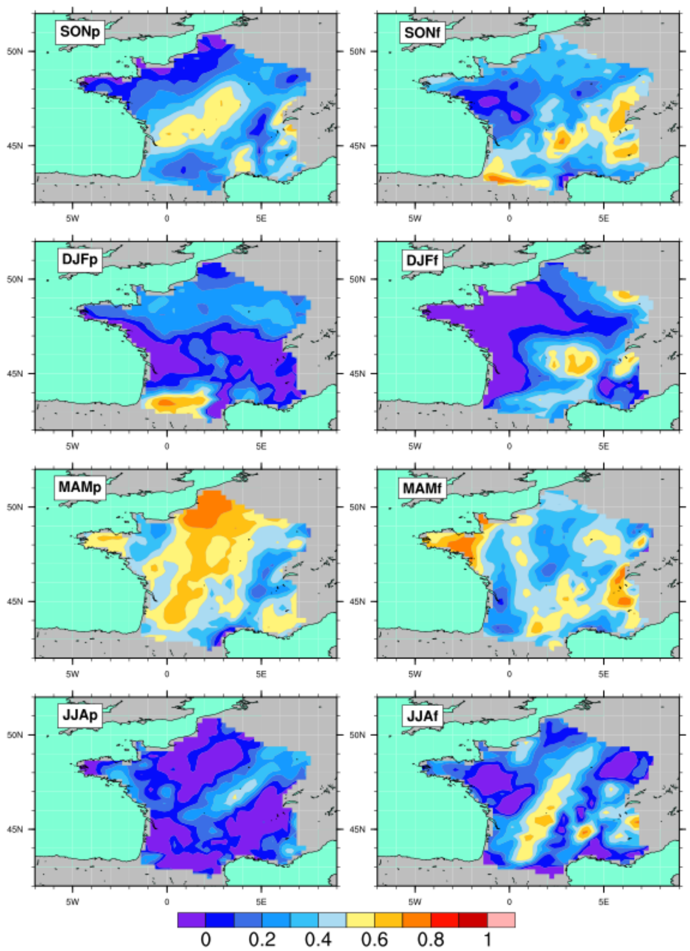


Figure 4: For each grid point, correlation between time series of pseudo-observed precipitations and time series of analog precipitations ; results are mapped for each season and for both past and future periods (the "p" and "f" written after the season's acronym stand for "past" and "future") for model C4I

[11]They perform well especially with the uniform selection method.

3.3.2 Spatial analysis

I will now focus on the spatial correlation between pseudo-observed and downscaled precipitations. For each day I compute this correlation over the whole of France. The results are quite better than the ones presented in the previous subsection. Indeed, as you can see it in figure 21, there is almost no negative correlation and a large share of them are above 0.5. The worst correlations can be found in SON for most of the models. DMI-BCM, DMI-ECHAM5, METNO-BCM and METNO-Had present particularly good spatial correlations. No period seems to be globally better than the other. Here again the results are not significantly different between the two selection methods for analog dates (see figure 33).

3.3.3 Statistical tests

Indices I test the equality of means, variances and distributions of pseudo-observed precipitations and analog precipitations for each grid point, each season and each period. For each test, I build a function that returns 0 if the null hypothesis (equality) is rejected and 1 otherwise. Then for each season and each period I compute the average of these results and get a number between 0 and 1. An average close to 1 indicates that the hypothesis of equality cannot be rejected for a majority of grid points, while an average close to zero indicates a bad performance of DSCLIM at reproducing the mean, variance, or distribution of the pseudo-observed precipitations.

- **Mean equality:** I use a Student's t-test for testing the equality of means. The results are variable depending on the model. For instance, while the index is greater than 0.6 for OURANOS, it goes down under 0.3 for DMI-BCM (JJA future) and even at 0.21 for KNMI (SON future) or 0.25 for MPI (DJF future). For this set of tests the results are quite better in the past than in the future period (cf table 3).

- **Variance equality:** I use a Fisher test for testing the equality of variances. The results are better than for the mean equality tests. Indeed, one never observes indices below 0.7. Here it is not obvious that a period is better than another (cf table 4).

- **Distribution adequation:** I finally use a Kolmogorov-Smirnov test for testing whether the pseudo-observed and analog precipitations are drawn from the same distribution or not. The results are even better than for means and variances: the indices are very often above 0.85 and even often above 0.9. Moreover, it is not scarce to obtain indices equal to 1, meaning that the distributions of the pseudo-observed and analog precipitations are very close to each other. The past period is a little bit better than the future period but the latter shows nevertheless very reasonable indices (cf table 5) despite two indices below 0.8 (0.79 for DMI-BCM JJA and 0.77 for KNMI MAM).

The results presented here (for stochastic resampling) are slightly better than the results observed with the uniform selection method for mean, variance and distribution equality tests (cf tables 6, 7 and 8).

3.3.4 Extreme rainfall analysis

I observed in the time series analysis that the precipitation peaks are most of the time underestimated by DSCLIM, whatever the model. It is therefore interesting to perform a deeper study of these high level rainfall. Precipitations are said to be extreme if they exceed the 99.9th percentile of the precipitation distribution (excluding precipitation levels that are below 1 mm/day). Therefore I compute for each grid point, for each season and each period the 99.9th quantile of the precipitation distribution of this particular grid point (one quantile for the observed precipitations and one for the analogs).

The 99.9th quantiles Not surprisingly the 99.9th quantile for analogs is globally smaller than the same quantile for the pseudo-observed precipitations distribution in both past and future period (cf figure 22 ; figure 5 provides a geographical illustration for C4I). Nevertheless, geographical patterns are globally respected by DSCLIM. Note that the quantiles are extremely different depending on the model we choose to work with. To be convinced one simply has to compare DMI-ECHAM5 and OURANOS: while the former shows severe extreme precipitations, the latter reports a drier weather. One can see in figure 22 that the underestimation of the extreme quantiles by DSCLIM is stronger in JJA than in other seasons (and even stronger for the future period's JJA). When one switches from uniform selection to stochastic resampling, the distribution of quantiles does not change a lot but variations can be observed concerning the spatial distribution of the 99.9th quantiles. One can look at figures 31 and 28 for a comparison of the two methods.

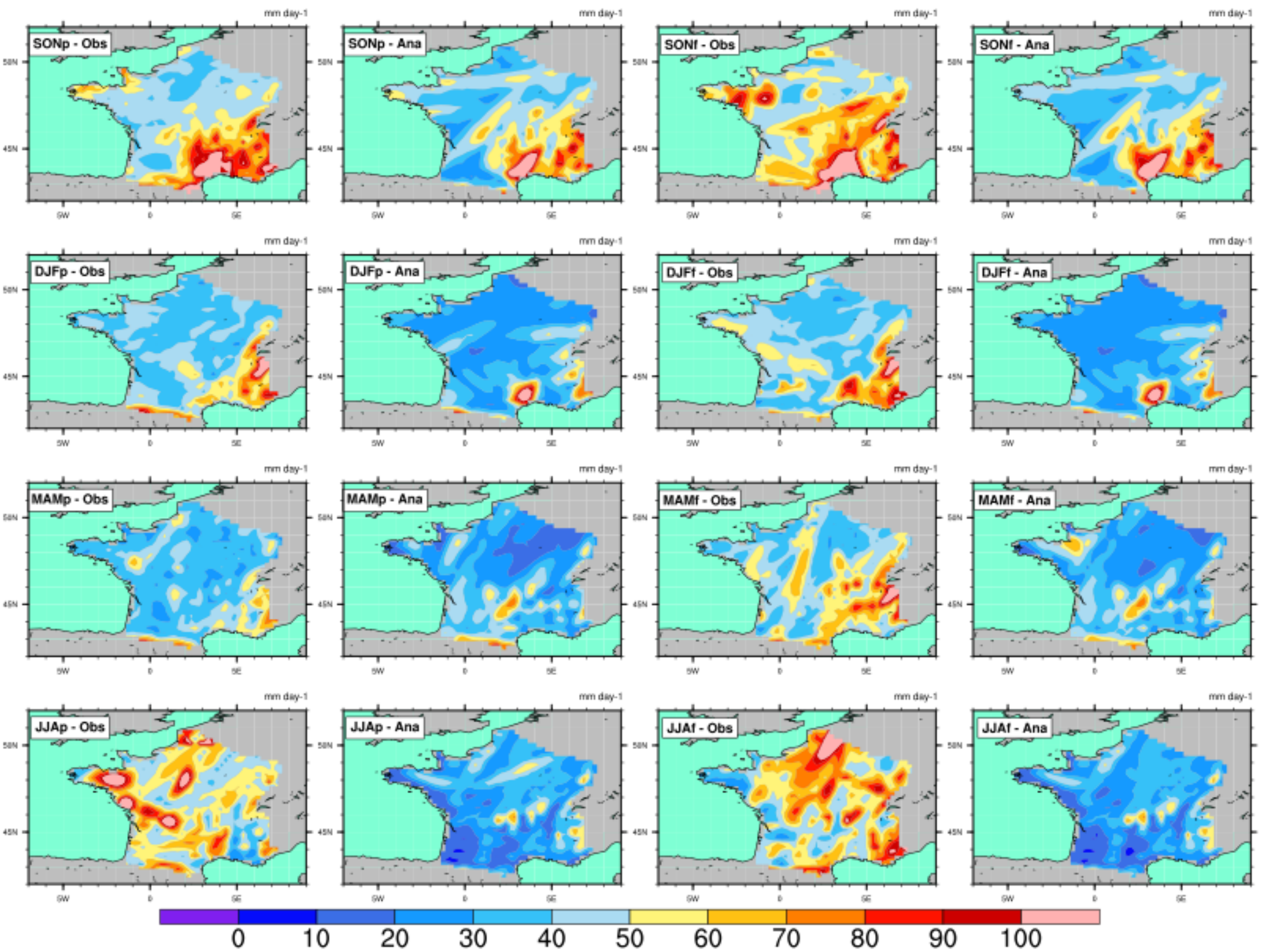


Figure 5: The 99.9th quantiles of the precipitation distribution (excluding precipitation levels that are below 1 mm/day) for each grid point, each season and each period (the "p" and "f" written after the season's acronym stand for "past" and "future", "Obs" and "Ana" stand for "pseudo-Observed" and "Analog") for model C4I.

Changes in extreme quantiles Another interesting point is the change of the extreme quantiles between past and future periods. One can already remark it by comparing both periods in figure 5: while the maps for pseudo-observed precipitations can be quite different, the maps for the analogs are not that different (you can see it for C4I in figure 5 but it is the case for all models). This result is confirmed when we consider the maps of absolute and relative changes in the quantiles between past and future (figure 23). While the absolute change in quantiles is most of the time greater than 6 mm/day for observed precipitations, it is most of the time smaller than 0.5 for analog precipitations (except for the Alps in several models and for other regions in few models where it can go up above 6 mm/day). As for relative changes, they are quite restricted (smaller than 1%) in most of the models, especially for the analog precipitations, with some high changes in few models (but isolated). One can also wonder whether precipitations considered as extreme in the past will still be considered so in the future. Figure 24 shows the projection of the past 99.9th quantile on future precipitations. The change is not very noticeable for pseudo-observed precipitations. Indeed, for all models, the precipitation level corresponding to the 99.9th quantile in the past corresponds to a quantile greater than 99 in the future. On the other hand, one observes more variation for analog precipitations. The 99.9th quantile today corrspond to a quantile between 96 and 99 in most of the models and even below 0.94 for some of them (CNRM DJF, DMI-ECHAM5 JJA, IACETH DJF and JJA, METO-HadQ16 DJF, MAM and JJA, MPI MAM are the most noticeable). Thus it seems that DSCLIM has a tendancy to overestimate precipitation increase in the future. Moreover, it seems that the stochastic resampling method increases this overestimation when one compares it to the uniform selection method. Indeed, one can rarely observe that the 99.9th quantile today correspond to a quantile below 0.96 in the future (one can see an illustration for C4I with figure 38 but it is the case for all models).

3.3.5 Drought analysis

The natural rest of the analysis is the study of extreme events that are at the opposite of extreme precipitation, that is to say the dry extremes (droughts). Here the definition of what a dry event is is not as obvious as in the previous section. The most used index to delimit dry events is a threshold of 1mm/day (cf [12]). Days for which precipitations are below this threshold are said to be dry. But in a perfect model framework analysis, things are not that simple. Indeed, the twelve models do not work the same way and physical links between climatic variables are not identical from one model to another. Therefore this definition does not take in account differences between models and can lead us to predict high drought rates for models that are by construction dryer than others (we saw with the study of extreme rainfall that some models show a 99.9th quantile higher than others), which will give us biased results. Thus I choose to use an alternative definition of drought in the perfect model framework: every day that shows a precipitation level below a given quantile of the precipitation distribution[12] is considered to be a dry day. I compute a quantile for each grid point, each season and both past and future period.

Choice of the quantile I compute for each model several quantiles to determine which one is the more relevant for our analysis: the 1st, 5th, 10th, 15th and 20th quantiles. For the first two, most of the grid points show a quantile of zero mm/day and one cannot see any differece between the models, except for OURANOS, which shows greater quantiles than other models. For the 10th quantile, one observes some differences between models and the dispersion of the different quantiles for a given model is not too large, but for seasons SON, MAM and JJA many quantiles are still equal to zero. For the 15th and 20th quantiles, the dispersion is wider and that's why I choose not to use these quantiles. Indeed, the difference in the threshold that determines dry days should not be too important between the grid points in a given model. Thus I think that the most relevant definition for the dry days is the 10th quantile. Nevertheless, on should be aware that this definition has some drawbacks: indeed, with this definition, the number of dry days is not necessarily more important in JJA than in DJF since the quantile is smaller in JJA than in DJF. This result is not a very intuitive, but the aim of this study is not to give a very realistic definition of drought but to compare pseudo-observed and analog precipitations (and thus pseudo-observed and analog drought).

[12] Now one considers the precipitation distribution including all precipitation levels, even those that are below 1 mm/day.

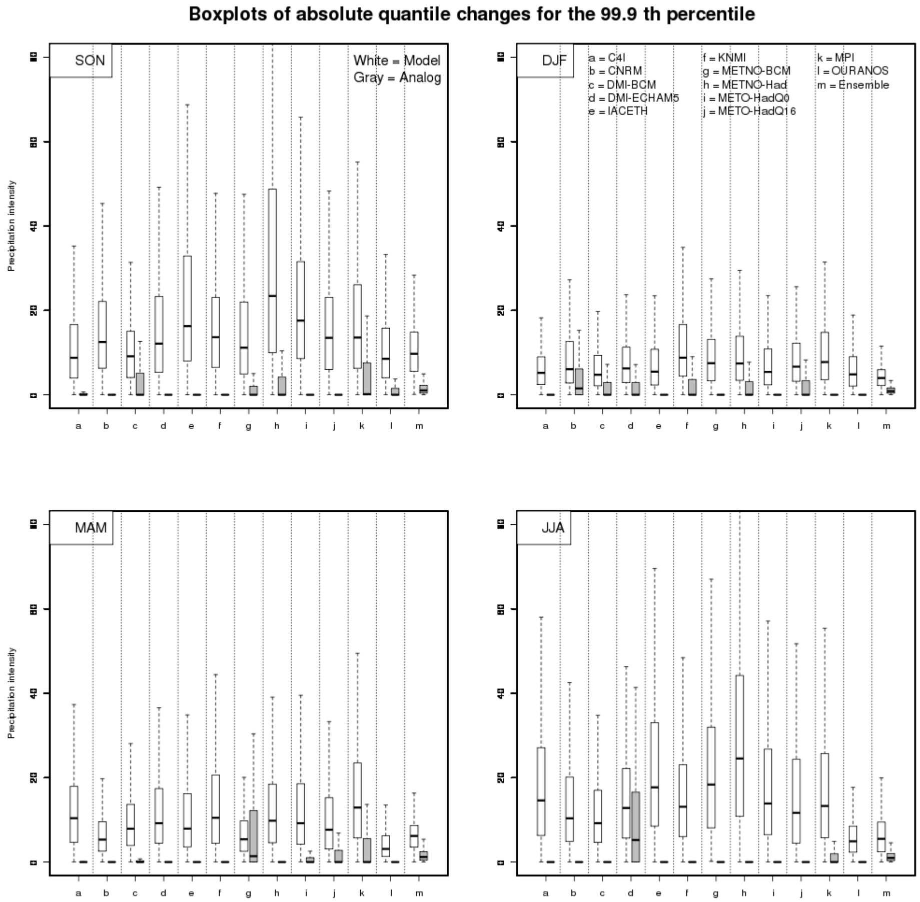


Figure 6: Absolute change between the 99.9th percentile of the past period's precipitation distributions and the 99.9th percentile of the future period's precipitation distributions (in absolute value) for each model and each season: boxplots for absolute changes in the pseudo-observed precipitation and for the absolute changes in the analog precipitations, for model C4I. Nota: outliers are not represented.

Therefore this definition will be sufficient to fill this objective despite its drawbacks. Boxplots of the 10th quantile are given with figure 25.

Surface of drought To assess the ability of DSCLIM to predict dry events, I consider the surface covered by dry events every day. For each day, I count the number of grid points that show a precipitation level below a given quantile of their distribution. Figure 26 presents the boxplots of the number of dry points for the quantiles 0.01, 0.05, 0.1, 0.15 and 0.2 for model C4I. The main conclusion that can be drawn from these results is that DSCLIM underestimates the surface covered by drought each day. Indeed, the number of dry points is on average greater for pseudo-observations than for analogs in most of the seasons for all models (there are some exceptions, for example C4I, DMI-BCM, MPI, METO-HadQ16, METO-HadQ0 for MAM or METNO-BCM for DJF future, METNO-Had for JJA past). Moreover, most of the time the dispersion of the pseudo-observed number of dry points is wider than the analog one: the minima are always the same but the observed maxima and third quartiles can be significantly higher than the analog maxima and third quartiles. One can see it on figure 26 for DJF (for C4I) but the difference is even noticeable with the same season for models IACETH or OURANOS, and for all seasons with METNO-BCM. Concerning the average number of dry grid points per quantile, the pseudo-observed and analog averages can be more or less close to each other depending on the season and the model one considers. For C4I, these averages are

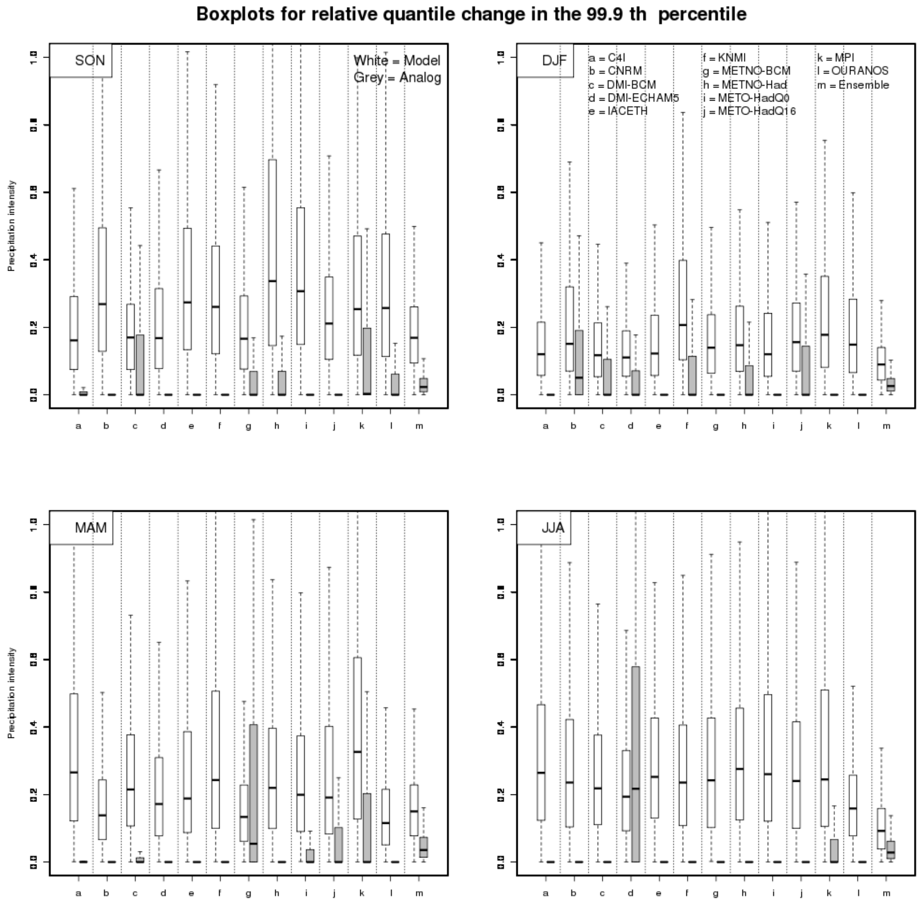


Figure 7: Relative change between the 99.9th percentile of the past period's precipitation distributions and the 99.9th percentile of the future period's precipitation distributions (in absolute value) for each model and each season: boxplots for relative changes in the pseudo-observed precipitation and for the relative changes in the analog precipitations, for model C4I. Nota: outliers are not represented.

almost identical for DJF and MAM, and more different for SON and JJA. The difference is more evident when one consider for instance DMI-ECHAM5 in JJA, where it is of more than 150 points for quantiles 0.05 and 0.10 and around 200 points for quantiles 0.15 and 0.20. Furthermore the results are very different depending on the model: one season can be well estimated by DSCLIM for a given model and very badly in another one. Thus it is difficult to state some definitive conclusions on the quality of climate projections produced by DSCLIM as it tightly depends on the model one chooses to work with.

3.3.6 Conclusions of the results' analysis

To conclude with this section, one can say that the performance of DSCLIM to make climate projections on future rainfall in a perfect model framework is strongly conditioned by the model that is used to build pseudo-observations. Most of the temporal correlation between pseudo-observed and downscaled data is puzzling, but the spatial correlation is better and the statistical tests show a good performance of DSCLIM in reproducing the precipitation distributions. The main weakness of the downscaling method is the climate projection of extreme rainfall and drought, as I explained in the last subsection: they are underestimated most of the time and their reproduction is more or less good depending on the model.

4 Assessment of the performance of DSCLIM with real data

Once we have assessed the advantages and drawbacks of DSCLIM in a perfect model framework, we can now have a look on its performance with real data. The period that is used to carry out the comparison of observed and analog precipitation is 1979-2010, which is also the learning period.

First of all I will assess the performance of DSCLIM to reproduce ERAI reanalyses. Then I will check whether one of the main hypothesis of DSCLIM holds or not, that is to say the preservation of variance when one projects PSL on EOFs. Finally I will investigate how DSCLIM performs with some of the CMPI5 models compared to SAFRAN observations.

4.1 Downscaling with the ERAI reanalyses

4.1.1 Learning phase : the regression

Coefficient estimates and residuals The results of the learning phase are quite good for ERAI reanalyses. As for the regression analysis in the perfect model framework, I perform a Student's t-test for testing the significance of coefficient estimates. They are all significant for all seasons, at the 99% level. Concerning the residuals, they are centered on zero and do not have a too wide variance. Some Shapiro-Wilk tests lead to the conclusion that they are normal for all seasons. Moreover, the autocorrelation of residuals is not too large on average (see table 2) and is comparable to the one obtained in the perfect model framework (see table 1). Concerning the spatial distribution of residuals, one cannot observe any specific pattern when mapping the residuals or their standard deviations (cf figure 8).

Other criteria for the quality of the regression As in the perfect model analysis one can observe an adjusted R^2 smaller than the original R^2 for all seasons, despite it is only slightly smaller. The colinearity problem reappears when one looks at the VIFs (cf table 2): indeed, all the VIFs are greater than 10 even if they are smaller than the ones observed in the perfect model framework.

	AUTOCORRELATION				**VIF**			
	SON	**DJF**	**MAM**	**JJA**	**SON**	**DJF**	**MAM**	**JJA**
Mean	0.23	0.24	0.23	0.24	226	397.60	1988	2378
Min	0.05	0.05	0.09	0.10	26.34	72.68	118.70	93.99
1st quartile	0.19	0.17	0.19	0.21	69.90	165	840.70	279.90
Median	0.23	0.24	0.23	0.25	187.90	254	1828	559.20
3rd quartile	0.28	0.29	0.29	0.27	344.20	550.20	2447	3497
Max	0.44	0.47	0.35	0.40	523.20	1109	5219	10480
	R^2				**ADJUSTED** R^2			
	SON	**DJF**	**MAM**	**JJA**	**SON**	**DJF**	**MAM**	**JJA**
Mean	0.46	0.53	0.44	0.38	0.40	0.48	0.38	0.32
Min	0.32	0.31	0.29	0.18	0.26	0.24	0.22	0.10
1st quartile	0.42	0.49	0.41	0.37	0.37	0.44	0.35	0.30
Median	0.46	0.53	0.45	0.39	0.40	0.49	0.39	0.33
3rd quartile	0.49	0.57	0.47	0.41	0.44	0.53	0.42	0.35
Max	0.56	0.64	0.56	0.49	0.51	0.60	0.52	0.44

Table 2: Summary statistics on the autocorrelation of regressions' residuals, the VIFs, the R^2s and the adjusted R^2s of the regressions for ERAI reanalyses by season

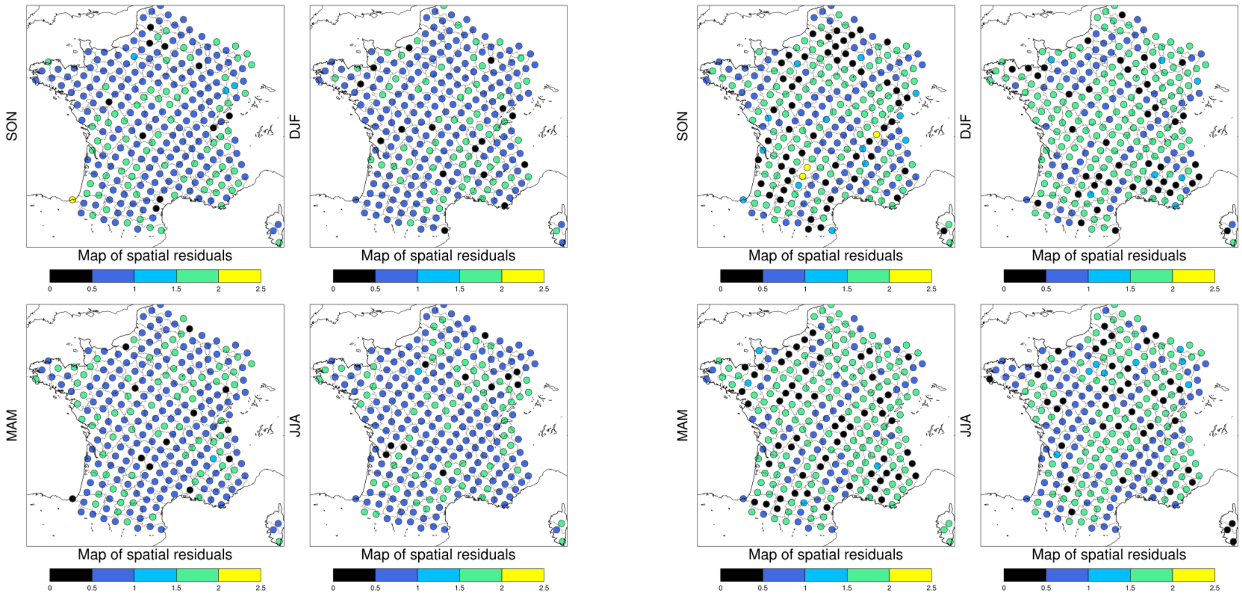


Figure 8: Spatial distribution of residuals (left) and their standard errors (right) obtained using regression estimates for each grid point and each season for ERAI reanalyses

4.1.2 Downscaling phase

Correlations and equality tests Not surprisingly temporal interannual correlations are not very high. As one can see in table 9, this correlation can go high (0.79) but 75% of its values are below 0.44 which is not very high, as well as the average temporal correlation which is of 0.3. Spatial correlations are better since a majority of them are above 0.6. Table 10 shows that DSCLIM is not very efficient in reproducing averages (except for DFJ) and that variances and distributions are quite well projected for the period 1979-2010[13].

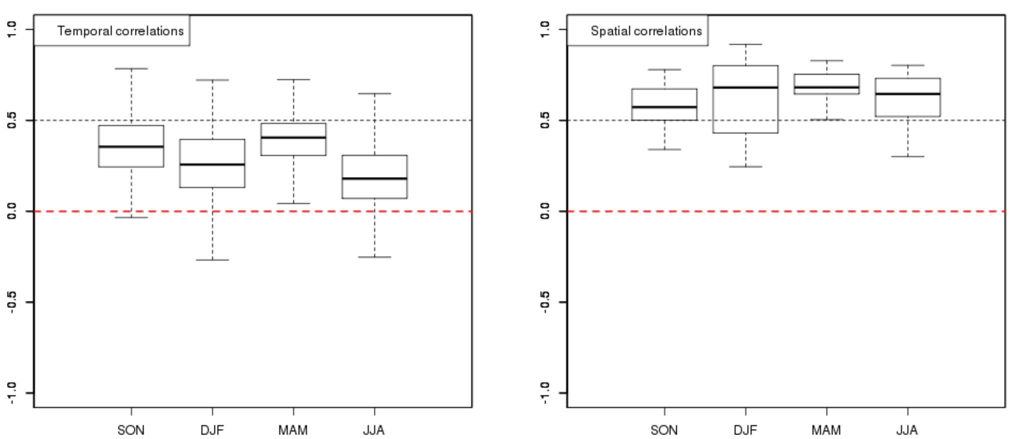


Figure 9: Boxplots of the temporal (left) and spatial (right) correlations between observed (SAFRAN) and downscaled precipitations (from ERAI reanalyses) by season. Nota : outliers are not represented

[13]**NB:** The analog precipitations are obtained using only the stochastic resampling method for selecting the final analog day.

Extreme events Concerning the extreme droughts, I try to compute different quantiles as I did above. But it seems that even the quantile 0.20 of both observed and analog precipitations is set to zero, so that it is impossible to observe any difference between the quantiles. Concerning the extreme rainfall, figure 10 shows that DSCLIM underestimates the 99.9th quantiles with ERAI reanalyses as well as in the perfect model framework. The season for which the underestimation is the lowest is MAM and it is the biggest in DJF.

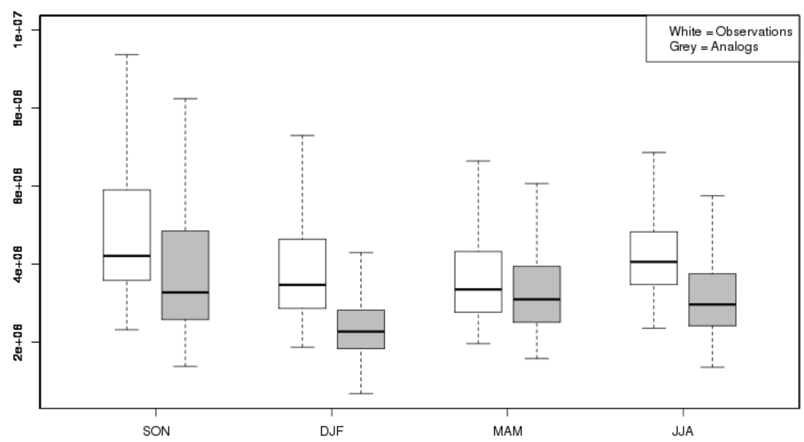


Figure 10: Boxplots of the 99.9th quantile of precipitation distributions (excluding precipitation levels below 1mm/day) for ERAI reanalysis by season ; white boxplots are for reanalyses and grey for analog precipitations. Nota : outliers are not represented.

Finally I drew the QQ-plots of observations against analogs for each season to have a global view of their respective distributions (cf figure 11). As expected, results are particularly bad in JJA and for other seasons the goodness of fit between observations and analogs is satisfying except for extreme rainfall. Moreover it seems that there is a ceiling for the values of downscaled precipitations. This highlights a sampling problem, that is to say that since the extreme precipitations are by definition more rare than the other precipitation levels[14], extreme events are underrepresented and therefore the fit between observations and analogs can not be as good as for precipitation levels that are better represented.

4.2 Loss of variance with EOFs

I explained in the first section of this report that the construction of weather types is based on the EOFs of the APSL. When one wants to associate a weather type to a given day, one looks at the APSL of this particular day. But to be able to associate this day to a weather type, the weather type and the APSL have to be defined on the same EOFs-basis, otherwise it has no sense to compare them. But if one computes the EOFs of the APSL for each model one wants to work with, they may be very different and then it will be hard to carry on comparisons. Thus we decided to compute principal components of APSL for each model and then to project them on the EOFs obtained with ERAI reanalyses, in order to get comparable APSL fields. Therefore an important hypothesis of DSCLIM is the following: the conservation of variance when one projects the APSL principal components on the ERAI EOFs. This hypothesis does not obviously hold and that is what we will verify in this section.

[14]Indeed, in our framework extreme rainfall only represent 0.1% of the rain events since we choose to define them as precipitation levels that exceed the 99.9th quantile of precipitation distributions.

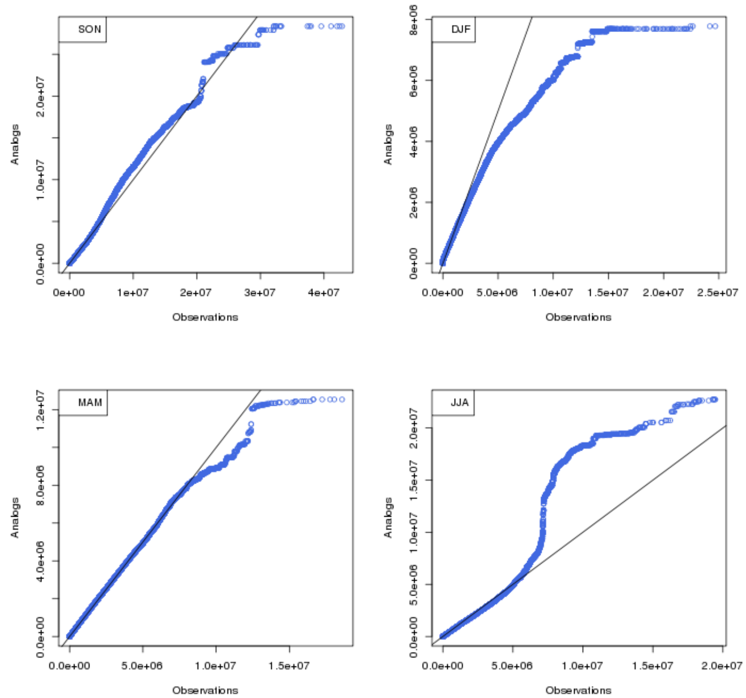


Figure 11: QQ-plots of the analog precipitations obtained with ERAI reanalyses against the SAFRAN observations for each season ; the black line represents the first bisector.

The computation of the projected APSL field is done in the following way. Let eof^{erai} be a vector of EOFs derived from ERAI reanalyses, let m be a particular climate model and pc^m the vector of the principal components of the APSL produced by m. eof^{erai} has dimensions (nd, lat, lon) where nd is the number of EOFs required (10 in the case of DSCLIM), lat and lon the latitude and longitude ; pc^m has dimensions $(time, nd)$. The projected APSL field of the model, denoted psl^m, is obtained as follows :

$$psl^m = \sum_{i=1}^{nd} pc_i \times eof_i \tag{2}$$

where $pc_i = pc(time, i)$ and $eof_i = eof(i, lat, lon)$ so that psl^m is of dimensions $(time, lat, lon)$.

Therefore I compute the variance of both the original and projected APSL[15] for the following models : ACCESS1-3, bcc-csm1-1-m, BNU-ESM, CanESM2, CMCC-CM, CNRM-CM5, GFDL-CM3, IPSL-CM5A-LR, IPSL-CM5A-MR, MIROC5, MPI-ESM-MR, MRI-CGCM3, NorESM1-M. The variance is computed for each of the 270 grid points. For each model, I map the ratio of the projected variance over the original variance, which gives maps of the explained variance for the different models. The mapping of the variances' ratios suggests that the loss of variance is not that important when projecting the principal components of APSL on the ERAI EOFs (cf figure 12 and for further details see tables 12 and 13). Indeed, at least 75% of the variance of each APSL is explained when projecting if on the ERAI EOFs. Moreover the biggest loss of variance is observed in the Alps for all models as well as for the ensemble mean and even there the loss of variance is not as important as expected.

[15] By "original" I mean the variance of the APSL of the model without any projection on the EOFs ; on the other hand, by "projected" I mean the variance of the APSL after its projection on the ERAI EOFs.

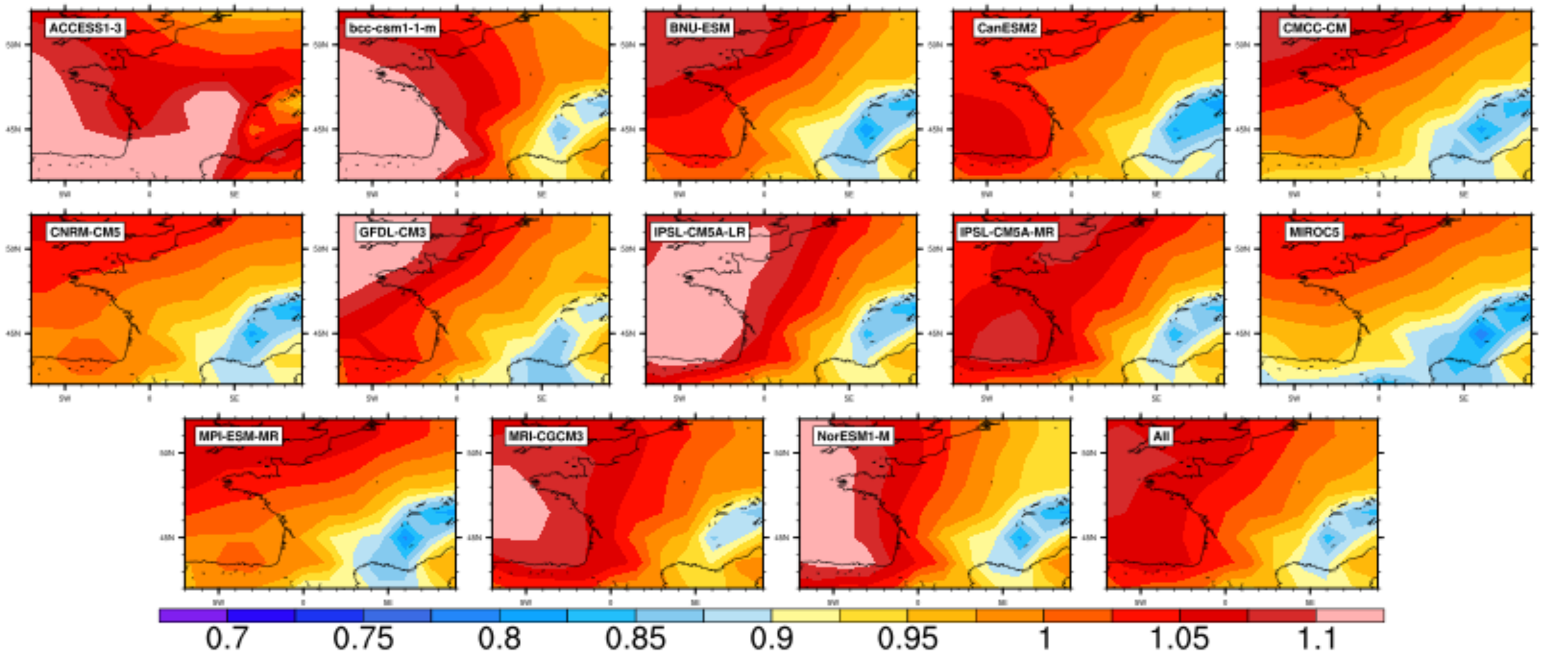


Figure 12: Ratio of the projected variance (variance of the APSL projected on ERAI EOFs) over the original variance (variance of the total field APSL) for each grid point, for each model and ensemble mean ("All").

4.3 Downscaling with CMIP5 models and SAFRAN observations

Now I will analyze the performance of DSCLIM in reproducing the long term rainfall using CMIP5 model data. I will compare the downscaled precipitations[16] obtained with these model data to the SAFRAN observations. But unfortunately it is impossible to compare them at the daily scale. Indeed, for these models, the historical climate simulations are initialized at a given date in the past and then allowed to evolve in the future (until 2010) in line with prescribed forcings (here the CO_2 forcing). These climate projections are not periodically adjusted with updated information about the state of the atmosphere (unlike reanalyses). Therefore these projections are not designed to reproduce the daily sequence of precipitations but only long-term (multi-decadal) climate statistics such as moments and distributions[17].

Therefore I perform several tests on observations and model data: as above, a Student test for testing mean equality, a Fisher test for testing variances equality and finally a Kolmogorov-Smirnov test for testing distributions adequation. The results are globally good, especially for variances and distributions (cf table 14). The results of the Student tests are variable depending on the model but one can notice that they are better in DJF than in other seasons, and especially JJA that shows the worst results. This pattern is less obvious for Fisher tests but one can see it with the Kolmogorov tests as well.

4.4 Conclusions of the real data framework

The main provision of this fourth part is to show that the loss of variance dragged by the projection of APSL principal components on ERAI EOFs is not too important, so that it is not dummy to state the hypothesis of

[16]**NB:** The analog precipitations are obtained using only the stochastic resampling method for selecting the final analog day.

[17]For a brief overview one can see [13] ; for further details, see [1]. The idea is that the forcings are the same in the models and the reality but there is no reason that the cycles (and especially the oceanic cycles) are the same: the climate is globally the same but the variations inside a cycle do not happen at the same time so that a multi-decadal comparison is the only one possible.

variance conservation. Besides the diagnostic tools applied to ERAI and other models show similar results as the ones obtained in the perfect model framework: the main difficulty encountered by DSCLIM in the regression phase is the colinearity ; in the downscaling phase it is the downscaling of extreme precipitations.

5 Assessment of an other downscaling method

The work I did until now confirmed some drawbacks we suspected on DSCLIM. That is why I am now interested in looking at the results obtained with an other downscaling method that rests upon the analog method too but without using the regression step. In this fifth part I will use the Dayon method described in [8].

5.1 The downscaling method

The general idea[18] of this second downscaling method is the following. The analog method is applied with four variables: instead of using only PSL as predictor, one uses PSL, the 2m air temperature (TAS), the totals total index (TTI[19]), the humidity flux at 850 hPa (HFLUX) and the specific humidity at 850 hPa (HUS[20]). To select the final analog date, one musts:

- compute the distance between the downscaled date and the days of the learning period for each predictor (one uses the Teweles and Wobus score (TWS) for PSL and the Euclidean distance for other predictors)[21],
- standardize these distances using means and standard errors of the predictors on the training period,
- sum all the distances for each date,
- and finally select the minimum distance among these sums, which gives the final analog date.

Note that there is no weather-type approach here and that the current year is excluded from the analog dates. Moreover, a window of more or less 60 days around the downscaled date is respected so that one can not associate days from too different periods (for instance associate an analog day in December to a day in March).

The main improvement of the method is that the use of several predictors instead of just one[22] allow to precise the statistical link between large-scale and finer-scale variables: indeed, several physical processes are involved in the link between these two scales and one easily guess that several processes can not be represented by a single predictor. Another important advantage is that one makes less hypothesis (which are often not verifiable) here than with the DSCLIM method. First of all there is no hypothesis made on the regression (is the regression linear or not, what variables should it take into account...) ; furthermore the absence of regression avoids biases in the coefficient estimates in the first step of downscaling and the colinearity problem does not matter anymore. Moreover, the potential problems arising from the weather type classification are overtaken here as this classification is not used. Indeed, the relevant number of weather types is not an easy question to solve: taking too little types does not give a proper description of reality and taking too many of them gives weather types very close to each

[18] For a further description of the method, see [8].

[19] The TTI consists in 2 components: the vertical totals (VT) and the cross totals (CT) where VT represents the static stability between 500 and 850 hPa and CT the 850 hPa dewpoint. We have $TTI = VT + CT$ where $VT = T(850hPa) - T(500hPa)$ and $CT = Td(850hPa) - T(500hPa)$ where T is the temperature and Td the dewpoint temperature. A TTI between 45 and 50 indicates possible thunderstorms, most likely and possibly severe thunderstorms when it is between 50 and 55 and most likely severe thunderstorms when it is between 55 and 60.

[20] The specific humidity is the ratio between water vapor mass and air parcel's total mass.

[21] When we compare the PSL fields we are not interested in the absolute field but in the gradient (which is more interesting in terms of atmospheric circulation), that is why the TWS is more interesting than a distance based on the absolute value. On the other hand, for the other predictors we are interested in their absolute value so the Euclidean distance is used.

[22] Note that it is not impossible to include several predictors in the DSCLIM method, but it is still not implemented in the code. Moreover, the way they should be included in the classification-regression-analog process is not obvious.

other and thus the classification is not discriminant (the statistical link between large scale and finer scale variables is weak). Moreover the absence of weather types implies that every day is potentially a good analog day since there is no frontier between weather types (with DSCLIM[23] the weather type classification implies that the days located at the frontier between two types are not good analog days since they do not clearly belong to one type). Furthermore, weather types have a regional aspect: they are different for the different regions, so that one has to recompute them and to find a new optimal number of types when working on another region. The absence of weather types also implies that one is not forced to separate the year between fixed seasons: with this new method a moving season is used, that is to say the 60-days window on either side of the downscaled day. Indeed the use of a moving season allows to maximize the sample size: if not enough data are available all it takes to enlarge this window, what is not possible to do with predetermined seasons. Last but not least, this method allows to preserve the spatial coherence of the downscaled variable and of the climate variables between each other, as well as DSCLIM.

But using a pure analog method has also some drawbacks. Like any resampling method, the loss of variance is an important matter. Moreover, events that did not happen in the training period are not reproducible with an analog method, which implies that one needs to perform some post-processing for temperatures as they are predicted to be higher in the future than today (but the available observations do not allow to project higher temperatures than the ones observed today). Another drawback of this method is that it is more sensitive to the bias in the simulated predictors (for instance PSL). Some areas are also not well predicted by the method as they are climatologically too different from the global area (the Mediterranean basin for instance). Furthermore, the allocation of the same weight to each predictor may be disputable: indeed, with this method one considers that each variable is as important as the others in the building of the statistical link between large-scale and finer-scale variables, which may not be the case depending on the region of interest. For instance in the Mediterranean basin LCL and HFLUX are often more relevant than PSL to explain the link between climatic variables at both scales.

We will see in the following if the results of this downscaling procedure are better than the ones obtained using DSCLIM.

5.2 Assessment of the method

5.2.1 Brief state of the assessment

The main objective of [8] was to test the transferability of the downscaling method in the future climate. The perfect model framework is used here again to test the stationarity hypothesis. The downscaling method is applied to the twelve models I used in the third part of this report[24]. It is shown that the simulated changes in precipitations are quite well reproduced by the downscaling method, with a high spatial correlation for all seasons. Moreover, the inter-model spread is also well simulated by this method. One of the main conclusions of the paper states that to be efficient, a statistical downscaling methods has to take into account at least dynamical and thermodynamical evolutions to be able to make consistent precipitation projections. Moreover, it is shown that the five predictors (PSL, TAS, TTI, HUS and HFLUX) are needed for the method to be transferable in the future climate, what was not done with DSCLIM.

[23]DSCLIM takes into account the fact that days are at the frontier of two weather types in the regression phase but not in the choice of the analog dates.

[24]Recall: the models are C4I, CNRM, DMI-BCM, DMI-ECHAM5, IACETH, KNMI, METNO-BCM, METNO-Had, METO-HadQ0, METO-HadQ16, MPI and OURANOS.

5.2.2 Other assessment tools

While the assessment made in [8] used different combinations of predictors (the ones I cited above being the best one according to the paper), I will focus on the method I described at the beginning of this section and its assessment in a perfect model framework, applying the same diagnostic tools I developed in the previous parts of this report.

Nota: I had some problems with the time definition of the data sets I used for this analysis ; I solved most of them but had no time to solve it for DMI-BCM (future period only) and OURANOS, that is why they are missing in the figures and tables that are presented in this section and the appendix. For the same reason I had no time to finish the extreme events analysis but I wrote the scripts so that it will be carried out by other members of the team soon.

Time series and temporal correlations The results from a temporal point of view are indisputably better than the ones obtained with DSCLIM. Indeed the values taken by the time series are on average only slightly different between pseudo-observations and analogs for all models. The dispersion of these values is roughly the same for both sets too (see figure 13). Moreover, temporal correlations are also better than the ones obtained DSCLIM (see figure 39) as they are on average above zero for all models (except METNO-Had future) and both periods are comparable.

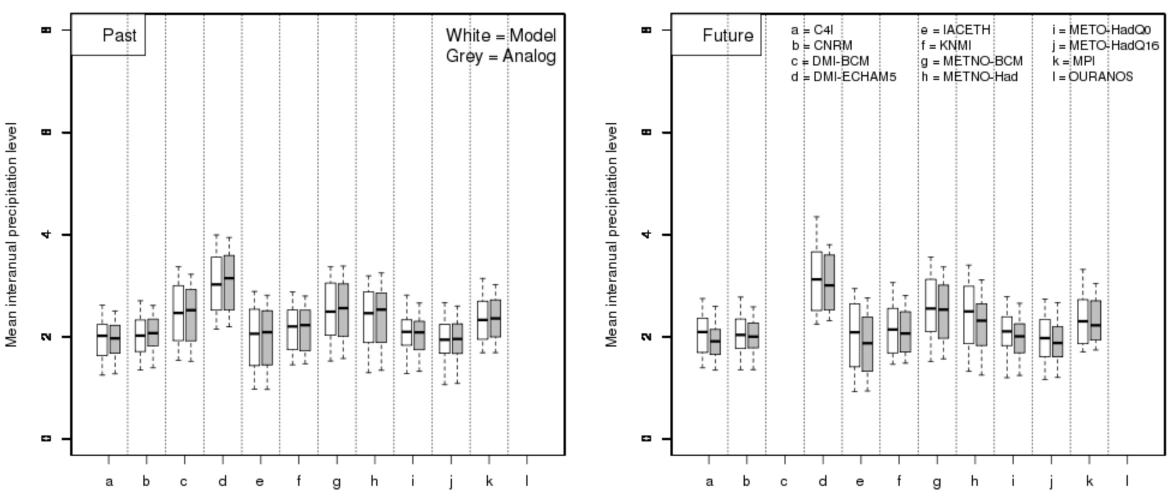


Figure 13: For each model and for both past and future periods, boxplots of the values of the observed and analog inter-annual precipitations computed using the Dayon method. Nota: outliers are not represented.

Spatial correlations The improvement of the results is confirmed by the results obtained with the statistical correlation between pseudo-observed and analog precipitations. As you can see with figure 40, most of the correlations show values above 0.8 and all of them are above 0.6.

Statistical tests The results of the statistical tests are presented in tables 15, 16 and 17. The way I performed these tests is the same than in the previous subsections. The results of the Student's t-tests are not better than the ones obtained for DSCLIM and the results of the Fisher tests are comparable, around 0.7-0.8. A more surprising result is shown by table 17 that presents the results of the Kolmogorv-Smirnov tests. Indeed, they are worst than the ones obtained for DSCLIM to a great extend: while one observed most of the time more than 90% of grid points reproducing well the precipitation distribution with DSCLIM, here this percentage fall down to around 50% for all models and seasons. The results are a little bit better for the future period but even smaller than 70% for most of them.

5.3 Conclusions of the pure analog method analysis

The results I obtained confirm some analyses that were performed by G. Dayon in his paper [8]. The innovation of my work is to provide results on general statistics (equality of means and variances) and on the reproduction of precipitation distributions by the downscaling method. The most interesting result is the quite bad reproduction of distributions by this method.

6 General conclusions

6.1 Conclusions on the basic DSCLIM method

Concerning the regression phase of the downscaling process, we have seen that the main problem is the colinearity between explanatory variables. There exist methods that enable to deal with this problem, for instance the method of partial least squares regression which bears some relation to principal components regression which is itself based on a principal component analysis. Another solution to be tested is to run a pure analog method, which would avoid the bias induced by a non-consistent regression phase. In the rest of the DSCLIM project, the use of several predictors like temperature or humidity indices in the learning phase will also be tested, instead of using only the APSL. An other way to deal with the colinearity problem could be to use the Aikake Information Criterion (AIC) to see what is the optimal number of explanatory variables (parsimony criterion).

Concerning final climate projections, we have seen that precipitations distributions are quite well reproduced by DSCLIM despite some irregular results exist in temporal and spatial correlations. Extreme precipitations are clearly underestimated (which was expected with this type of method), that's why one should now look for the existence of systematic biases and for ways to fit wet extremes better (and their changes between past and future periods in a perfect model framework). Indeed, not only the 99.9th quantile but also its relative change between past and future periods is underestimated by DSCLIM in PMF, when they are important issues for impact studies. The underestimation of drought by the downscaling method is an important issue for impact studies too as underestimating the dry episodes and the surface they cover can have severe consequences, especially in terms of agriculture and public health. Moreover, besides the search of systematic biases and other ways to fit the dry events, a more subtle definition of dry events should be set up so that it can be used for impact studies.

Concerning the application of the method to real data, it has been shown above that the most interesting result is the consistency of the variance preservation hypothesis, which was one of the strongest ones made by DSCLIM.

6.2 Conclusions on the pure analog method

The Dayon method has strong advantages on DSCLIM as it makes less strong hypothesis, in particular on the regression step and weather types. Therefore this method has chances to be more relevant than DSCLIM since the latter lays on non-verifiable hypothesis. The use of a moving season presents a clear advantage too. This pure analog method presents better results than DSCLIM concerning the temporal and spatial correlations of pseudo-observed and analog precipitations, but the distributions reproduction is worst than the one observed with DSCLIM. Thus a further step of the analysis could be to look these distributions deeper in order to find ways to improve this fit.

6.3 Global assessment of the internship

The main new knowledge provision I acquired during my internship at CERFACS was the set of knowledge about climate in general and in particular climate projections in the climate change framework. But beyond this I learned a new programming language (NCL) and I have been trained to optimize my codes in R and NCL in order to make them faster and more efficient. Concerning statistics, I have applied some tools and tests I saw (or not) during the year but with which I never really worked before (Shapiro-Wilk test, Kolmogorov-Smirnov test, Principal Component Analysis, analog method...). Finally this internship has been educational in terms of time management as I had a lot to do and some quite heavy scripts to run (several hours of run).

References

[1] Catherine Jeandell and Rémy Mosseri (dir), *Le climat à découvert*, CNRS Editions, 2011

[2] Christian Pagé, Laurent Terray and Julien Boé, *DSCLIM : A software package to downscale climate scenarios at regional scale using a weather-typing based statistical methodology*, Climate Modeling and Global Change, CERFACS, September 2010

[3] P.A. Michelangeli, P. Vautard and B. Legras, *Weather Regimes: Recurrence and quasi stationarity, Journal of the Atmospheric Sciences*, 52 (8), 1237-1256, 1995

[4] P. Vautard, *Multiple Weather Regimes over the North Atlantic: Analysis of Precursors and Successors, Monthly Weather Review*, 118, 2056-2081, 1990

[5] Julien Boé, *Changement global et cycle hydrologique : une étude de régionalisation sur la France*, Thèse de l'Université Toulouse III Paul Sabatier, November 2007

[6] Julien Boé, Laurent Terray, Florence Habets and Eric Martin, *Projected changes in components of the hydrological cycle in French river bassins during the 21st century*, Water Resour.Res?, 45, W08426, 2009

[7] Gildas Dayon, *Evolution du cycle hydrologique continental en France au cours des prochaines décennies*, Thèse de l'Université Toulouse III Paul Sabatier, 2015

[8] Gildas Dayon, Julien Boé and Eric Martin, *Transferability in the future climate of a statistical downscaling method for precipitations in France*, Journal of Geophysical research: Atmospheres, 120, 1023-1043, 2015

[9] Gaelle Albertus, *Etude de la performance de DSCLIM sur l'Europe, un outil de désagrégation statistique pour les études d'impact du changement climatique*, Rapport technique CERFACS, August 2013

[10] Ed Hawkins, Jon Robson, Rowan Sutton, Doug Smith and Noel Keenyside, *Evaluating the potential for statistical decadal predictions of sea surface temperatures with a perfect model approach*, Springer, 2011

[11] Douglas Maraun et al., *VALUE: A framework to validate downscaling approaches for climate change studies*, *Earth's Future*, 3, 1-14, 2015

[12] J. Sillman, V.V. Kharin, X. Zhang, F.W. Zwiers and D. Bronaugh, *Climate extremes indices in the CMIP5 multimodel ensemble: Part 1. Model evaluation in the present climate*, *Journal of Geophysical Research : Atmospheres*, 118, 1716-1733, 2013

[13] Climate change in Australia Website, *How does a climate model work ?*:
http://www.climatechangeinaustralia.gov.au/en/climate-campus/modelling-and-projections/climate-models/theory-and-physics/

[14] S.P. Lloyd, *Least square quantization in PCM, Bell Telephone Laboratories Paper*, 1957, published in *IEEE Transactions on Information Theory*, vol. 28, 2, 129-137, 1982

[15] E.W. Forgy, *Cluster analysis of multivariate data: efficiency versus interpretability of classifications, Biometrics*, vol.21, 768-769, 1965

Documentation for programming languages:

[16] NCAR Command Language Website: http://www.ncl.ucar.edu/

[17] Pavel Michna, *Interface to NetCDF Datasets*, CRAN, February 2016
https://cran.r-project.org/web/packages/RNetCDF/RNetCDF.pdf

7 Appendix 1 - Scripts and output files

This appendix quickly presents the scripts I wrote to carry out my analyses, in order to facilitate their use for the next persons who may want to use them when working with DSCLIM. They are all located in /home/globc/vialaret/. One can find further details of their documentation directly in the codes.

NB: The scripts are quite generic except for the sizes of past and future periods that are explicitly given in the scripts. Moreover, for the scripts concerning the pure analog method, the data are not uniform (filenames, dimensions of variables, calendars...) so the scripts are less generic. Some of the files used to perform the perfect model analysis with the stochastic resampling method should be handled carefully too because of time-unit differences.

7.1 Scripts for the perfect model analysis

- **pmf_reganalysis_DSCLIM.R:** R script for the analysis of the regression phase of DSCLIM. This script contains the study of coefficient estimates, residual's behavior and statistics for the regression's quality (R^2, $\bar{R}^2$ and VIF) for all models considered in the perfect model framework (C4I, CNRM, DMI-BCM, DMI-ECHAM5, IACETH, KNMI, METNO-BCM, METNO-Had, METO-HadQ0, METO-HadQ16, MPI and OURANOS).
- **pmf_residmap_DSCLIM.ncl:** NCL script that contains the code for mapping the regressions' residuals and their standard deviations for all models.
- **pmf_resanalysis_DSCLIM.R:** R script for the analysis of the downscaling procedure's results. It contains the code for obtaining the downscaled precipitations using the uniform random selection of one analog day among eleven and to compute the various indicators I used in the dedicated section (temporal and spatial correlations, equality tests for means, variances and distributions) and to plot the figures I used to present my results for all models (boxplots, time series).
- **pmf_extranaylsis_wet_DSCLIM.R:** R script for the analysis of extreme precipitations. It contains the code for all comparison tools I used to study extreme rainfall: computation of the 99.9th quantile for pseudo-observations and analogs (for both past and future periods), absolute and relative changes in the quantiles between past and future, projection in the future period of the precipitation level corresponding to the 99.9th quantile of the past precipitation distributions (excluding precipitation levels that are below 1 mm/day). It also contains the code for plotting the different figures (boxplots) except the maps and for writing some variables I obtained in NetCDF files.
- **pmf_extrmaps_wet_DSCLIM.ncl:** NCL script that contains the code for mapping the 99.9th quantiles, their absolute and relative changes, and the projection in the future of the precipitation levels corresponding to the 99.9th quantile of the past precipitation distributions. The mapping of temporal correlations can be found here too.
- **pmf_extranaylsis_dry_1mm_DSCLIM.R:** R script for the extreme droughts' analysis when one considers that a dry day shows a precipitation level beyond 1mm/day. It contains the code for computing the numbers and proportions of dry days among all the days in the periods considered and for plotting some figures (boxplots).
- **pmf_extranaylsis_dry_quantile_DSCLIM.R:** R script for the extreme droughts' analysis. It contains the same tools as the previous script, but here days are said to be dry if they show a precipitation level beyond a given quantile of the precipitation distribution. It also computes and writes in a NetCDF file the number

of dry grid points for each day. Note that the quantile has to be changed manually at the beginning of the script before running it.

- **pmf_extranaylsis_dry_quantile_surface_DSCLIM.R:** R script for plotting the number of dry grid points per day. It draws a figure to compare the number of dry points per day for each quantile in the following list : 1%, 5%, 10%, 15% and 20%. Note that one has to check filenames before running the script (they are not generic).
- The following scripts are the same than for the analysis of the results of the downscaling procedure, but using the stochastic resampling method to select the final analog date (note that they are pretty longer to run):
  - **pmf_resanalysis_DSCLIM_SR.R**,
  - **pmf_extranaylsis_wet_DSCLIM_SR.R**,
  - **pmf_extrmaps_wet_DSCLIM_SR.ncl**,
  - **pmf_extranaylsis_dry_quantile_DSCLIM_SR.R**,
  - **pmf_extranaylsis_dry_quantile_surface_DSCLIM_SR.R**.

7.2 Scripts applying the diagnostics for the real data framework analysis

- **reganalysis_ERAI.R:** R script for the analysis of the regression phase of DSCLIM using ERAI reanalyses. It contains the tools for the study of coefficients, residuals' behavior and statistics for the regressions' quality.
- **residmap_ERAI.ncl:** NCL script for mapping the regressions' residuals and their standard deviations.
- **resanalysis_ERAI.R:** R script for the analysis of the downscaling procedure's results. It contains the code for deriving the analog precipitations (using the stochastic resampling method) and computing the tools I used above to compare observations and analogs (temporal and spatial correlations, equality tests for means, variances and distributions, computation of the wet and dry extremes).
- **eofs_variance_ERAI.ncl:** NCL script that compares the variance of PSL on the total PSL field and on the field projected on the 10 EOFs derived from ERAI reanalyses. It computes the ratio of explained variance and maps it for the following models : ACCESS1-3, bcc-csm1-1-m, BNU-ESM, CanESM2, CMCC-CM, CNRM-CM5, GFDL-CM3, IPSL-CM5A-LR, IPSL-CM5A-MR, MIROC5, MPI-ESM-MR, MRI-CGCM3, NorESM1-M and an ensemble mean.
- **resanalysis_CMIP5.R:** R script that computes the tests for equality of means ans variances and for distribution adequation between SAFRAN observations and CMIP5 model data (for models bcc-csm1-1-m, BNU-ESM, CNRM-CM5, CSIRO-Mk3-6-0, GFDL-CM3, IPSL-CM5A-MR, MIROC5, MPI-ESM-MR, MRI-CGCM3 and NorESM1-M).

7.3 Scripts for the pure analog method's analysis

NB: The data are not as uniform as the ones I used for DSCLIM so one should check whether the data have the right format since there are a lot of particular cases in the following scripts.

- **pmf_resanalysis_GD.R:** R script for the analysis of the pure analog downscaling procedure's results. It contains the code for obtaining downscaled precipitations with the pure analog method and to compute the various indicators I used in the dedicated section (temporal and spatial correlations, equality tests for means, variances and distributions) and to plot the figures I used to present my results for all models (boxplots).

- **pmf_extranalysis_wet_GD.R:** R script for the extreme precipitations' analysis. It contains the code for all comparison tools I used to study extreme rainfall: computation of the 99.9th quantile for pseudo-observations and analogs (for both past and future periods), absolute and relative changes in the quantiles between past and future, projection in the future period of the precipitation level corresponding to the 99.9th quantile of the past precipitation distributions (excluding precipitation levels that are below 1 mm/day). It also contains the code for plotting the different figures (boxplots) except the maps and for writing some variables I obtained in NetCDF files.

- **pmf_extrmaps_wet_GD.R:** NCL script that contains the code for mapping the 99.9th quantiles, their absolute and relative changes, and the projection in the future of the precipitation levels corresponding to the 99.9th quantile of the past precipitation distributions. The mapping of temporal correlations can also be found here. (**Nota:** This script has not been run).

- **pmf_extranalysis_dry_quantile_GD.R:** R script for the extreme droughts' analysis when one considers that a dry day shows a precipitation level beyond a given quantile of the precipitation distribution (0.01, 0.05, 0.10, 0.15 and 0.20). It contains the code for computing the dry quantiles in the periods of interest and for plotting some figures (boxplots). It also computes and writes in a NetCDF file the number of dry points per day. (**Nota:** This script has not been run).

- **pmf_extranalysis_dry_quantile_surface_GD.R:** R script for plotting the number of dry grid points per day. It draws a figure to compare the number of dry points per day for each quantile in the following list: 1%, 5%, 10%, 15% and 20%. (**Nota:** This script has not been run).

7.4 Other scripts

- **my_functions.R:** R script containing the functions used in all of the other R scripts: functions for the significance tests, equality tests, analog precipitations' computation, time conversions, etc.

7.5 Output files

They contain the NetCDF and text output files produced with the scripts described above and the figures.

NB: Check the names of output files that are given in the scripts before running them. For example the files produced on 12th July 2016 are automatically written in

```
/home/globc/vialaret/output/20160712_output/20160712_filename.extension
```

so one has to check whether the file "current-date_output" already exists in the "output" directory or change the file name directly in the script.

In the "output" directory files have been renamed in order to facilitate their use for potential future users. I use the following abbreviations: "pmf" for "perfect model framework", "SR" for "stochastic resampling", "tcorrel" and "scorrel" for "temporal/spatial correlation", "qextr" for "extreme quantile", "chabs" for "absolute change", "chrel" for "relative change", "GD" for "Dayon method", "qdry" for "dry quantile", "ts" for "time series".

8 Appendix 2 - Additional tables and figures

This appendix includes tables and figures I didn't include in the main text to avoid heaviness. The captions that include the mention **INTERMEDIATE REPORT** indicate tables and figures I took from the intermediate internship report of the 15th June. They are provided here as comparison tools for the new tables and figures presented in this final report.

	PAST				FUTURE			
	SON	**DJF**	**MAM**	**JJA**	**SON**	**DJF**	**MAM**	**JJA**
C4I	0.84	0.76	0.75	0.67	0.73	0.79	0.74	0.59
CNRM	0.49	0.93	0.62	0.71	0.36	0.85	0.68	0.28
DMI-BCM	0.74	0.90	0.82	0.60	0.83	0.47	0.80	0.28
DMI-ECHAM5	0.45	0.59	0.78	0.80	0.52	0.52	0.56	0.76
IACETH	0.88	0.72	0.78	0.58	0.86	0.62	0.69	0.65
KNMI	0.79	0.62	0.78	0.52	0.21	0.64	0.43	0.55
METNO-BCM	0.86	0.76	0.53	0.61	0.86	0.89	0.57	0.54
METNO-Had	0.71	0.94	0.75	0.53	0.50	0.79	0.72	0.83
METO-HadQ0	0.92	0.88	0.86	0.69	0.84	0.97	0.81	0.56
METO-HadQ16	0.91	0.65	0.92	0.70	0.81	0.74	0.87	0.57
MPI	0.75	0.39	0.99	0.88	0.74	0.25	0.92	0.80
OURANOS	0.80	0.71	0.65	0.87	0.61	0.62	0.61	0.83

Table 3: Tests for equality between the means of observed and analog precipitations: for both periods the average result of a Student t-test for mean equality is computed (the test returns 0 if the null hypothesis of mean equality is rejected, 1 if not)

	PAST				FUTURE			
	SON	**DJF**	**MAM**	**JJA**	**SON**	**DJF**	**MAM**	**JJA**
C4I	0.84	0.91	0.88	0.89	0.72	0.96	0.88	0.92
CNRM	0.93	0.92	0.78	0.89	0.87	0.89	0.75	0.88
DMI-BCM	0.88	0.83	0.88	0.82	0.93	0.84	0.88	0.76
DMI-ECHAM5	0.87	0.83	0.79	0.96	0.81	0.70	0.79	0.95
IACETH	0.94	0.87	0.87	0.93	0.90	0.82	0.86	0.88
KNMI	0.93	0.83	0.84	0.88	0.76	0.86	0.80	0.90
METNO-BCM	0.93	0.84	0.87	0.89	0.89	0.93	0.91	0.89
METNO-Had	0.83	0.80	0.86	0.88	0.84	0.66	0.91	0.88
METO-HadQ0	0.93	0.77	0.92	0.93	0.85	0.74	0.89	0.88
METO-HadQ16	0.91	0.90	0.98	0.94	0.86	0.97	0.97	0.92
MPI	0.80	0.84	0.96	0.95	0.76	0.76	0.95	0.93
OURANOS	0.84	0.86	0.96	0.93	0.80	0.91	0.94	0.90

Table 4: Tests for equality between the variances of observed and analog precipitations: for both periods the average result of a Fisher test for variance equality is computed (the test returns 0 if the null hypothesis of variance equality is rejected, 1 if not)

	PAST				FUTURE			
	SON	**DJF**	**MAM**	**JJA**	**SON**	**DJF**	**MAM**	**JJA**
C4I	0.99	1	0.99	0.98	0.99	1	0.98	0.95
CNRM	0.99	1	1	0.99	0.97	1	1	0.88
DMI-BCM	0.99	1	1	0.99	0.99	0.99	0.99	0.79
DMI-ECHAM5	0.99	0.98	0.98	1	0.96	0.90	0.88	1
IACETH	0.99	1	0.99	0.96	0.99	0.99	0.98	0.99
KNMI	0.99	0.99	1	0.99	0.95	0.98	0.77	0.99
METNO-BCM	0.99	0.99	0.99	0.96	0.99	1	0.98	0.85
METNO-Had	1	0.99	1	0.99	0.99	0.99	1	0.99
METO-HadQ0	0.99	1	1	0.99	1	1	1	0.99
METO-HadQ16	1	1	1	1	1	1	1	1
MPI	1	0.99	1	1	0.99	0.80	1	0.99
OURANOS	0.99	1	1	1	0.99	1	0.99	1

Table 5: Tests for adequation between the distributions of observed and analog precipitations: for both periods the average result of a Kolmogorov-Smirnov test for distribution adequation is computed (the test returns 0 if the null hypothesis of distribution matching is rejected, 1 if not)

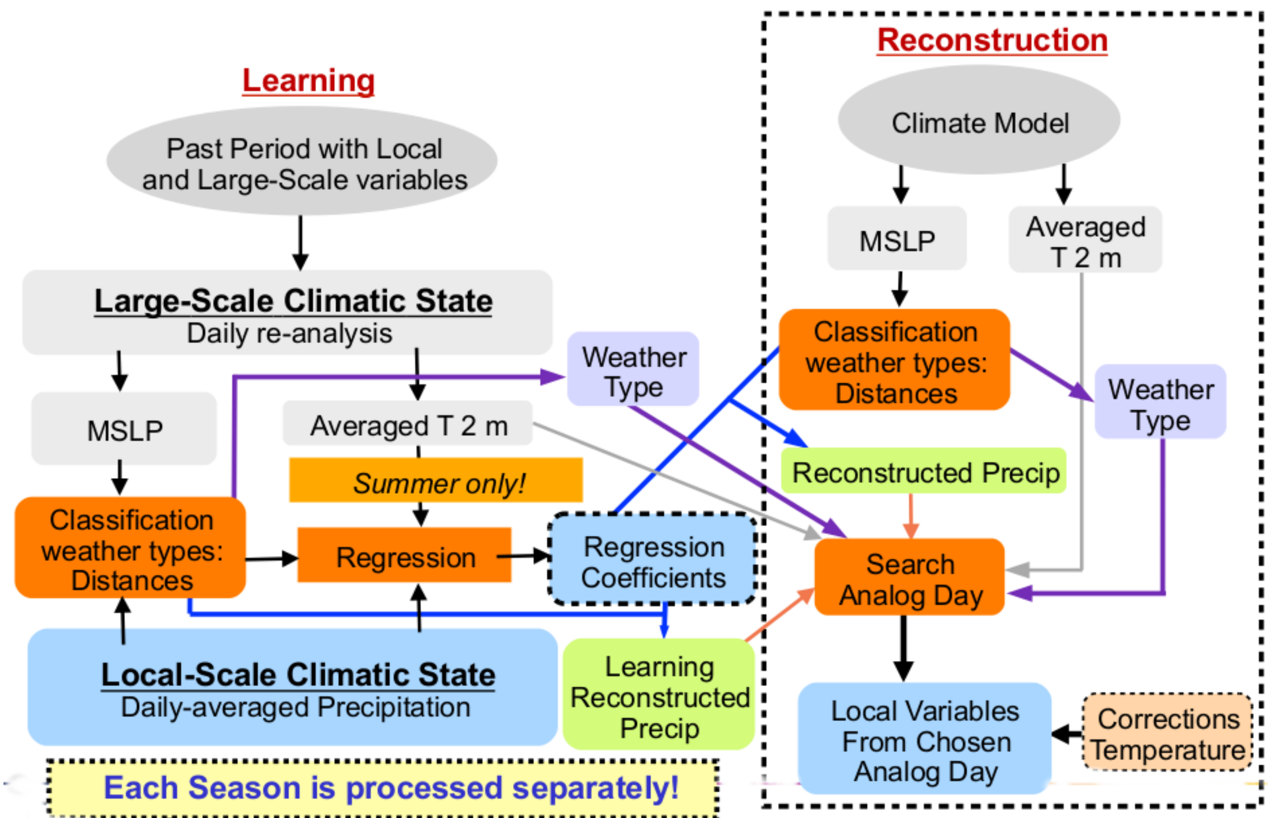


Figure 14: Summary of different steps of the DSCLIM method [2]

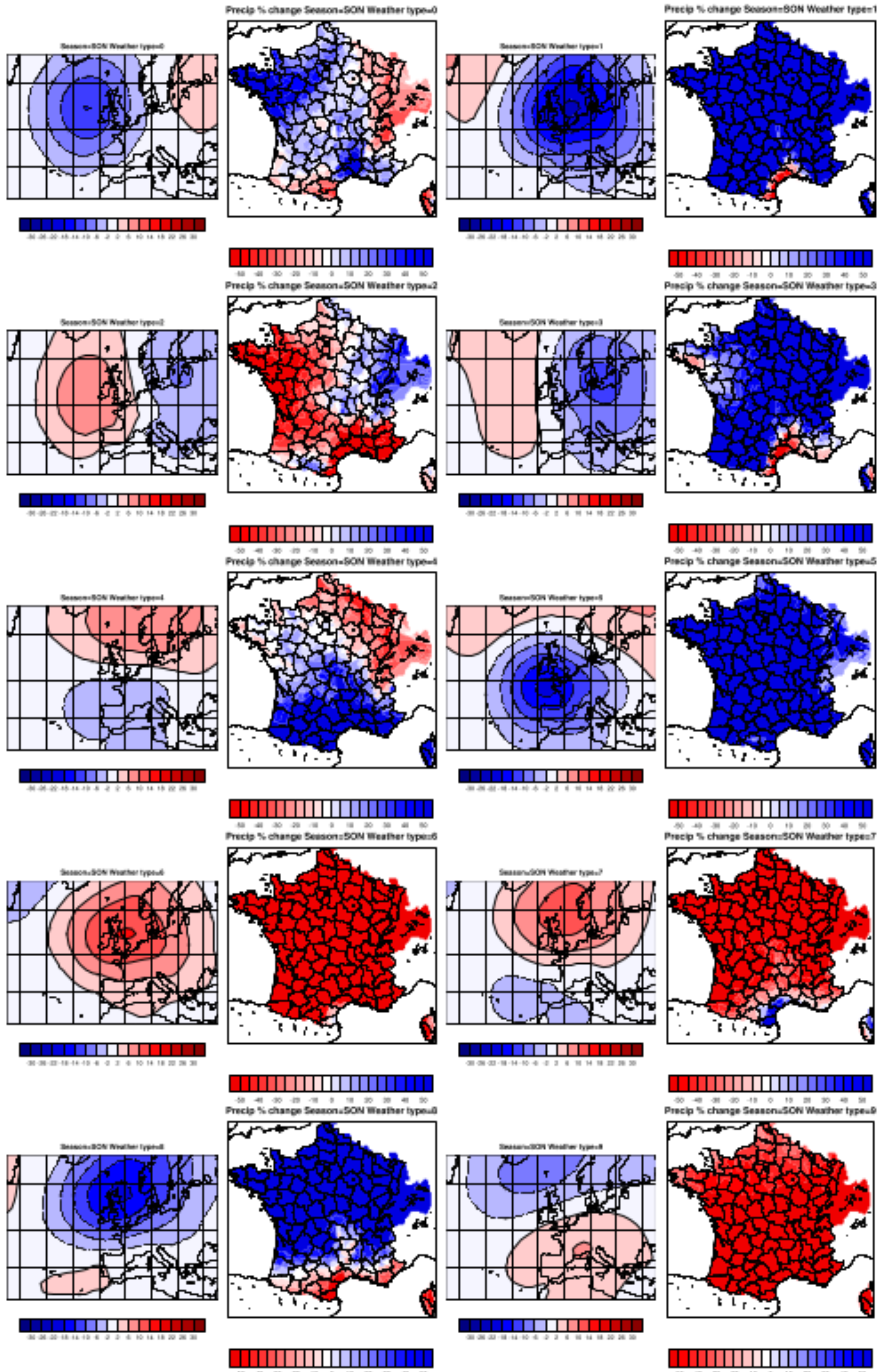


Figure 15: SON Weather types: APSL (hPa) and precipitations (%), for SAFRAN reanalyses. This figure is provided as an illustration of what a weather type represent, but the weather types I use in the perfect model framework are different for each of the twelve models and each season

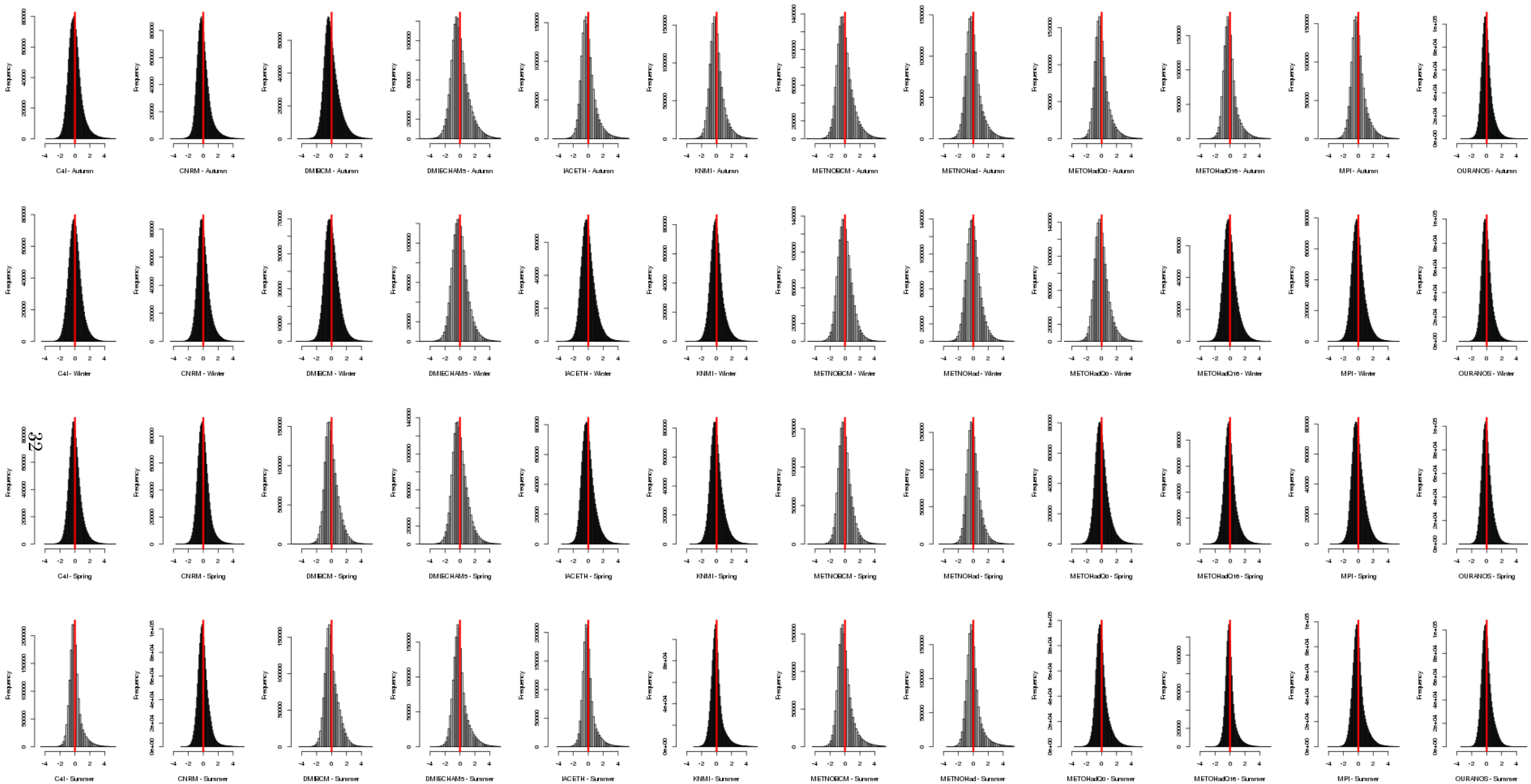


Figure 16: Histograms of residuals of the regressions, for all models and all seasons (line 1 = SON, line 2 = DJF, line 3 = MAM and line 4 = JJA ; each column represent a given model : C4I, CNRM, DMI-BCM, DMI-ECHAM5, IACETH, KNMI, METNO-BCM, METNO-Had, METO-HadQ0, METO-HadQ16, MPI and OURANOS)

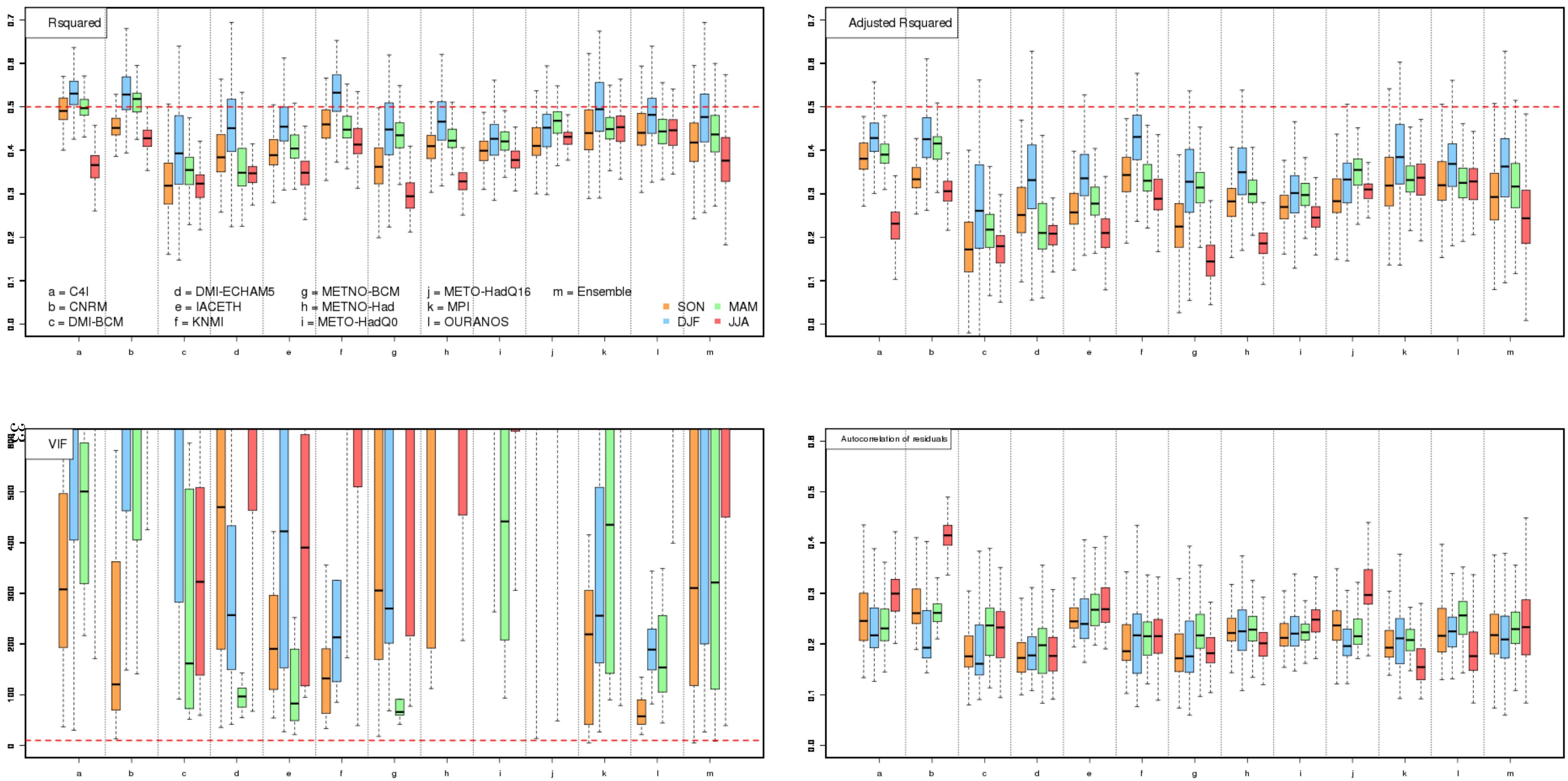


Figure 17: Boxplots summarizong the indices for the quality of each regression : R^2, $\bar{R}^2$, VIF and residual's autocorrelation (by model and by season). Nota: the outliers are not represented in these boxplots.

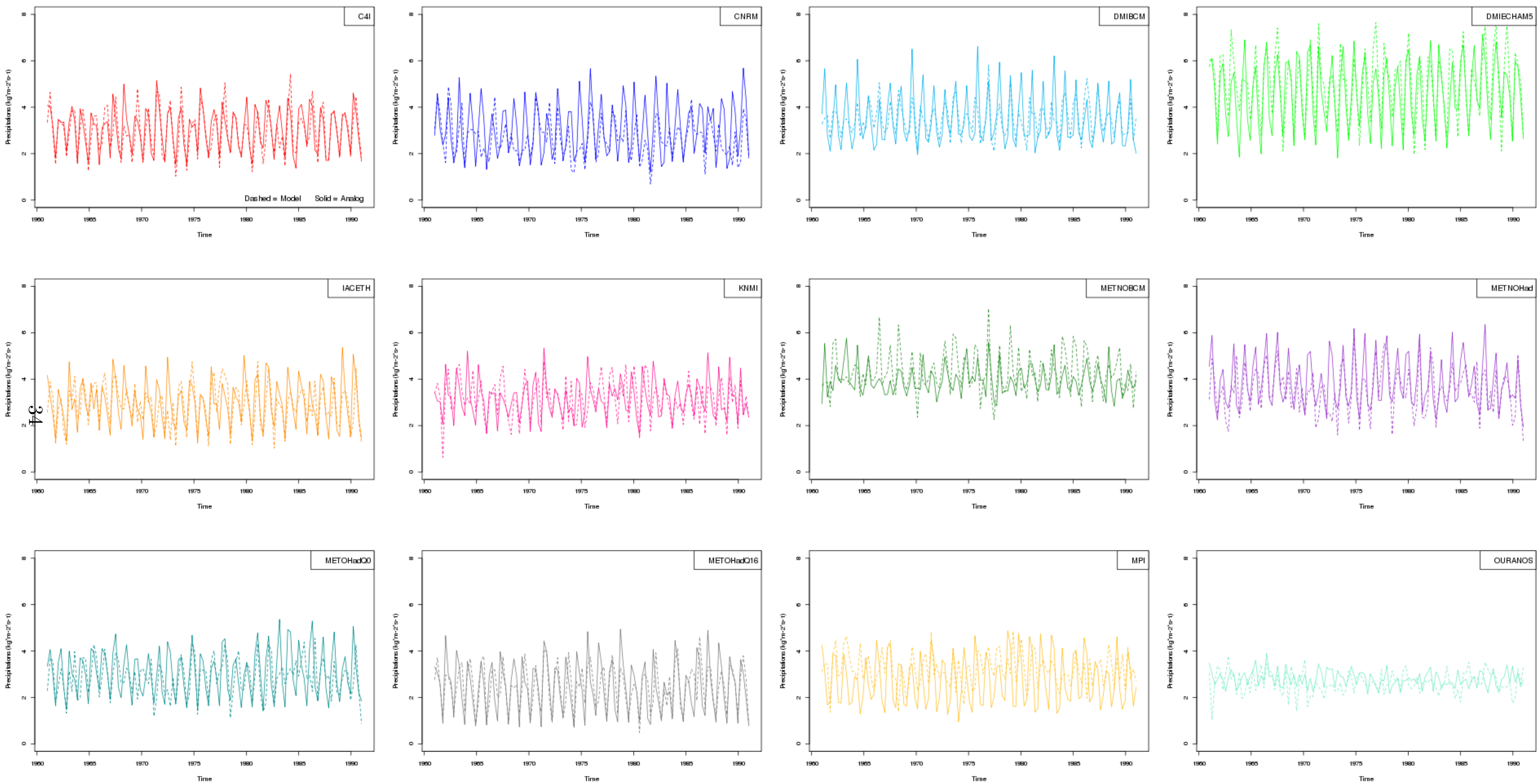


Figure 18: Time series of inter-annual observed (dashed) and analog (solid) precipitations for the different models (past period)

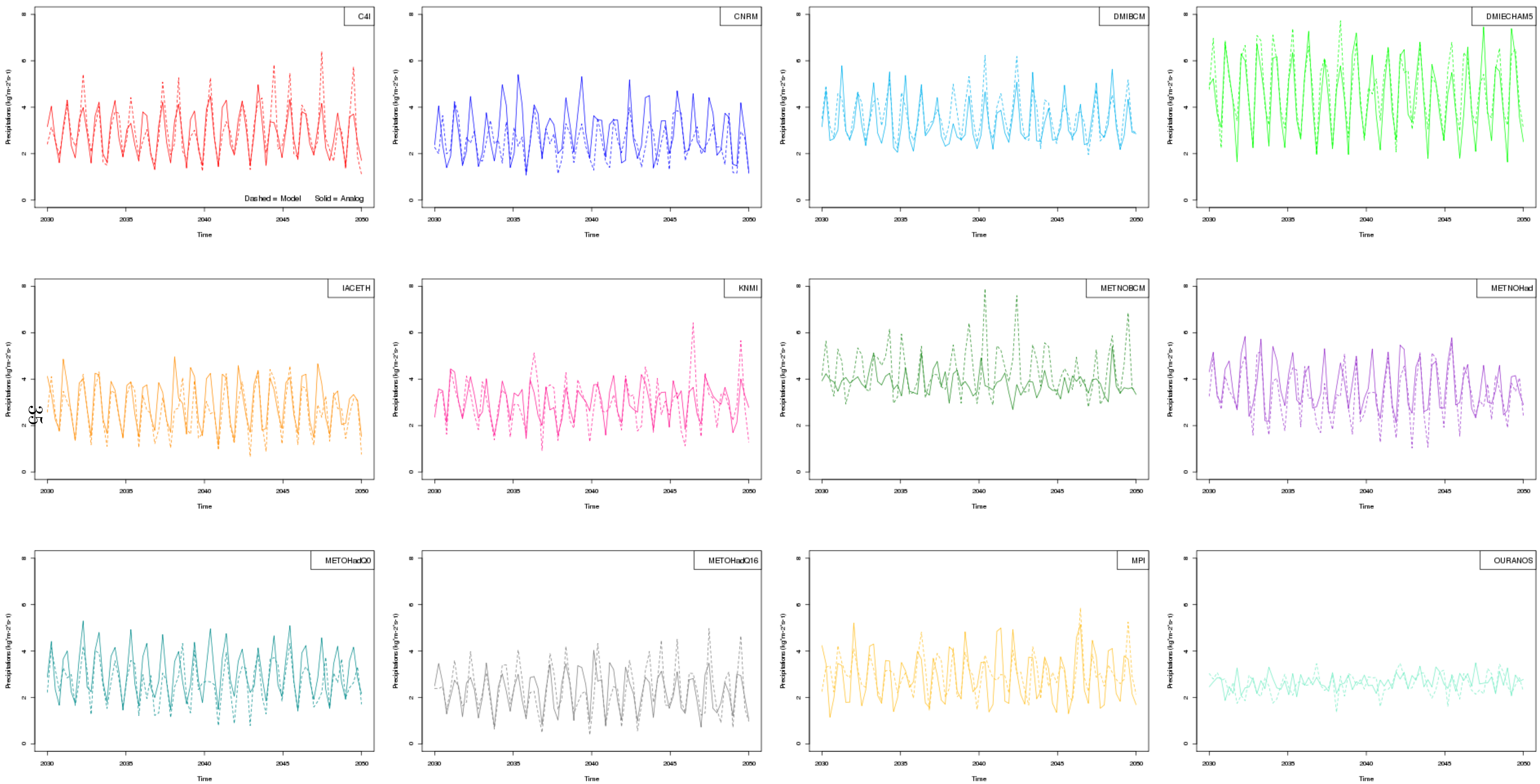


Figure 19: Time series of inter-annual observed (dashed) and analog (solid) precipitations for the different models (future period)

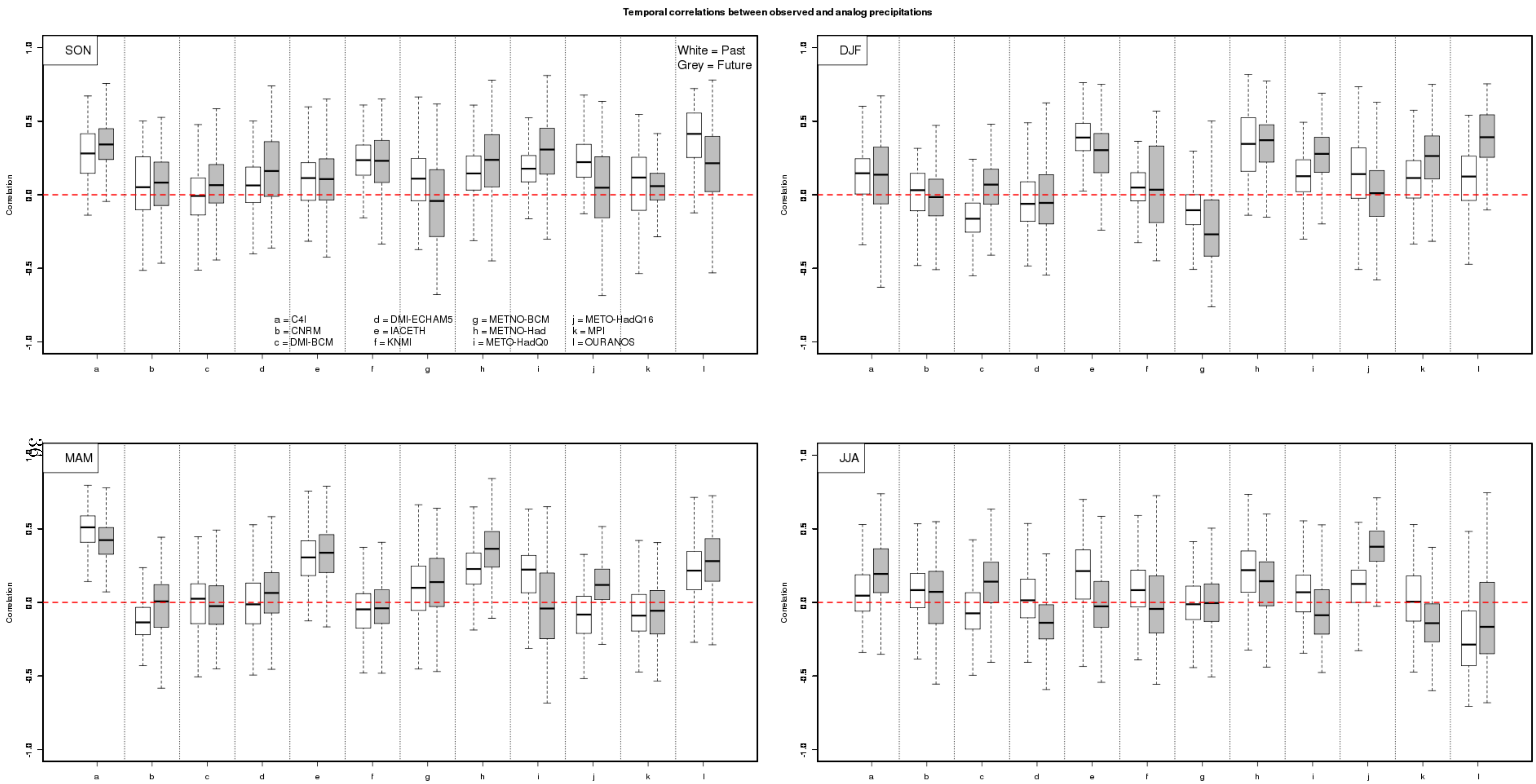


Figure 20: Boxplots for temporal correlations at each grid point between observed and analog precipitations by model and by season, for both past (white) and future (grey) periods, obtained with the stochastic resampling method. Nota: the outliers are not represented in these boxplots.

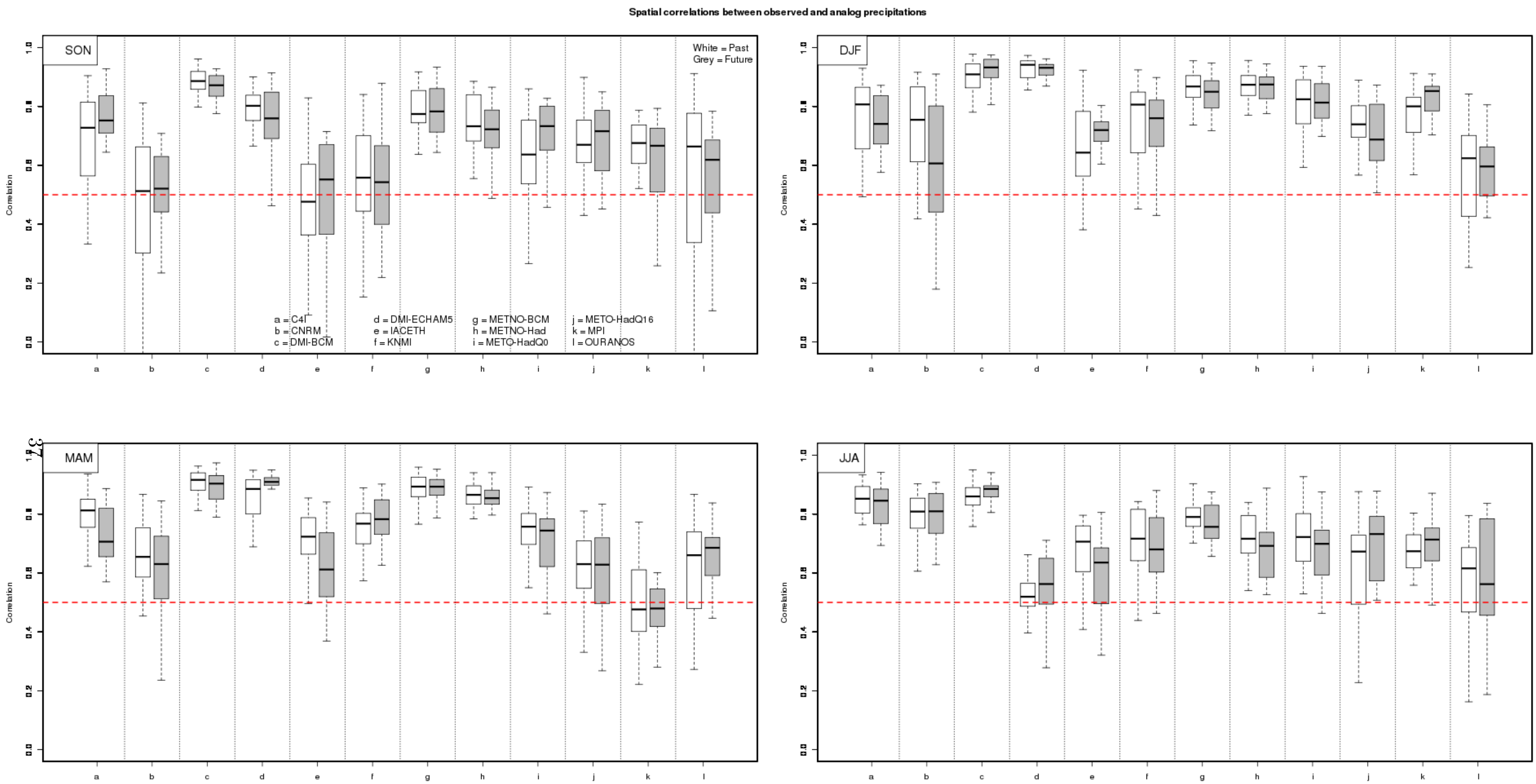


Figure 21: Boxplots for spatial correlations over the whole of France between observed and analog precipitations by model and by season, for both past (white) and future(grey) periods, obtained with the stochastic resampling method. Nota: the outliers are not represented in these boxplots.

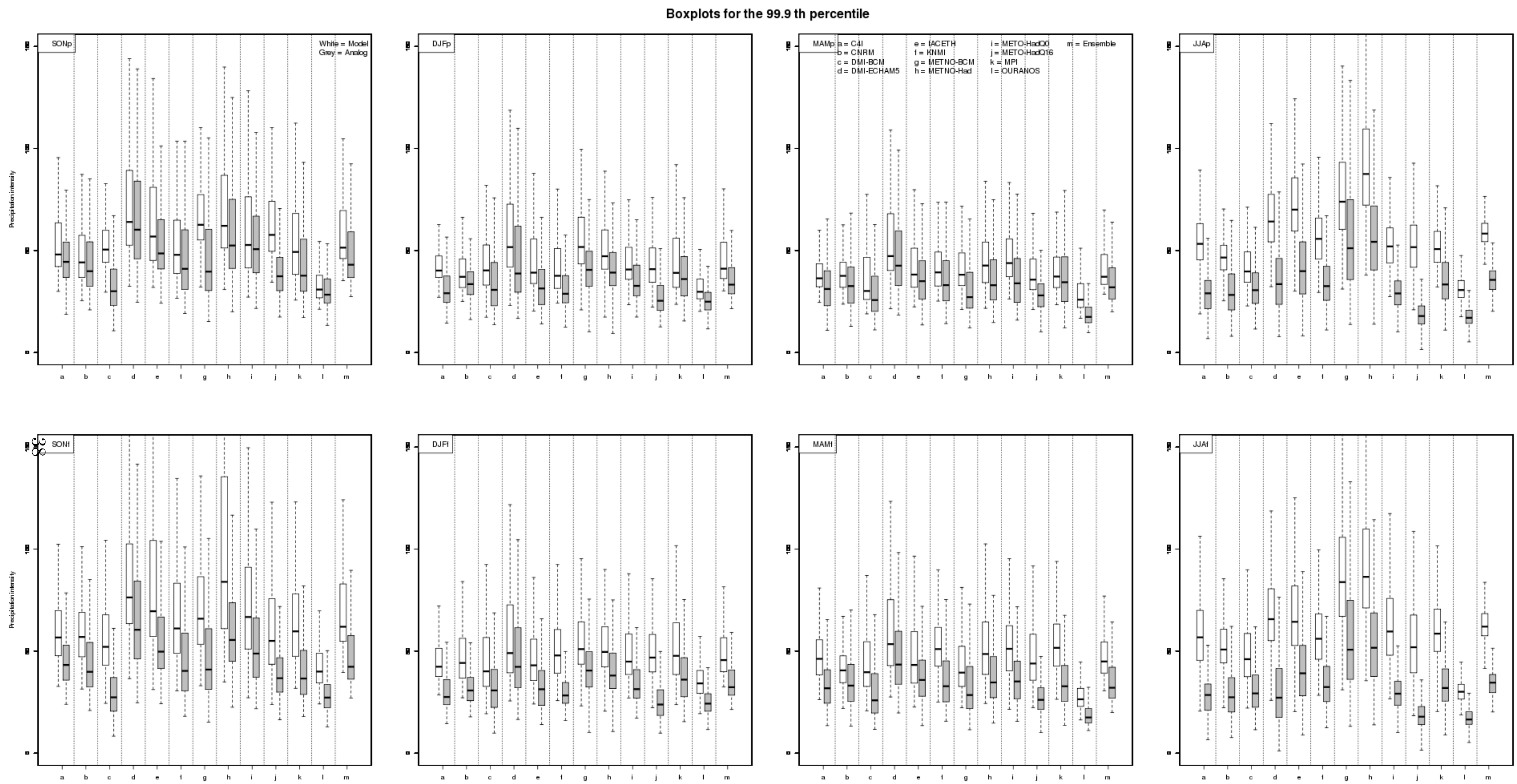


Figure 22: Boxplots for 99.9th quantiles computed for each grid oint, each model, each season and both past (line 1) and future (line 2) periods, for observed (white) and analog (grey) precipitations. Nota: the outliers are not represented in these boxplots.

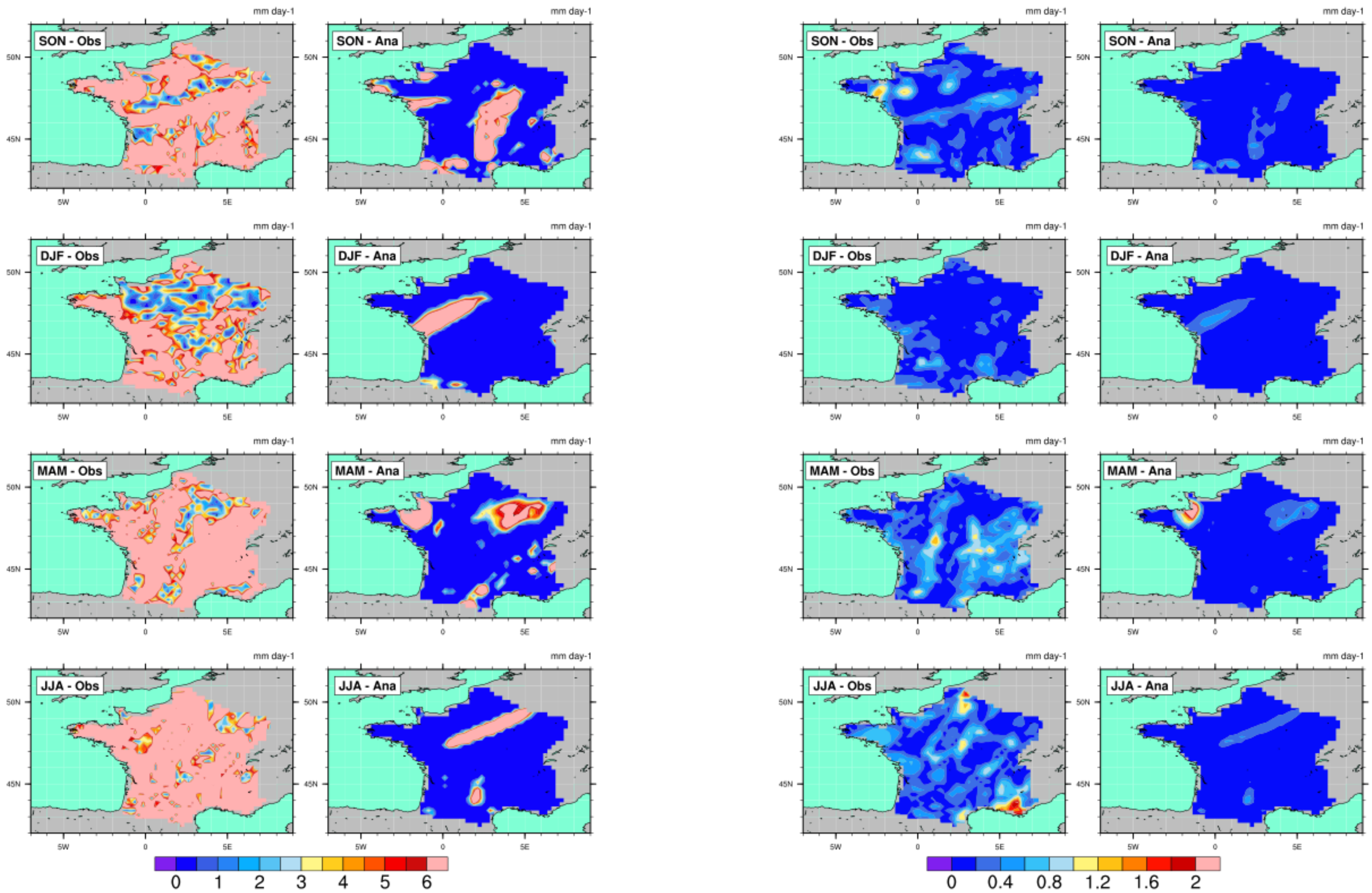


Figure 23: Absolute (left) and relative (right) changes between the 99.9th percentile of the past period's precipitation distributions and the 99.9th percentile of the future period's precipitation distributions (in absolute value) for each grid point and each season ("Obs" and "Ana" stand for "Observed" and "Analog") for model C4I

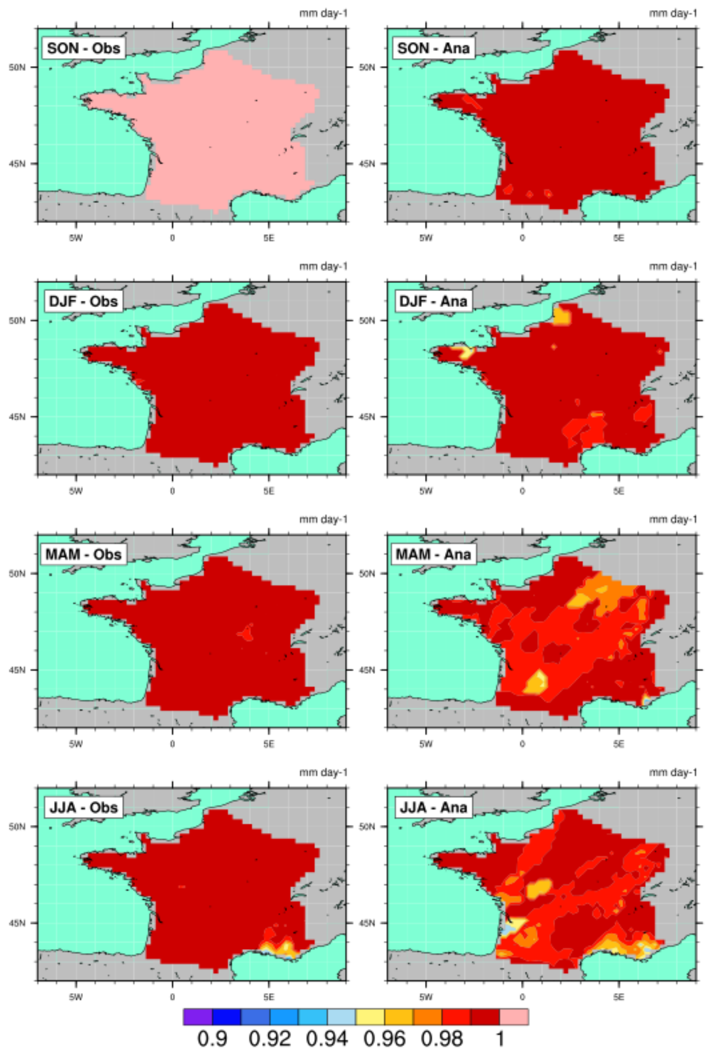


Figure 24: Projection of the 99.9th quantiles of the past period on future precipitations for each grid point and each season ("Obs" stands for "Observed" and "Ana" for "Analog") for model C4I

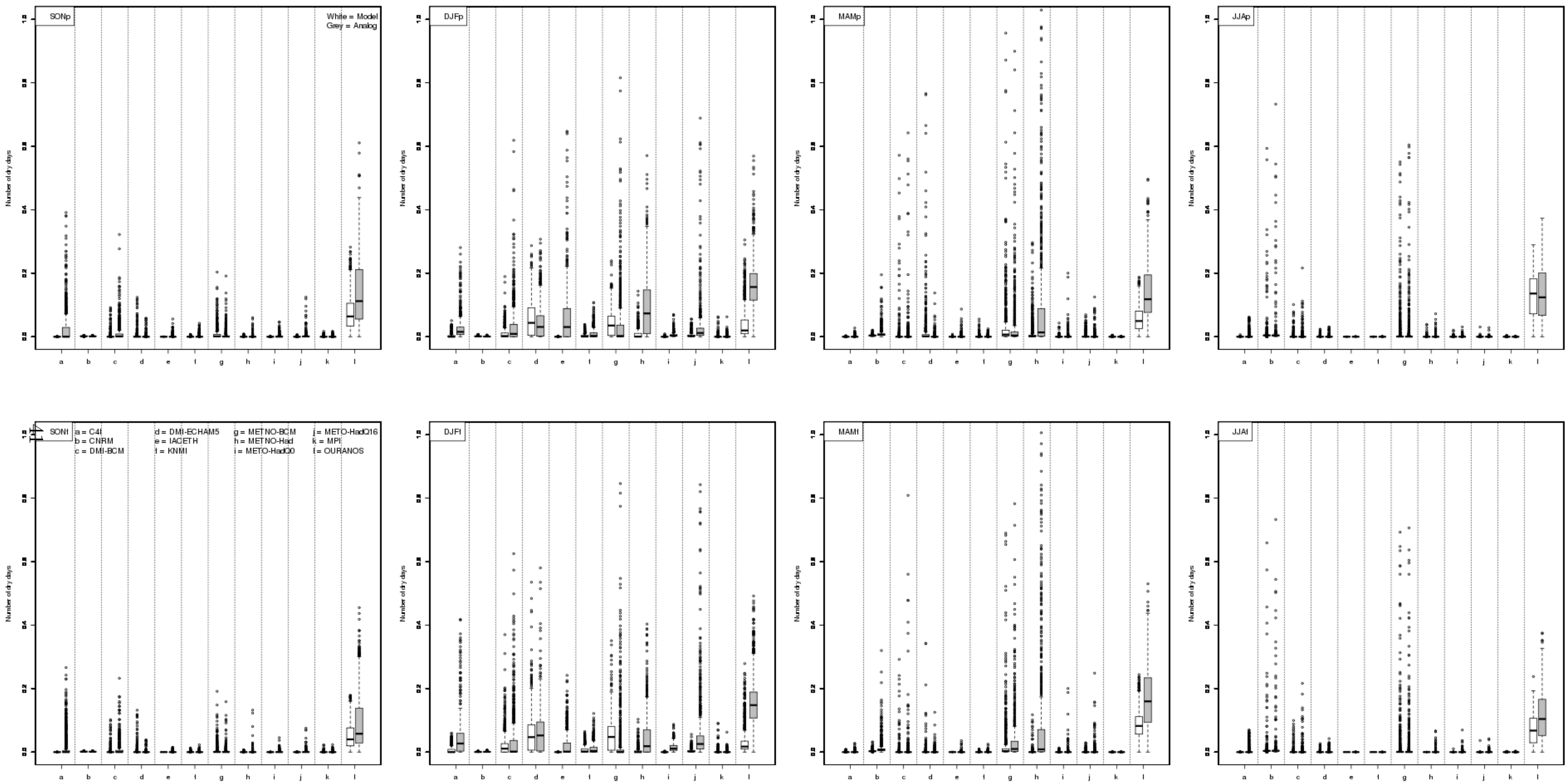


Figure 25: Boxplots of the 10% percentile of the precipitation distributions for each model, each seasons and both past and future periods. Observed quantiles are in white, analogs are in grey. The dots represent the outliers.

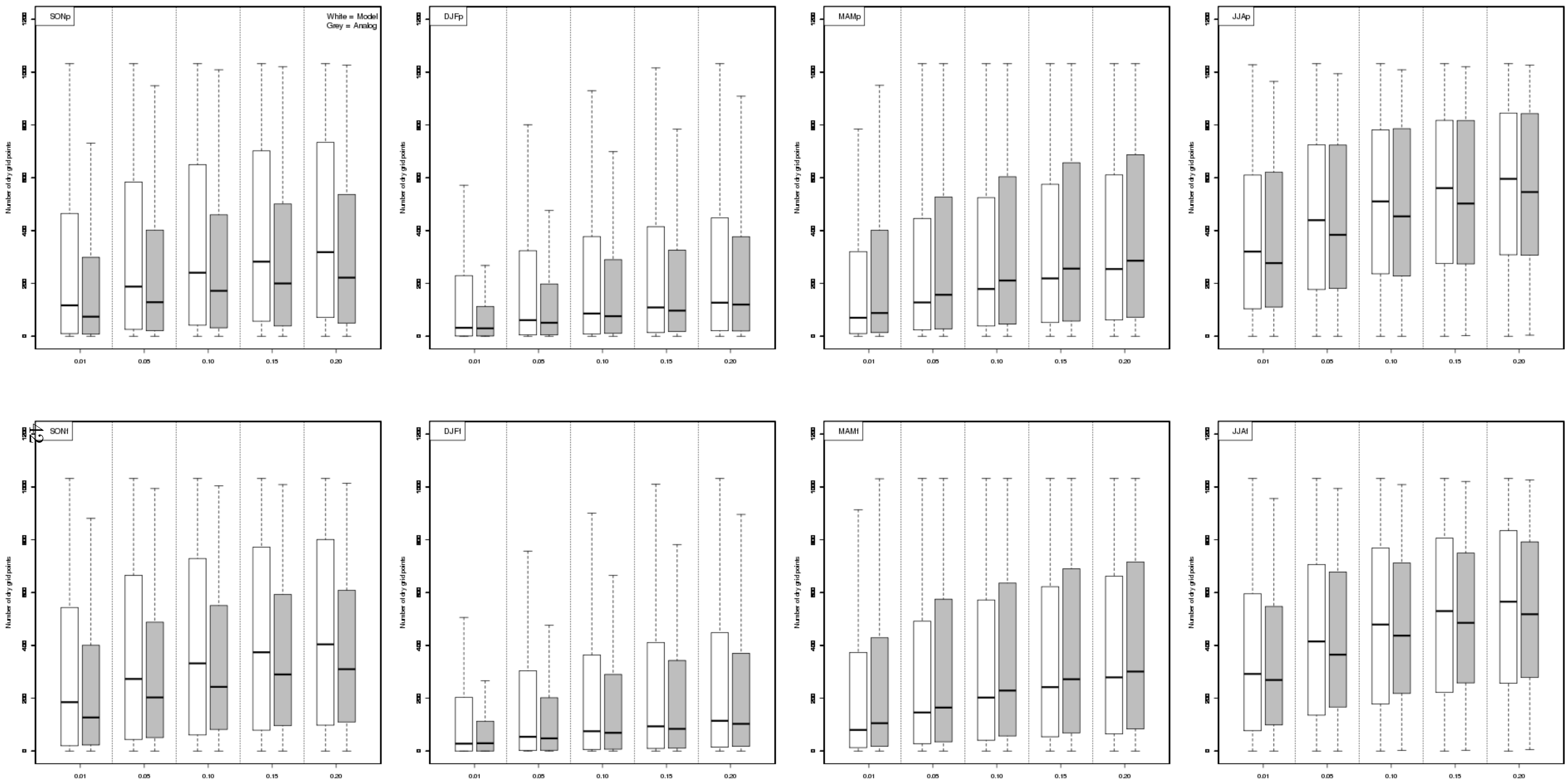


Figure 26: Boxplots of the number of dry points per day for model C4I. The dry points are counted for each day and each quantiles among 0.01, 0.05, 0.1, 0.15 and 0.2, for both observed and analog precipitations and both past and future periods.

	PAST				FUTURE			
	SON	**DJF**	**MAM**	**JJA**	**SON**	**DJF**	**MAM**	**JJA**
C4I	0.61	0.84	0.75	0.59	0.47	0.81	0.65	0.64
CNRM	0.28	0.64	0.75	0.63	0.12	0.81	0.71	0.36
DMI-BCM	0.80	0.66	0.76	0.55	0.27	0.35	0.17	0.96
DMI-ECHAM5	0.49	0.68	0.62	0.95	0.52	0.52	0.56	0.76
IACETH	0.50	0.83	0.68	0.74	0.58	0.74	0.63	0.51
KNMI	0.54	0.72	0.88	0.43	0.47	0.62	0.55	0.54
METNO-BCM	0.45	0.78	0.68	0.51	0.62	0.81	0.49	0.38
METNO-Had	0.49	0.81	0.77	0.69	0.44	0.54	0.60	0.49
METO-HadQ0	0.86	0.59	0.66	0.47	0.68	0.74	0.70	0.21
METO-HadQ16	0.78	0.82	0.72	0.89	0.56	0.74	0.73	0.57
MPI	0.54	0.52	0.91	0.68	0.39	0.78	0.91	0.41
OURANOS	0.83	0.83	0.71	0.90	0.79	0.60	0.62	0.85

Table 6: **INTERMEDIATE REPORT** - Tests for equality between the means of observed and analog precipitations: for both periods the average result of a Student t-test for mean equality is computed (the test returns 0 if the null hypothesis of mean equality is rejected, 1 if not)

	PAST				FUTURE			
	SON	**DJF**	**MAM**	**JJA**	**SON**	**DJF**	**MAM**	**JJA**
C4I	0.83	0.86	0.87	0.89	0.91	0.84	0.84	0.98
CNRM	0.78	0.90	0.89	0.91	0.77	0.91	0.88	0.89
DMI-BCM	0.96	0.90	0.82	0.83	0.94	0.96	0.83	0.87
DMI-ECHAM5	0.89	0.84	0.75	0.98	0.79	0.74	0.99	0.98
IACETH	0.85	0.93	0.86	0.95	0.89	0.91	0.82	0.89
KNMI	0.83	0.83	0.85	0.85	0.77	0.79	0.82	0.84
METNO-BCM	0.80	0.87	0.85	0.89	0.82	0.91	0.83	0.86
METNO-Had	0.84	0.77	0.88	0.94	0.88	0.79	0.80	0.93
METO-HadQ0	0.92	0.89	0.87	0.91	0.88	0.86	0.87	0.90
METO-HadQ16	0.93	0.93	0.94	0.99	0.93	0.93	0.92	0.96
MPI	0.84	0.78	0.92	0.93	0.85	0.88	0.95	0.86
OURANOS	0.92	0.90	0.90	0.94	0.93	0.88	0.90	0.93

Table 7: **INTERMEDIATE REPORT** - Tests for equality between the variances of observed and analog precipitations: for both periods the average result of a Fisher test for variance equality is computed (the test returns 0 if the null hypothesis of variance equality is rejected, 1 if not)

	PAST				FUTURE			
	SON	**DJF**	**MAM**	**JJA**	**SON**	**DJF**	**MAM**	**JJA**
C4I	1	0.99	0.99	0.97	0.98	1	0.97	0.89
CNRM	0.99	1	1	0.99	0.88	0.89	0.88	1
DMI-BCM	0.99	1	1	0.97	0.64	0.99	0.99	0.99
DMI-ECHAM5	0.99	0.99	0.99	1	0.98	0.82	0.72	1
IACETH	0.99	1	0.99	0.95	0.93	1	0.99	0.99
KNMI	0.98	0.99	1	0.91	0.89	1	0.89	0.92
METNO-BCM	0.99	0.99	0.99	0.95	0.87	1	0.97	0.73
METNO-Had	0.99	1	1	0.99	0.99	0.99	0.99	0.99
METO-HadQ0	0.99	1	1	0.99	1	1	1	0.97
METO-HadQ16	0.99	1	1	0.99	1	1	1	0.99
MPI	0.99	0.99	1	1	0.94	0.99	1	0.99
OURANOS 1	1	1	1	0.98	1	0.99	1	

Table 8: **INTERMEDIATE REPORT** - Tests for adequation between the distributions of observed and analog precipitations: for both periods the average result of a Kolmogorov-Smirnov test for distribution adequation is computed (the test returns 0 if the null hypothesis of distribution matching is rejected, 1 if not)

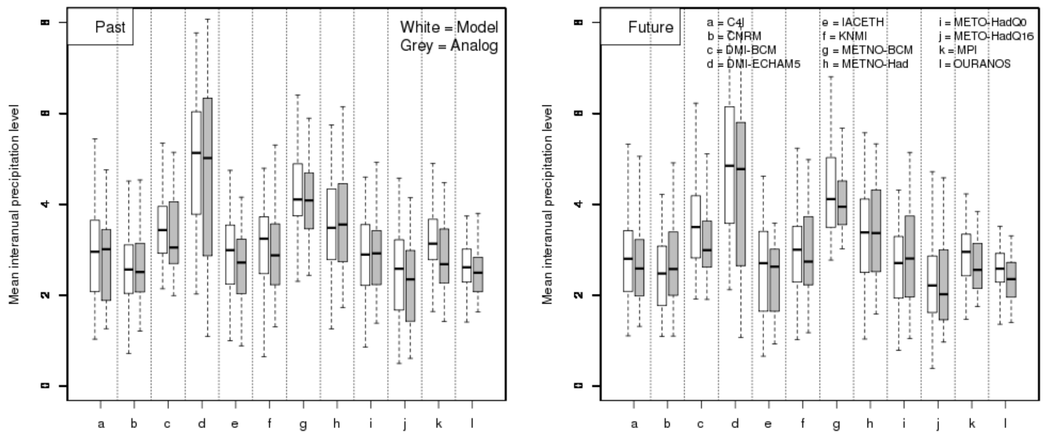


Figure 27: **INTERMEDIATE REPORT** - For each model and for both past and future periods, boxplots of the values of the observed and analog inter-annual precipitations. Nota: the outliers are not represented in these boxplots.

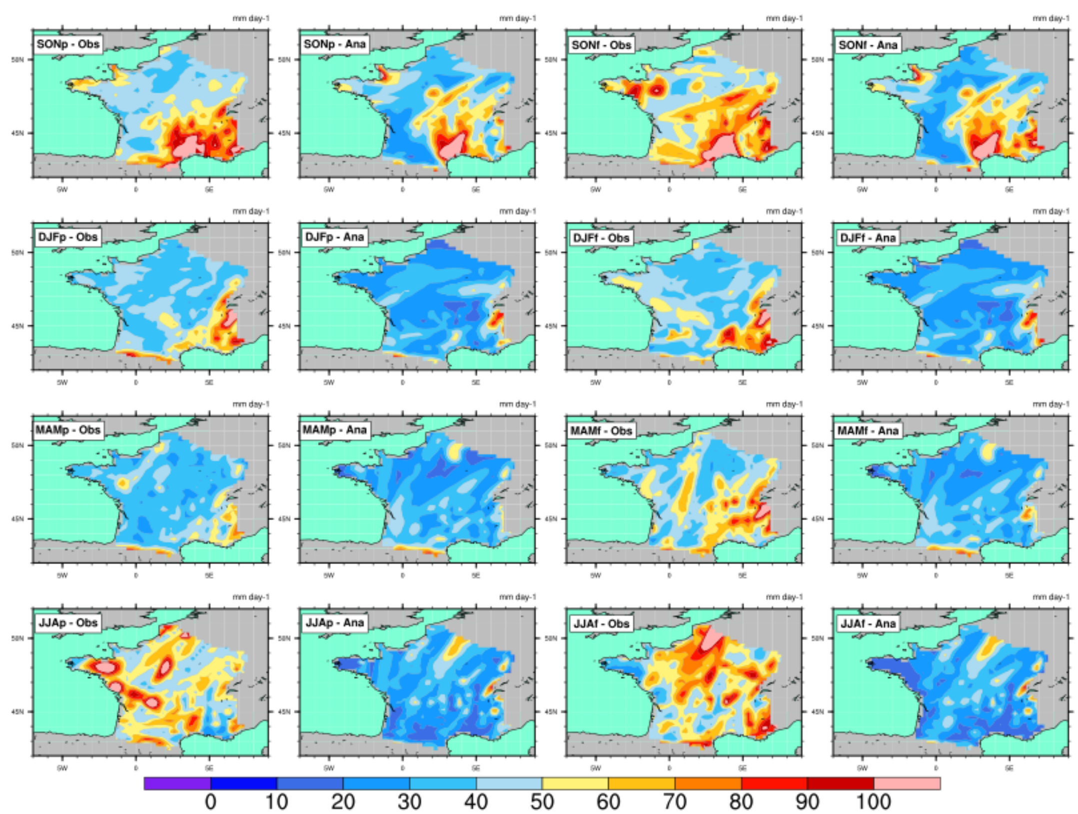


Figure 28: **INTERMEDIATE REPORT** - The 99.9th quantiles of the precipitation distribution for each grid point, each season and each period ("p" and "f" after the season's acronym stand for "past" and "future", "Obs" and "Ana" stand for "Observed" and "Analog") for model C4I

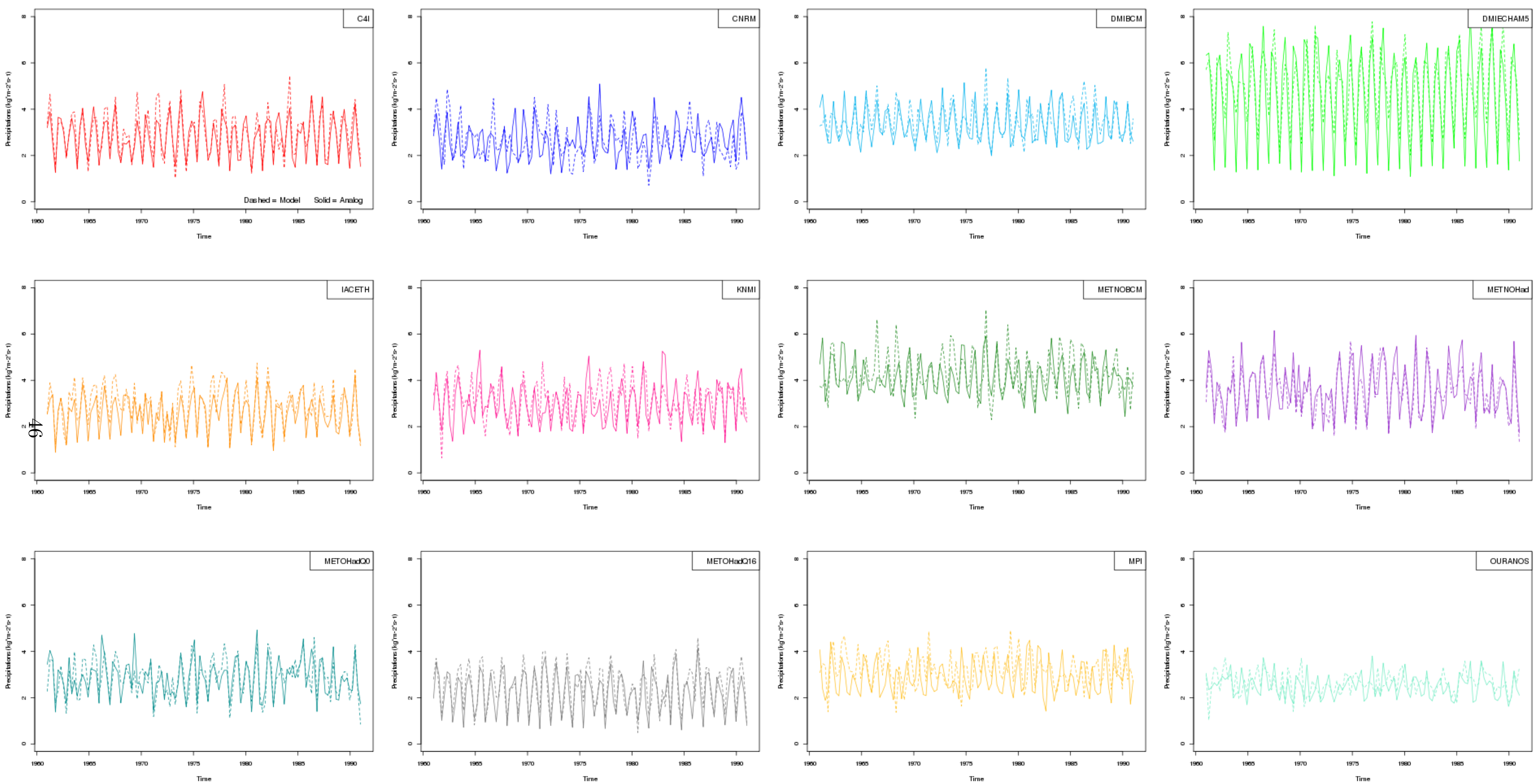


Figure 29: **INTERMEDIATE REPORT** - Time series of inter-annual observed (dashed) and analog (solid) precipitations for the different models (past period)

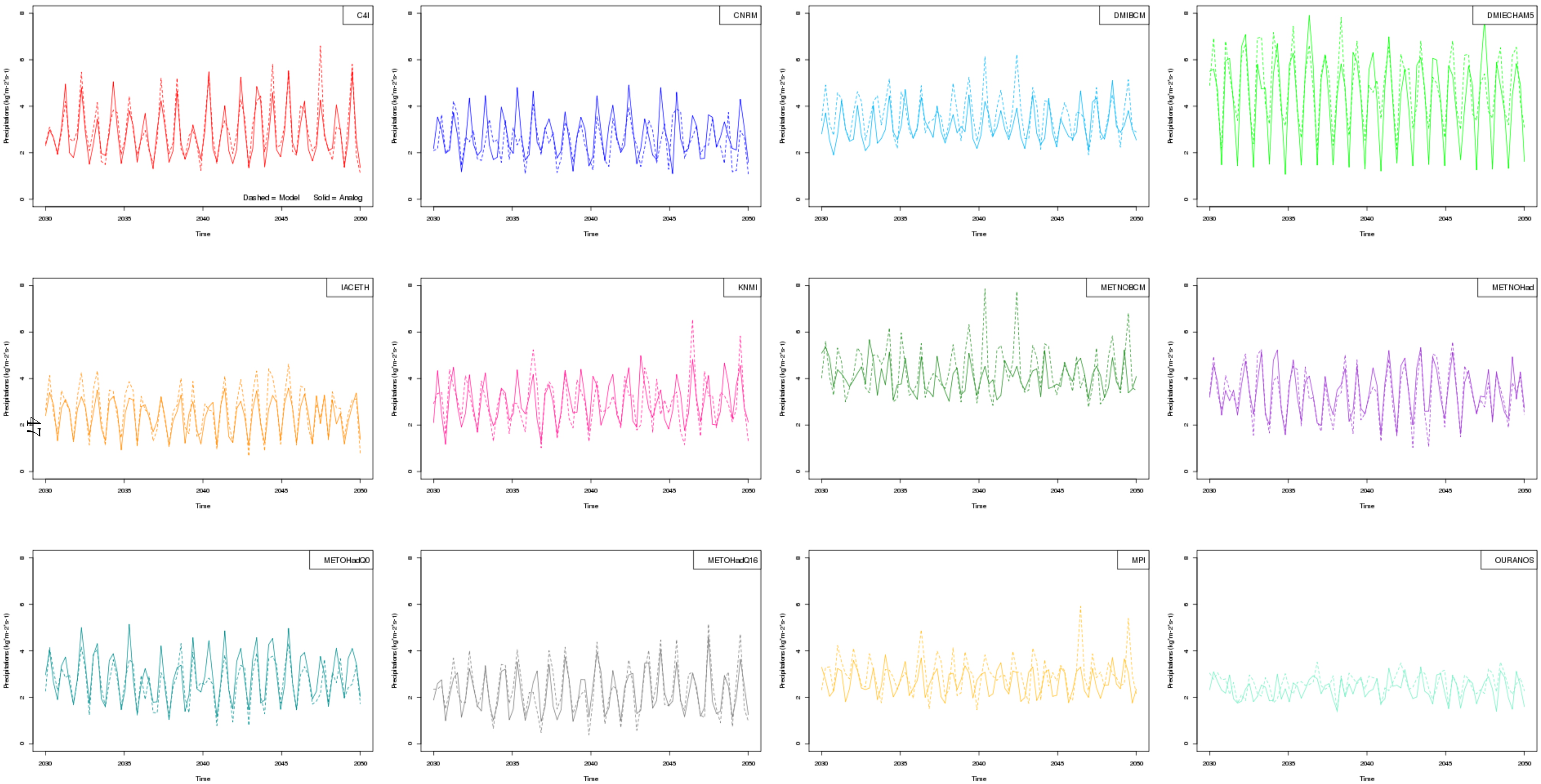


Figure 30: **INTERMEDIATE REPORT** - Time series of inter-annual observed (dashed) and analog (solid) precipitations for the different models (future period)

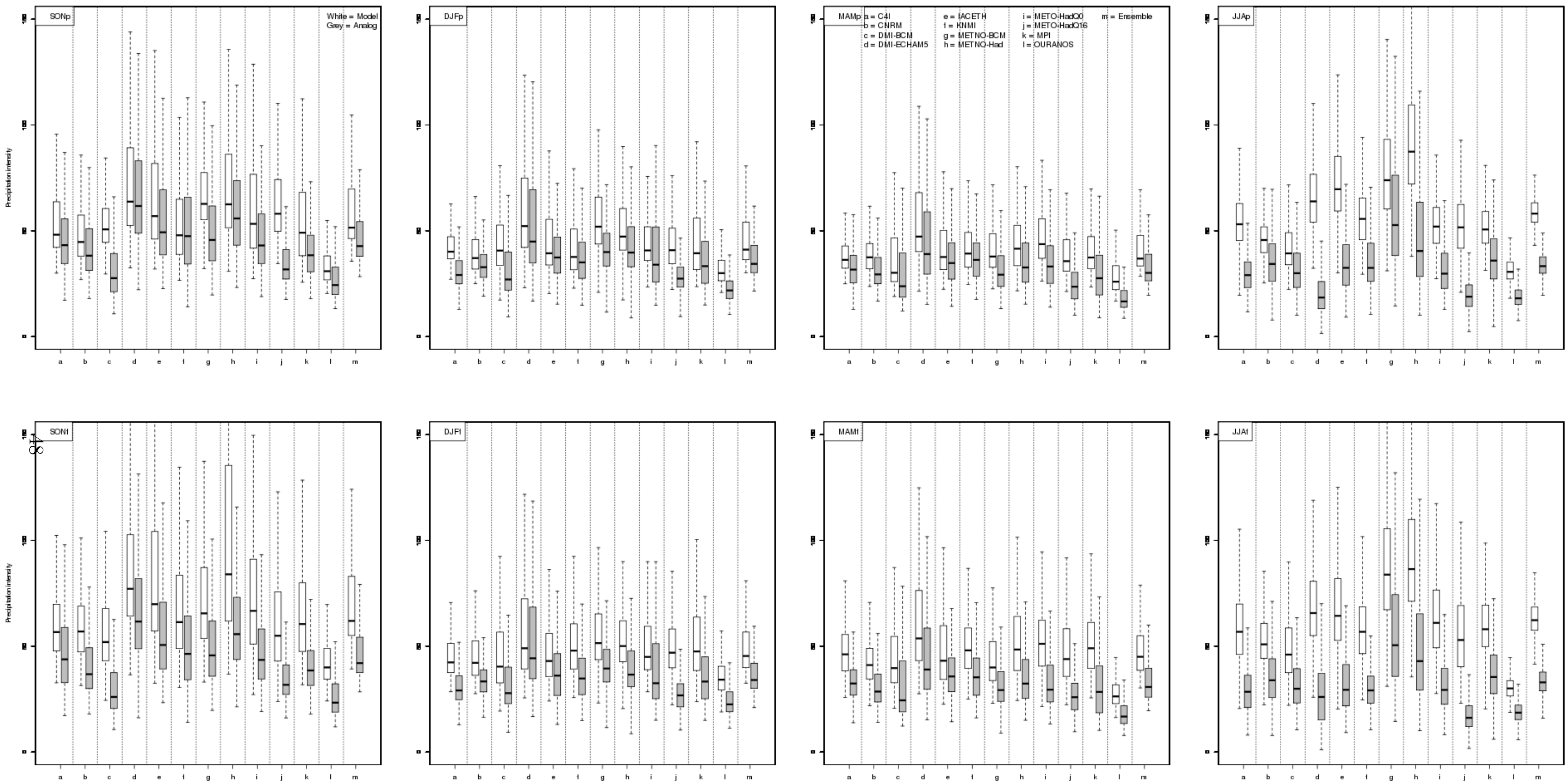


Figure 31: **INTERMEDIATE REPORT** - Boxplots for 99.9th quantiles computed for each grid point, each model, each season and both past (line 1) and future (line 2) periods, for observed (white) and analog (grey) precipitations. Nota: the outliers are not represented in these boxplots.

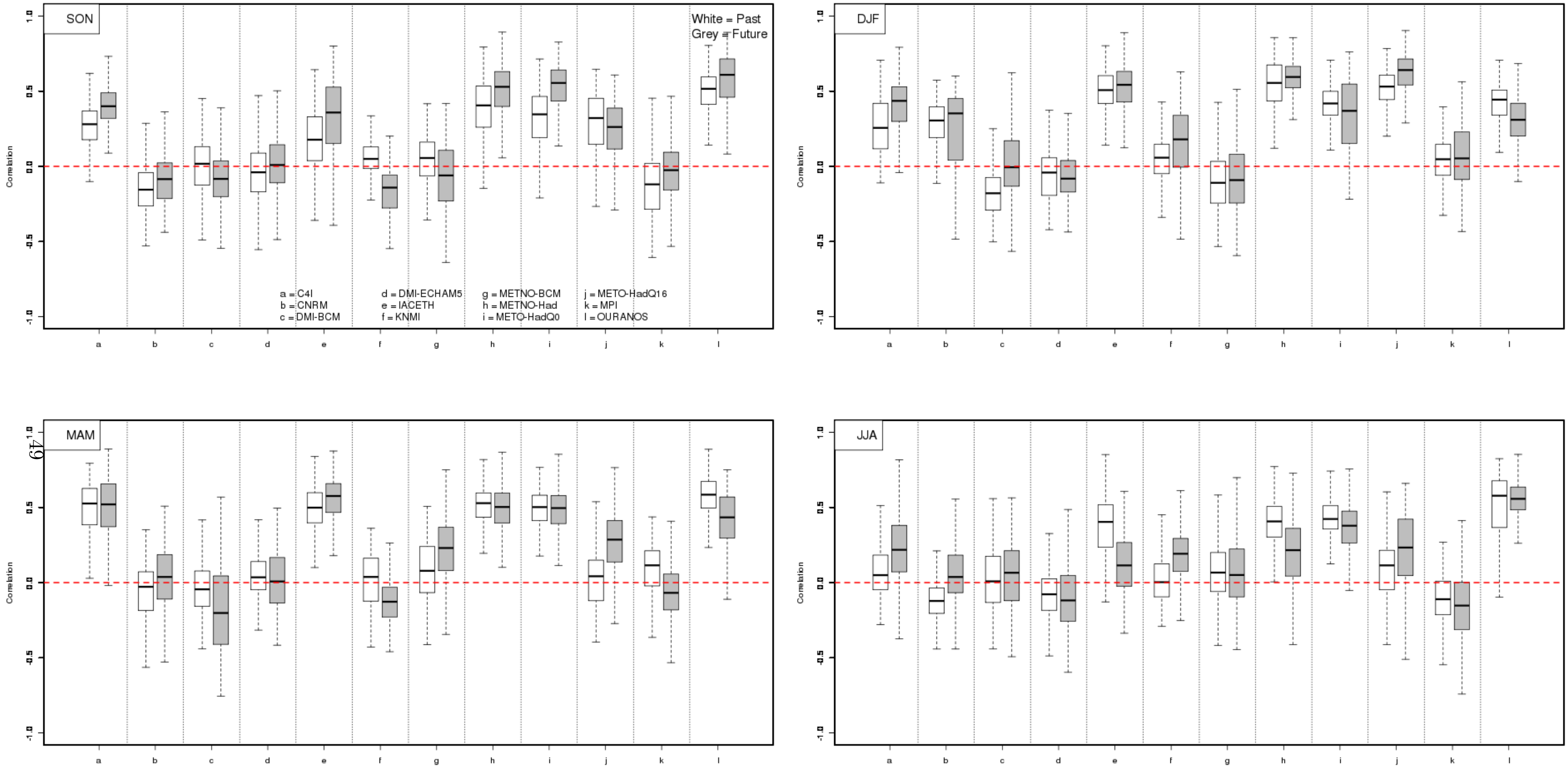


Figure 32: **INTERMEDIATE REPORT** - Boxplots for temporal correlations at each grid point between observed and analog precipitations by model and by season, for both past (white) and future (grey) periods, obtained with the uniform selection of the analog day. Nota: the outliers are not represented in these boxplots.

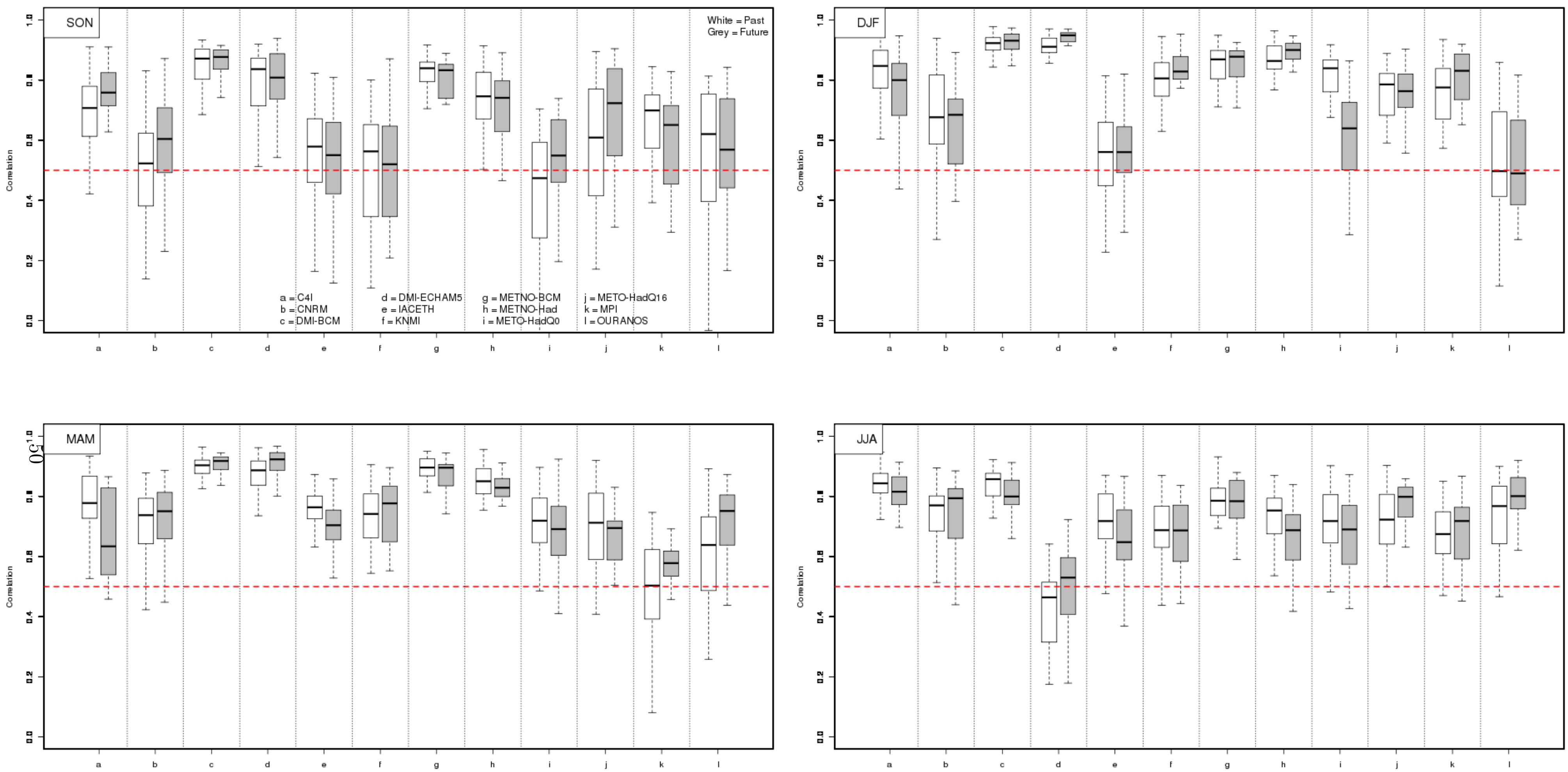


Figure 33: **INTERMEDIATE REPORT** - Boxplots for spatial correlations over the whole of France between observed and analog precipitations by model and by season, for both past (white) and future(grey) periods, obtained with the uniform selection of the analog day. Nota: the outliers are not represented in these boxplots.

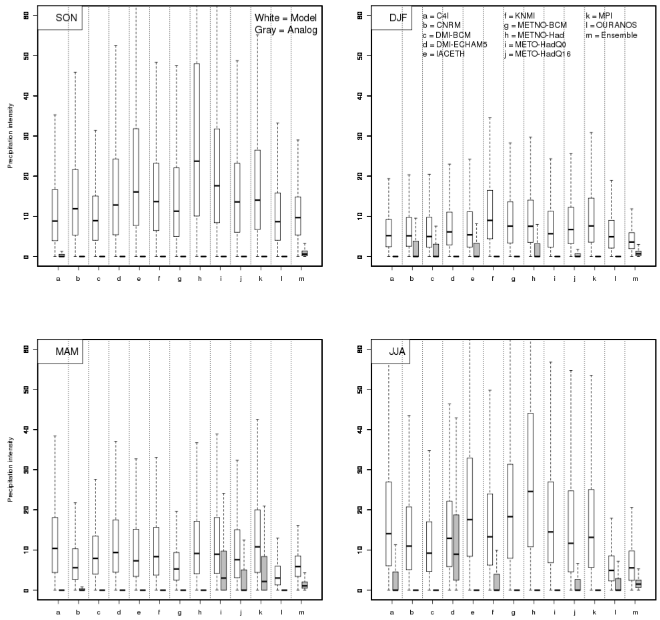


Figure 34: **INTERMEDIATE REPORT** - Absolute change between the 99.9th percentile of the past period's precipitation distributions and the 99.9th percentile of the future period's precipitation distributions (in absolute value) for each model and each season: boxplots for absolute changes in the observed precipitation and for the absolute changes in the analog precipitations, for model C4I. Nota: the outliers are not represented in these boxplots.

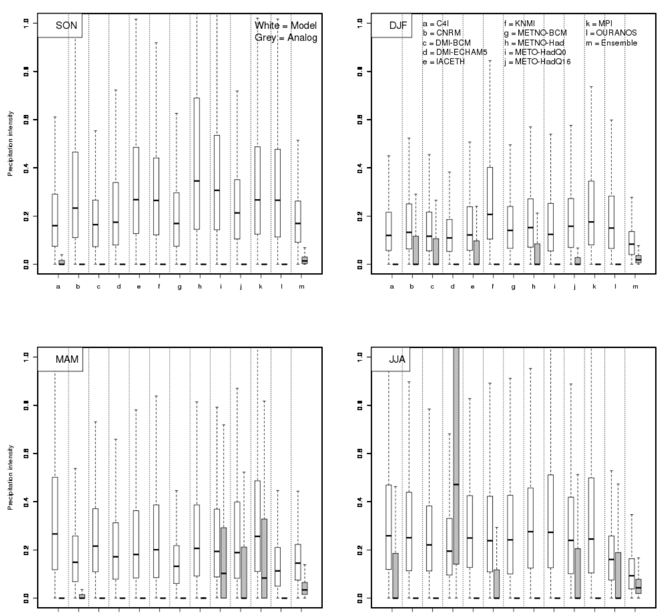


Figure 35: **INTERMEDIATE REPORT** - Relative change between the 99.9th percentile of the past period's precipitation distributions and the 99.9th percentile of the future period's precipitation distributions (in absolute value) for each model and each season: boxplots for relative changes in the observed precipitation and for the relative changes in the analog precipitations, for model C4I. Nota: the outliers are not represented in these boxplots.

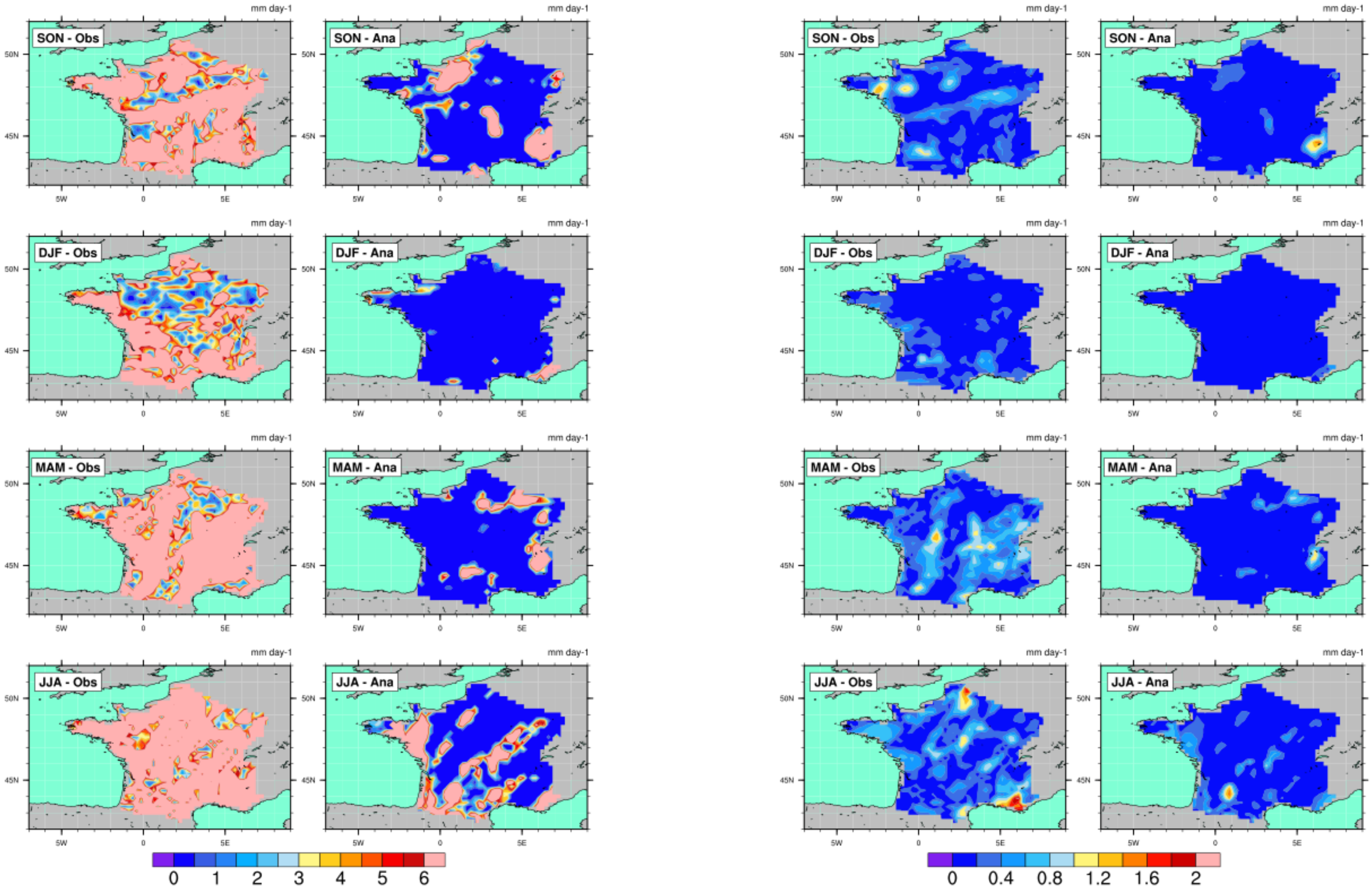


Figure 36: **INTERMEDIATE REPORT** - Absolute (left) and relative (right) changes between the 99.9th percentile of the past period's precipitation distributions and the 99.9th percentile of the future period's precipitation distributions (in absolute value) for each grid point and each season ("Obs" and "Ana" stand for "Observed" and "Analog") for model C4I

	Mean	**Min**	**1st quartile**	**Median**	**3rd quartile**	**Max**
Temporal correlation	0.30	-0.38	0.17	0.31	0.44	0.79

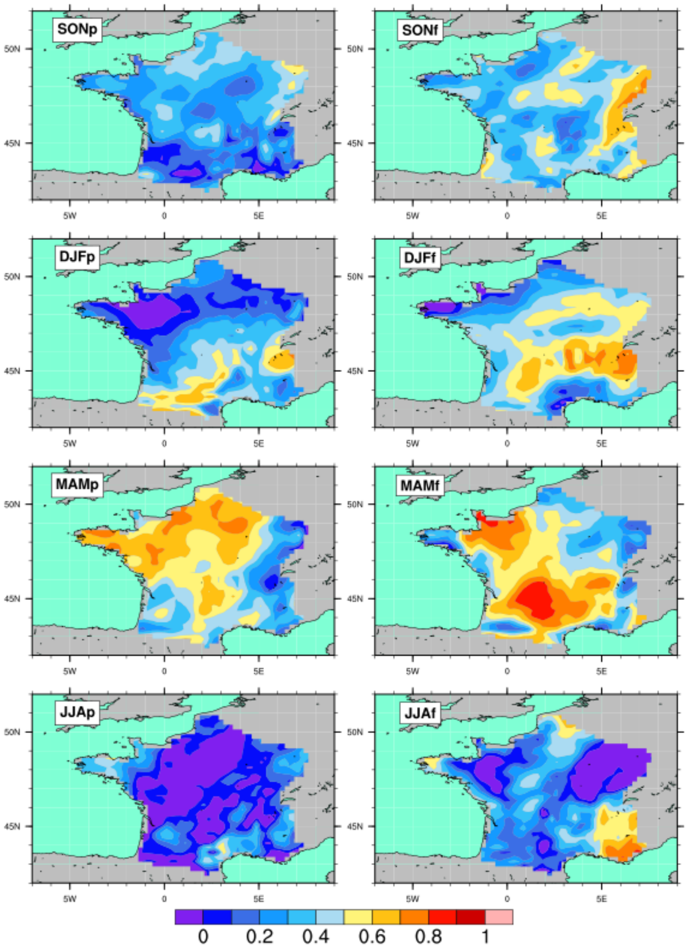
Table 9: Summary statistics for the interannual temporal correlation over the period 1979-2010 computed for each grid point for ERAI reanalyses

	SON	DJF	MAM	JJA
Mean	0.58	0.61	0.68	0.62
Min	0.34	-0.14	0.35	0.30
1st quartile	0.51	0.45	0.65	0.53
Median	0.57	0.68	0.68	0.65
3rd quartile	0.67	0.80	0.75	0.73
Max	0.78	0.92	0.83	0.80

Table 10: Summary statistics for the spatial correlation over France computed for each day of the period 1979-2010 for ERAI reanalyses by season

	SON	DJF	MAM	JJA
Mean	0.45	0.77	0.38	0.47
Variance	0.83	0.88	0.71	0.87
Distribution	0.78	0.99	0.67	0.70

Table 11: Mean equality tests for mean, variance and distribution for ERAI reanalyses by season



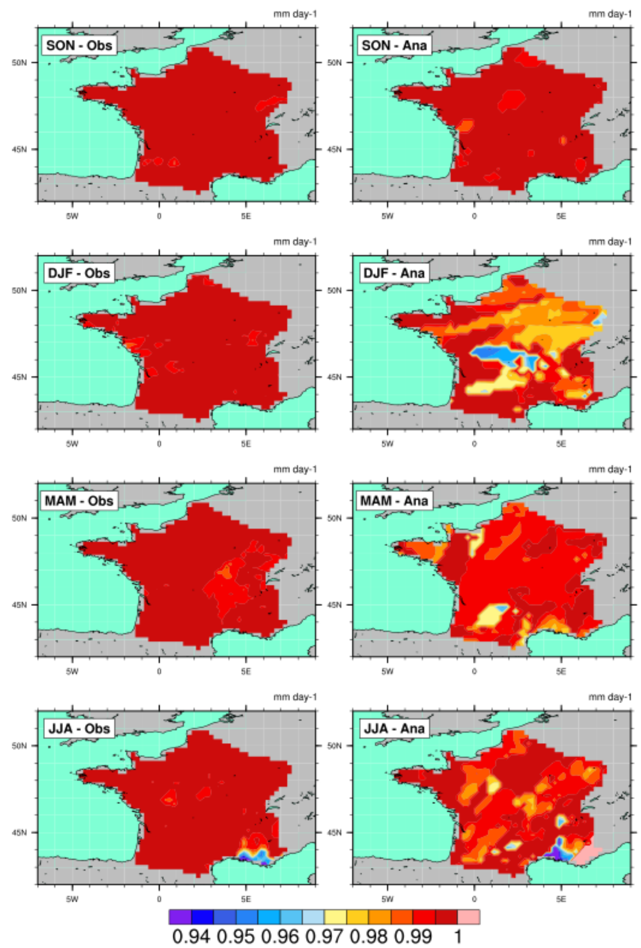


Figure 37: (left) **INTERMEDIATE REPORT** - For each grid point, correlation between the time series of observed precipitations and time series of analog precipitations ; results are mapped for each season and for both past and future periods ("p" and "f" after the season's acronym stand for "past" and "future) for model C4I

Figure 38: (right) **INTERMEDIATE REPORT** - Projection of the 99.9th quantiles of the past period on future precipitations for each grid point and each season ("Obs" stands for "Observed" and "Ana" for "Analog") for model C4I

	Mean	Min	Q1	Median	Q3	Max	σ
ACCESS1-3	804842	362557	498737	692137	1062730	1736880	1.3508e+11
bcc-csm1-1-m	1199930	473158	760031	1066370	1617330	2291620	2.52175e+11
BNU-ESM	1022420	325003	673618	979000	1365280	1889400	1.76573e+11
CanESM2	856922	312929	621575	819890	1068700	1685420	1.04423e+11
CMCC-CM	965619	413739	668287	925580	1237570	1800010	1.21394e+11
CNRM-CM5	834978	411135	586894	778529	1051920	1547610	8.22184e+10
GFDL-CM3	926498	481975	663181	840882	1161600	1681220	9.76343e+10
IPSL-CM5A-LR	915221	441093	650168	837366	1134560	1728780	9.73132e+10
IPSL-CM5A-MR	985095	384429	673753	911623	1242700	1993780	1.39934e+11
MIROC5	563884	259725	402734	531661	694778	1053430	3.80778e+10
MPI-ESM-MR	972880	442043	637710	940338	1264000	1824070	1.43119e+11
MRI-CGCM3	1078710	513328	734725	974012	1378240	2196060	1.76396e+11
NorESM1-M	880424	404498	587932	810836	161970	1571740	1.07312e+11

Table 12: Summary statistics for the variance of the APSL for different models ; the variance is computed for each grid point

	Mean	Min	Q1	Median	Q3	Max	σ
ACCESS1-3	798560	364221	540247	714915	1023700	1552010	1.0016e+11
bcc-csm1-1-m	1191590	493132	763916	1095480	1581480	2241120	2.38657e+11
BNU-ESM	1018800	387014	633486	953188	1363600	1881790	1.88152e+11
CanESM2	852330	385546	578730	759798	1097160	1652230	1.13297e+11
CMCC-CM	956658	420668	630182	865166	1252930	1813590	1.47103e+11
CNRM-CM5	828697	378012	551300	745569	1095070	1597090	1.05923e+11
GFDL-CM3	919695	427150	620225	821780	1187170	1784650	1.31304e+11
IPSL-CM5A-LR	911396	397202	607734	826280	1175450	1740100	1.34581e+11
IPSL-CM5A-MR	978504	406200	631888	900508	1284450	1836690	1.61483e+11
MIROC5	557937	263519	370037	502557	736193	1085130	4.73192e+10
MPI-ESM-MR	966465	378172	596767	897306	1299440	1795150	1.64637e+11
MRI-CGCM3	1072080	461887	715509	964577	1383330	2045130	1.85309e+11
NorESM1-M	873826	379931	573994	795512	1147490	1656350	1.23739e+11

Table 13: Summary statistics for the variance of the APSL projected on ERAI EOFs for several models ; the variance is computed for each grid point

	MEAN					VARIANCE					DISTRIBUTION				
	All	**SON**	**DJF**	**MAM**	**JJA**	**All**	**SON**	**DJF**	**MAM**	**JJA**	**All**	**SON**	**DJF**	**MAM**	**JJA**
bcc-csm1-1-m	0.71	0.41	0.79	0.82	0.41	0.95	0.79	0.90	0.95	0.86	0.99	0.84	1	1	0.75
BNU-ESM	0.70	0.58	0.82	0.79	0.62	0.92	0.80	0.87	0.92	0.94	0.99	0.92	0.99	1	0.84
CNRM-CM5	0.62	0.46	0.84	0.78	0.64	0.90	0.79	0.86	0.92	0.87	0.99	0.74	1	0.99	0.86
CSIRO-Mk3-6-0	0.63	0.47	0.80	0.77	0.43	0.93	0.73	0.87	0.91	0.86	0.99	0.83	1	1	0.68
GFDL-CM3	0.65	0.49	0.80	0.75	0.63	0.94	0.79	0.87	0.89	0.93	0.99	0.83	1	1	0.82
IPSL-CM5A-MR	0.66	0.46	0.81	0.77	0.46	0.94	0.81	0.91	0.95	0.91	0.99	0.79	1	1	0.68
MIROC5	0.56	0.40	0.84	0.74	0.45	0.89	0.80	0.85	0.89	0.87	0.99	0.81	0.99	0.99	0.68
MPI-ESM-MR	0.57	0.37	0.80	0.75	0.47	0.95	0.73	0.89	0.90	0.86	0.99	0.82	1	0.99	0.74
MRI-CGCM3	0.67	0.49	0.79	0.75	0.56	0.93	0.82	0.89	0.90	0.89	0.99	0.73	1	0.99	0.82
NorESM1-M	0.67	0.48	0.81	0.76	0.53	0.92	0.80	0.87	0.93	0.88	0.99	0.92	1	0.99	0.75

Table 14: Tests for equality of means, variances and distributions of the CMIP5 model data and the SAFRAN observations over the period 1979-2010 ('All' column shows the results obtained without distinguishing between seaons). For each model the average (over all grid points) result of a Student test, Fisher test and Kolmogorov-Smirnov test is computed (each test returns 0 if the null hypothesis of equality is rejected, 1 otherwise, for each grid point)

	PAST				FUTURE			
	SON	**DJF**	**MAM**	**JJA**	**SON**	**DJF**	**MAM**	**JJA**
C4I	0.39	0.45	0.34	0.24	0.58	0.44	0.29	0.40
CNRM	0.24	0.35	0.32	0.39	0.34	0.33	0.27	0.43
DMI-BCM	0.26	0.34	0.28	0.30	-	-	-	-
DMI-ECHAM5	0.31	0.37	0.32	0.24	0.39	0.32	0.28	0.33
IACETH	0.27	0.33	0.26	0.22	0.63	0.41	0.36	0.29
KNMI	0.25	0.35	0.29	0.29	0.36	0.32	0.25	0.32
METNO-BCM	0.30	0.35	0.27	0.29	0.37	0.32	0.31	0.34
METNO-Had	0.25	0.31	0.34	0.28	0.58	0.53	0.39	0.49
METO-HadQ0	0.27	0.31	0.29	0.37	0.51	0.35	0.29	0.33
METO-HadQ16	0.36	0.35	0.34	0.43	0.47	0.38	0.35	0.44
MPI	0.27	0.32	0.27	0.31	0.33	0.35	0.31	0.34
OURANOS	-	-	-	-	-	-	-	-

Table 15: Tests for equality between the means of observed and analog precipitations obtained with the Dayon method: for both periods the average result of a Student t-test for mean equality is computed (the test returns 0 if the null hypothesis of mean equality is rejected, 1 if not)

	PAST				FUTURE			
	SON	**DJF**	**MAM**	**JJA**	**SON**	**DJF**	**MAM**	**JJA**
C4I	0.79	0.76	0.76	0.82	0.88	0.83	0.77	0.88
CNRM	0.73	0.72	0.74	0.82	0.79	0.75	0.75	0.87
DMI-BCM	0.75	0.71	0.73	0.81	-	-	-	-
DMI-ECHAM5	0.73	0.75	0.75	0.79	0.79	0.70	0.75	0.84
IACETH	0.77	0.74	0.74	0.85	0.88	0.81	0.83	0.90
KNMI	0.73	0.69	0.74	0.81	0.79	0.67	0.75	0.83
METNO-BCM	0.76	0.73	0.74	0.83	0.79	0.72	0.76	0.85
METNO-Had	0.74	0.74	0.75	0.84	0.85	0.76	0.77	0.90
METO-HadQ0	0.76	0.70	0.77	0.84	0.87	0.79	0.79	0.88
METO-HadQ16	0.75	0.70	0.74	0.84	0.86	0.78	0.81	0.90
MPI	0.74	0.70	0.75	0.80	0.79	0.73	0.75	0.82
OURANOS	-	-	-	-	-	-	-	-

Table 16: Tests for equality between the variances of observed and analog precipitations obtained with the Dayon method: for both periods the average result of a Student t-test for mean equality is computed (the test returns 0 if the null hypothesis of mean equality is rejected, 1 if not)

	PAST				FUTURE			
	SON	**DJF**	**MAM**	**JJA**	**SON**	**DJF**	**MAM**	**JJA**
C4I	0.47	0.49	0.51	0.37	0.63	0.49	0.45	0.51
CNRM	0.60	0.50	0.51	0.60	0.66	0.50	0.49	0.66
DMI-BCM	0.69	0.53	0.46	0.49	-	-	-	-
DMI-ECHAM5	0.72	0.68	0.54	0.63	0.64	0.50	0.60	0.80
IACETH	0.39	0.48	0.45	0.30	0.65	0.49	0.45	0.36
KNMI	0.32	0.47	0.42	0.39	0.46	0.42	0.47	0.48
METNO-BCM	0.48	0.49	0.45	0.41	0.50	0.41	0.47	0.48
METNO-Had	0.45	0.36	0.50	0.40	0.70	0.66	0.57	0.71
METO-HadQ0	0.45	0.50	0.48	0.42	0.67	0.56	0.54	0.56
METO-HadQ16	0.60	0.49	0.49	0.57	0.68	0.49	0.62	0.63
MPI	0.43	0.44	0.43	0.53	0.51	0.45	0.58	0.61
OURANOS	-	-	-	-	-	-	-	-

Table 17: Tests for adequation of the distributions of observed and analog precipitations obtained with the Dayon method: for both periods the average result of a Student t-test for mean equality is computed (the test returns 0 if the null hypothesis of mean equality is rejected, 1 if not)

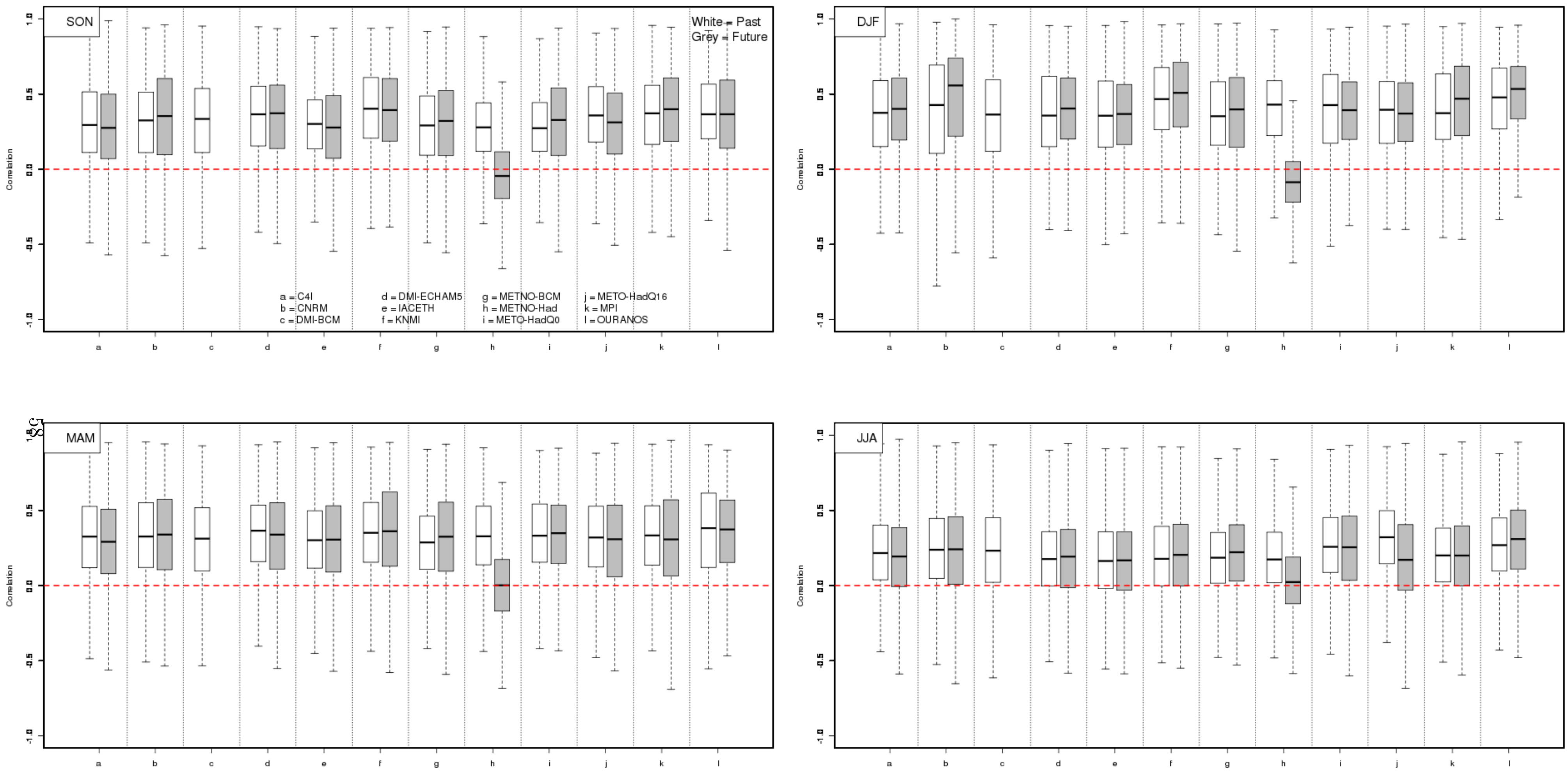


Figure 39: Boxplots for temporal correlations at each grid point between observed and analog precipitations by model and by season, for both past (white) and future (grey) periods, obtained with the Dayon method. Nota: the outliers are not represented in these boxplots.

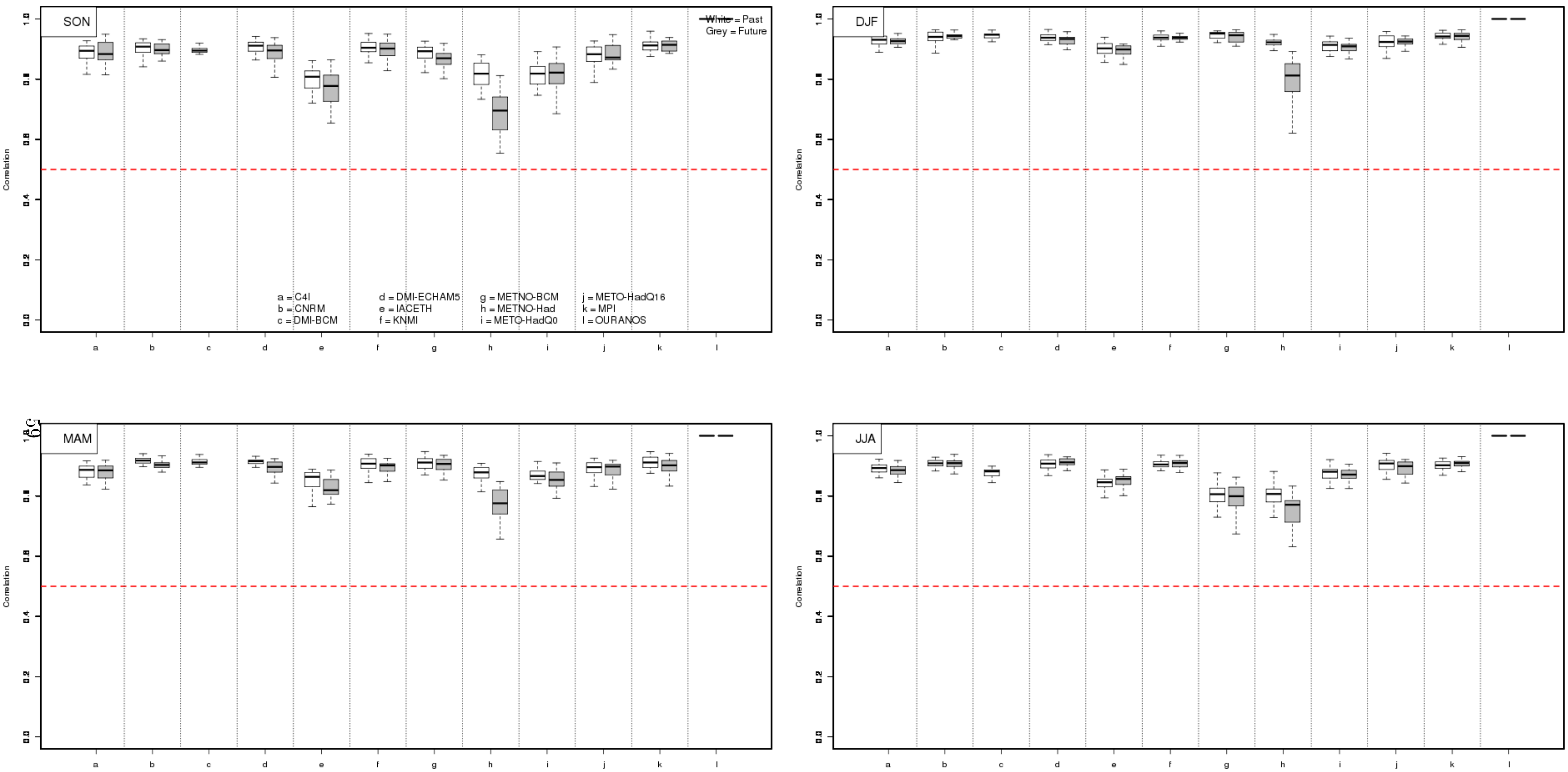


Figure 40: Boxplots for spatial correlations at each grid point between observed and analog precipitations by model and by season, for both past (white) and future (grey) periods, obtained with the Dayon method. Nota: the outliers are not represented in these boxplots.