An adaptive two-stage analog/regression model for probabilistic prediction of small-scale precipitation in France.

Jérémy Chardon[1], Benoit Hingray[1], and Anne-Catherine Favre[1]

[1]Univ. Grenoble Alpes, CNRS, IRD, Grenoble INP, IGE, F-38000 Grenoble, France

Correspondence to: Benoit Hingray (benoit.hingray@univ-grenoble-alpes.fr)

Abstract. Statistical Downscaling Models (SDMs) are often used to produce local weather scenarios from large scale atmospheric information. SDMs include transfer functions which are based on a statistical link identified from observations between local weather and a set of large scale predictors. As physical processes driving surface weather vary in time, the most relevant predictors and the regression link are likely to vary in time too. This is well known for precipitation for instance and the link is thus often estimated after some seasonal stratification of the data. In this study, we present a two-stage analog/regression model where the regression link is estimated from atmospheric analogs of the current prediction day. Atmospheric analogs are identified from fields of geopotential heights at 1000 and 500 hPa. For the regression stage, two Generalized Linear Models are further used to model the probability of precipitation occurrence and the distribution of non-zero precipitation amounts respectively. The two-stage model is evaluated for the probabilistic prediction of small-scale precipitation over France. It noticeably improves the skill of the prediction for both precipitation occurrence and amount. As the analog days vary from one prediction day to another, the atmospheric predictors selected in the regression stage and the value of the corresponding regression coefficients can vary from one prediction day to another. The model allows thus for a day-to-day adaptive and tailored downscaling. It can also reveal specific predictors for peculiar and non-frequent weather configurations.

1 Introduction

Statistical Downscaling Models (SDMs) have been widely used to generate local weather scenarios for past or future climates from outputs of climate models (e.g. Wilby et al., 1999; Hanssen-Bauer et al., 2005; Boé et al., 2007; Lafaysse et al., 2014) and to produce local weather forecasts from outputs of numerical weather prediction models (e.g. Obled et al., 2002; Gangopadhyay et al., 2005; Marty et al., 2012; Ben Daoud et al., 2016). For the recent years, they have been also used to reconstruct past weather conditions from atmospheric reanalysis data (e.g. Kuentz et al., 2015; Caillouet et al., 2016).

Among the different SDM approaches presented over the last decades (see Maraun and al., 2010, for a review), perfect prognosis SDMs make use of the physical relationships that exist between some large scale atmospheric parameters and local weather variables. Local weather scenarios can then be produced for any prediction day conditionally on the large scale atmospheric configuration observed or simulated for this day, where the "prediction day" refers here to some future, past or present simulation day, depending on the application context at hand (e.g. forecasting, simulation, reconstruction). Perfect prognosis SDMs include transfer functions, weather-type based models and methods based on atmospheric analogs. In the latter case,

atmospheric analog days of the current prediction day are searched for on the basis of some atmospheric similarity criterion in the historical database. The weather variables observed for the most similar day, for one similar day chosen randomly or for a selection of the k-most similar days are then used as a weather scenario for the prediction day of interest (e.g. Dayon et al., 2015; Raynaud et al., 2016).

5   Transfer functions mainly consist of regression models where the expected value of the predictand for time t is expressed as a linear or non-linear function of a set of predictors. For precipitation, the regression can be achieved with Multiple Linear Regressions or Generalized Linear Models (GLMs) which extent the linear regression to non-Gaussian data (e.g. Nelder and Wedderburn, 1972; Stern and Coe, 1984; Asong et al., 2016). Transfer functions can also make use of classification and regression trees (CART) (e.g. Gaitan et al., 2014) artificial neural networks or least squares support vector machines (e.g.
10   Campozano et al., 2016).

The downscaling relationship used in transfer functions is usually established empirically between a selection of large-scale predictors and the predictand (e.g. precipitation occurrence) from a set of observations available for recent decades. As physical processes driving surface weather vary in time, the most relevant predictors and the downscaling link are however expected to vary in time too. When inferred from all observations available for a given period, the downscaling relationship – which is thus
15   likely inferred from a heterogeneous ensemble of weather configurations – is consequently likely to be sub-optimal. To reduce this potential limitation, the parameterization of the relationship is often estimated after some data stratification. In the usual calendar stratification, one parameter set is for instance optimized for each calendar month or season (e.g. Nasseri et al., 2013). The stratification can also be based on some weather type information. In this case, a set of parameters is usually estimated for each weather type of a given pre-established weather type classification (e.g. Enke and Spegat, 1997). Often applied, this
20   weather-type based approach is expected to allow for a better identification of the most important driving large-scale variables and consequently for a more relevant downscaling. An obvious limitation however remains for prediction days that do not clearly belong to one specific weather type (e.g. prediction days that are close to the "weather frontiers" delimiting two or more weather types). Those days are indeed likely to be rather dissimilar to the weather configurations that each weather type is expected to characterize, making the downscaling relationships to be used not suited anymore or, at least, sub-optimal.

25   A smoother weather type like approach consists in defining the weather type from all atmospheric situations that are similar to the situation of the prediction day. The ensemble of days from which the downscaling link can be identified is thus expected to be rather homogeneous and to rather well inform the large to small scale link sought for the considered prediction day. This is in turn expected to make the link stronger and to improve the prediction (e.g. Woodcock, 1980). Such an approach can be actually achieved identifying for each prediction day the transfer function from atmospheric analogs of that day. To our
30   knowledge, this two-stage downscaling approach, combining in turn the two popular analog and transfer function methods, has been only explored in few previous studies. In Ribalaygua et al. (2013), it was found to improve the probabilistic prediction of local surface temperature in the Spanish Iberian Peninsula. The multiple linear regression of the regression stage, estimated from the 150 most similar atmospheric analogs of the prediction day of interest, uses forward and backward stepwise selection of predictors from a set of four potential predictors (thickness of the air column and three temperature indexes of previous
35   days). For precipitation, the authors did not test the potential of the two-stage combination, building directly the predictions

from the precipitation observations of the 30 most similar atmospheric analogs. In the deterministic approach presented by Ibarra-Berastegi et al. (2011), incorporating the regression stage (with 79 potential atmospheric predictors) was found to allow a clear though not overwhelming improvement of precipitation prediction over the simple analog based predictions. A multiple linear regression model was also applied here for the regression stage.

5   In the present study, we present a two-stage analog/regression downscaling model for the probabilistic prediction of small-scale daily precipitation: for each prediction day, the statistical downscaling link between some large scale atmospheric predictors and small-scale precipitation is estimated from large scale and local scale observations available from an ensemble of days which are atmospheric analogs to the prediction day. The analog model (AM) used for the analog stage is based on developments from different studies initially focusing on the probabilistic quantitative precipitation forecasts in southern France (e.g.

10   Bontron and Obled, 2005; Marty et al., 2012) and extended to the prediction of precipitation on larger spatial domains (e.g. Chardon et al., 2014). The statistical distribution of daily precipitation is strongly non-gaussian with a non-negligible mass in zero (corresponding to the probability of a dry day), and a skewed distribution for non-zero daily amounts. For the regression stage, we thus do not use a multiple linear regression model as in previous studies but a 2-part GLM approach where the probability of precipitation occurrence and the distribution of wet-days amounts are modeled separately following Chandler

15   and Wheater (2002) and Mezghani and Hingray (2009). Conversely to the work of Ibarra-Berastegi et al. (2011), this allows predicting the full distribution of precipitation, including the probability of a wet day. In this two-stage analog/regression approach, the analogs change from one prediction day to the other. This makes the statistical downscaling link potentially adaptive, i.e. the predictors and the regressions parameters are likely to vary from one day to the other.

     As mentioned above, SDMs are used for the simulation of local weather scenarios in different contexts, e.g. local weather

20   forecasts, reconstructions or for climate impact studies. No specific context is considered here and the two-stage model could be further considered for either forecasting, reconstruction or future projections. Depending on its intended use, some specific issues would obviously apply, calling for specific focused analyses and developments. For instance, the large scale atmospheric parameters to be considered as predictors would depend on the dataset considered (e.g. atmospheric reanalyses, climate model or numerical weather prediction models) as a result of their intrinsic quality (e.g. Caillouet et al., 2016). The development of

25   climate projections would require to check the temporal transferability of the model in a modified climate context and would thus likely also condition the selection of the predictors as highlighted by (e.g. Dayon et al., 2015). These context specific issues are not considered here. Our main objectives are to present the principles of the two-stage analog/regression approach developed for the prediction of small-scale precipitation, to assess its predictive power for both precipitation occurrence and amount, and to give some insight on its adaptive behavior and thus on the temporal variability of the downscaling link. For

30   this, we explore the model skill and behavior for the prediction of daily precipitation for a large number of sites in France.

     The paper is structured as follows: Section 2 describes the data and section 3 the two-stage downscaling model. Section 4 presents the skill of the model for the prediction of both precipitation occurrence and amount. The adaptive behavior of the model is considered in section 5 and section 6 concludes.

2 Data

The predictand is the daily small-scale precipitation estimated for the 1982-2001 period over 8,981 grid cells of $8 \times 8 \text{ km}^2$ covering the continental French territory. The predictand is "local" precipitation, i.e. precipitation at a given grid cell. Each of the 8,981 grid cells is thus considered in turn in the following independently of the other cells. In other words, the predictions do not target precipitation fields. Small-scale precipitation data are obtained from the SAFRAN analysis produced for several surface variables at hourly time step by MeteoFrance (Quintana-Segui and al., 2008; Vidal et al., 2010). SAFRAN precipitation estimates are obtained each day from the closest measurement stations. They are considered as pseudo-observations.

Atmospheric predictors are taken from the European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis (ERA-40, Uppala and al., 2005). This global meteorological re-analysis is available on a $1.125° \times 1.125°$ grid with a 6-hourly temporal resolution.

For the analog stage, predictors are the 1000hPa and 500hPa geopotential height fields over a large spatial domain (roughly $\text{Lat} = 10°$, $\text{Lon} = 8°$) centered on the target location. These predictors have been found to be the most informative large scale predictors to be used in this context for France (e.g. Guilbaud and Obled, 1998; Obled et al., 2002; Radanovics et al., 2013). They also correspond to the best large scale predictors of daily precipitation for different regions in Europe with contrasted meteorological regimes (Raynaud et al., 2016).

For the regression stage, 22 other predictors were also considered. The selection gathers most predictors considered in previous studies over Europe (e.g. Hanssen-Bauer et al., 2005; Wetterhall et al., 2009; Horton et al., 2012; Raynaud et al., 2016). They include predictors characterizing the thermal state of the atmosphere, its dynamics, the atmospheric water content, its thermo-dynamical instability (see Table 1). As potential predictor, we also consider the occurrence of precipitation on the previous day. All predictors are here scalar variables. Atmospheric predictors are estimated on a daily time step (mean of the four values available at 6, 12, 18 and 24h00 UTC) from the four ERA-40 grid cells surrounding the prediction grid cell (inverse distance interpolation).

To avoid the multi-colinearity in the predictors for the regression, we identified a subset of uncorrelated predictors. The cross-correlations between all predictor pairs were first estimated on an annual basis from all available data. The correlation structure can however differ from one atmospheric configuration to the other. The set of uncorrelated predictors could thus differ from one prediction day to the other. We thus repeated the correlation analysis for each prediction day, using for this estimation the predictor values observed for the 100 nearest atmospheric analogs identified for this day. The main features of the inter-variable correlations were found to be roughly independent of the day (not shown). The final subset of uncorrelated predictors is highlighted in Table 1. These predictors are tested for the prediction of both precipitation occurrence and amount.

A large number of different possible predictor sets can be built from these predictors. In the present work, for the sake of robustness, we consider that a maximum of four predictors can be integrated in a given regression model. Predictors are obviously expected to be both day and location specific. In the present work, for the sake of simplicity and readability, we select them from a unique set of four potential predictors. This allows us reducing the degrees of freedom in the model and better highlighting its skill and adaptive behavior.

For each predictand, the set of the four potential predictors was selected as follows. For 12 SAFRAN grid cells uniformly distributed over the French territory, we first identified with a standard iterative forward/backward algorithm the four-predictor set which lead to the best prediction skill for the all-days configuration. From the twelve different sets respectively obtained for the 12 grid cells, we finally retained the set which leads in average to the best prediction skill for the 8,981 SAFRAN grid cells.

For precipitation occurrence, this best four-predictor set is constituted from the relative humidity R_{700}, the helicity H, the vertical velocity of the air at 700hPa W_{700} and the precipitation occurrence $Occ-1$ of the day before the prediction day. For precipitation amount, the best four-predictor set is similar except that the occurrence of the previous day $Occ-1$ is replaced by the 700 hPa air temperature T_{700}. Note that the selection of predictors R_{700}, W_{700} and T_{700} is consistent with results of several past studies in the region (e.g. Ben Daoud et al., 2016).

Predictors considered for the analog and regression stages obviously inform about different features of the atmosphere state for different scales. Geopotential fields, by their spatial extent, characterize the large-scale atmospheric circulation configuration (the spatial domain of several thousands of kilometers, includes a part of North-Eastern Atlantic and covers France and a part of neighboring countries) whereas scalar predictors used in the regression stage are descriptive of a more local (and mostly thermodynamic) state of the atmosphere (the spatial domain of several hundreds of kilometers is roughly centered above the target location).

3 The two-stage analog/regression model (SCAMP)

As illustrated in Fig. 1, the cumulative distribution function (cdf) F_Y of precipitation Y at a given site (grid cell) can be expressed for any given day as the composition of the no-precipitation occurrence probability $1-\pi$ and the cdf F_Q of the precipitation amount Q for non-zero precipitation:

$$F_Y(y) = (1-\pi) + \pi \cdot F_Q(q=y), \tag{1}$$

where π is the precipitation occurrence probability, y and q correspond to the precipitation value with regard to the whole precipitation distribution and to the non-zero precipitation distribution respectively.

In the present work, the cdf of precipitation is modeled for each grid cell and each prediction day with GLMs (Coe and Stern, 1982; Stern and Coe, 1984), estimated for this specific day from atmospheric analogs of the day. The probability of precipitation occurrence and the cdf of the non-zero precipitation amount are modeled separately.

In the following, we first describe the Analog Model (AM) used to identify atmospheric analog days (section 3.1) and the GLMs applied in the regression stage (section 3.2).

As discussed later, one can face prediction days where the regression stage fails, i.e. where the regression parameters are not significantly different from zero at the chosen significance level ($\alpha = 5\%$). For such days, we use the Analog Model as backup prediction model. The backup model can be used for precipitation occurrence probability, for non-zero precipitation amount or for both predictands simultaneously.
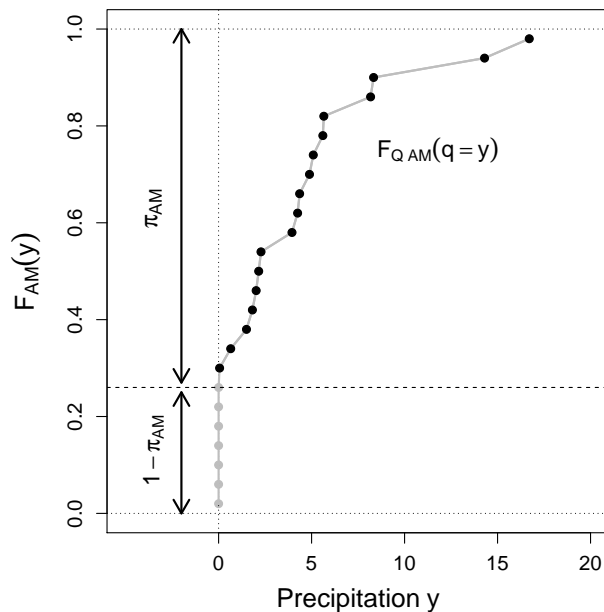
Figure 1. Cumulative distribution function (cdf) of precipitation amount for a given prediction day (in grey) at a given grid cell. For illustration, the prediction here corresponds to the empirical cdf achieved with the Analog Model (AM) mentioned in section 3.3. The contribution of the precipitation amount $F_{Q,AM}$ cdf to the overall cdf is highlighted in black (c.f. Eq. 1).

The way these different models are combined to finally give, for the current prediction day, a probabilistic prediction of precipitation is presented in section 3.3. In the following, this two-stage analog/regression model is further referred to as SCAMP (SCAMP stands for Sequential Constructive atmospheric Analogs for Multivariate weather Prediction and refers to the model presented by Raynaud et al. (2016) for the multivariate prediction of precipitation/temperature/radiation/wind).

3.1 Atmospheric analogs

The atmospheric analog days retained for the regression stage are identified with an Analog Model defined from the developments of several past studies in France (e.g. Obled et al., 2002; Marty et al., 2012; Chardon et al., 2014).

For any given prediction day (e.g. May-31's, 2018), the analog days retained for the regression are the N_d days that are most similar to that day in terms of large scale atmospheric circulation. The similarity is assessed using the Teweless-Wobus Score (TWS, Teweless and Wobus, 1954) applied to the geopotential height at 1000 hPa and 500 hPa at 12h and 24h UTC respectively. The TWS compares the shapes of geopotential fields, and thus informs on the localization of low and high pressure systems and on the origin of air masses. Note that the N_d analog days are identified within a restricted pool of candidate days, namely all days of the archive that are included in a calendar window of $\pm$ 30 calendar days centered on the prediction day (for the prediction of May-31's, 2018, candidates are all May 1st to all June 30's from all years of the archive). The prediction day (e.g. May-31's, 2018) and its 5 preceding and following days are excluded from the candidates. In the present work, the archive

period corresponds to 1982-2001 (20 years) and we used the 100-nearest atmospheric analog days to estimate the GLMs in the regression stage.

Following Chardon et al. (2014), the domain considered to estimate the atmospheric similarity was optimized for each target location. A different analog model was thus considered for each of the 8,981 SAFRAN grid cells. For each prediction day, the analog days thus likely differ from one SAFRAN grid cell to the next (see Chardon et al., 2014, for illustration).

3.2 Regression stage with GLMs

The cdf of precipitation is then modeled for each prediction day with GLMs estimated for this specific day from the atmospheric analogs of the day. GLMs make the cdf depending on some covariates, atmospheric predictors in the present case. For each prediction day, the probability of precipitation occurrence π was modeled with a GLM in the form of a logistic regression as:

$$\log\left(\frac{\pi}{1-\pi}\right) = \mathbf{x^{o T}}\beta^{\mathbf{o}}, \tag{2}$$

where $\mathbf{x^o}$ is the scalar vector of the K_o predictors $(x_1^o, x_2^o, ..x_{K_o}^o)$ and $\beta^{\mathbf{o}}$ the scalar vector of the K_o corresponding regression coefficients $(\beta_1^o, \beta_2^o, ..\beta_{K_o}^o)$.

For the non-zero precipitation amount, we used a GLM with the gamma distribution and the log link function. The expected amount μ of non-zero precipitation is therefore here expressed as:

$$\log\left(\mu\right) = \mathbf{x^{q T}}\beta^{\mathbf{q}}, \tag{3}$$

where $\mathbf{x^q}$ denotes the scalar vector of the K_q predictors $(x_1^q, x_2^q, ..x_{K_q}^q)$ and $\beta^{\mathbf{q}}$ the scalar vector of the corresponding regression coefficients $(\beta_1^q, \beta_2^q, ..\beta_{K_q}^q)$. The shape parameter ν of the gamma distribution is computed from the variance σ^2 of non-zero precipitation amounts estimated from Pearson's residuals (McCullagh and Nelder, 1989) as:

$$\sigma^2 = \frac{1}{\{N_q - (K_q + 1)\}} \sum_{i=1}^{N_q} \frac{(q_i - \mu)^2}{\mu^2}, \tag{4}$$

where N_q is the number of non-zero precipitation data q_i considered in the analysis. As the shape parameter ν equals the inverse of the variance $1/\sigma^2$, the gamma distribution F_Q modeling the precipitation amount thus follows a gamma distribution of this type: $\Gamma\left(\nu, \alpha = \mu/\nu\right)$.

For any given prediction day, the estimation of both GLM models practically proceeds as follows:

- The precipitation state (wet or dry), the precipitation amount and the value of the different potential predictors are extracted for the N_d nearest analogs of the day. The precipitation state of a given day is considered to be wet if the precipitation amount for this day is higher or equal to 0.1 mm. It is described with a binary precipitation occurrence variable $\mathbf{O}$, set to 1 for the wet case, 0 for the dry case.

- For occurrence probability, different sets of predictors are considered in turn. For each set, the parameters of the occurrence GLM are estimated from the predictors/occurrence values available for the N_d analogs.

– For precipitation amount, different sets of predictors are again considered in turn. For each set, the parameters of the GLM are estimated from the predictors/amount values available from the analog days which are wet (N_q, the number of days considered here for the regression, is therefore smaller or equal to N_d and varies a priori from one target day to another).

For the considered prediction day, the different sets of predictors considered in turn are built from the four potential predictors identified in the preliminary work (cf. section 2). For occurrence probability (resp. precipitation amount), the four potential predictors actually allow to build fifteen different sets of predictors, further denoted as "regressive structures" in the following (cf list in Table 2). For each regressive structure, the regression coefficients of corresponding GLMs are estimated using the Iterative Re-weighted Least Squares algorithm (IRLS, Nelder and Wedderburn, 1972). The prediction skills of the different regressive structures are then compared and the regressive structure (predictor set) which minimizes the Bayesian Information Criterion is retained for the prediction (Schwarz, 1978; Akaike, 1974) (only the regressive structures for which all coefficients are significant at a 5% level are compared; the significance is estimated with the Z-test (resp. the Student t-test)).

The prediction of the occurrence probability (resp. the expected precipitation amount) for the prediction day is finally obtained from the best occurrence (resp. amount) GLM, using the values of the predictors observed for that prediction day. The final distribution of precipitation F_Y is obtained by combining the issued occurrence probability π and the amount distribution F_Q according to Eq. 1.

3.3 The Analog Model as benchmark and backup prediction model

The N_d-nearest analog days identified with the AM can also be directly used, without further regression stage, for a probabilistic prediction. In the following, we also consider predictions obtained with the 25 nearest analog days (for the AM considered here, 25 was found to give the best prediction skill for France by Chardon et al. (2014)). In this case, the precipitation cdf for the prediction day is simply the empirical distribution of the precipitation values observed for these 25 analogs. The predictions obtained with this analog model, further called AM$_{25}$, are used as a benchmark to assess the prediction skill of the two-stage analog/regression approach. In addition, they were used as backup prediction for days for which the regression stage failed in the two-stage approach. One can actually face the situation where no GLM satisfies the significance conditions required for the regression coefficients. This can occur for precipitation occurrence probability, for non-zero precipitation amount or for both predictands simultaneously. In such cases, AM$_{25}$ is applied as backup prediction model.

If the significance conditions cannot be satisfied for the precipitation occurrence GLM, the occurrence probability π is set to that obtained with AM$_{25}$. It thus simply corresponds to the empirical probability $\pi_{\mathrm{AM}_{25}}$ of precipitation occurrence derived from the 25 analog days of AM$_{25}$ as:

$$\pi \equiv \pi_{\mathrm{AM}_{25}} = \frac{1}{25} \sum_{i=1}^{25} o_i. \tag{5}$$

Similarly, if the significance conditions cannot be satisfied for the precipitation amount GLM, the distribution F_Q is estimated with the empirical distribution $F_{Q,\text{AM}_{25}}$ derived with AM_{25} as:

$$F_Q(q) \equiv F_{Q,\text{AM}_{25}}(q) = \frac{F_{\text{AM}_{25}}(q) - (1 - \pi_{\text{AM}_{25}})}{\pi_{\text{AM}_{25}}}, \tag{6}$$

where $F_{\text{AM}_{25}}$ corresponds to the empirical cdf estimated from all precipitations (null and positive) related to the 25 analog days. Note also that if the number N_q of humid analog days is low ($N_q < 10$), the estimation of a GLM is not expected to be robust. When this case appears, F_Q is also set to the cdf obtained with AM_{25}.

As illustrated in Fig. 2, four prediction cases are thus achieved with the two-stage approach. They correspond respectively to cases where AM_{25} is used to backup the prediction of the whole precipitation distribution (case 1), where AM_{25} is applied to backup the amount cdf prediction (case 2), where AM_{25} is used to backup the occurrence probability prediction (case 3) and where the regression stage could be activated for both occurrence and amount (case 4).

Note that the regression stage achieved with GLMs can also be seen as a way to refine the estimation of the cdf that could have been obtained directly with the backup (and benchmark) AM_{25} analog model. The refinement leads to update the occurrence probability and/or the cdf of non-zero precipitation amount.

As described previously, the two-stage analog/regression prediction process is repeated for each prediction day in turn. As the analog days vary from one prediction day to another, the predictors selected in the regression stage and the value of the corresponding regression coefficients are expected to vary from one prediction day to the other. The two-stage model SCAMP allows thus for a day-to-day adaptive and tailored downscaling.

3.4 Model evaluation

The prediction skill of the downscaling model is assessed with probabilistic scores usually used to evaluate Ensemble Prediction Systems (EPS). Let us consider a given EPS, denoted as $\mathcal{P}$.

The Brier Score (Brier, 1950; Murphy, 1973) first evaluates the ability of EPS $\mathcal{P}$ to predict precipitation occurrence. When estimated over M prediction days, the mean Brier Score $\overline{\text{BS}}$ reads:

$$\overline{\text{BS}} = \frac{1}{M} \sum_{i=1}^{M} [p_i - o_i]^2, \tag{7}$$

where, for a given prediction day i, p_i is the occurrence probability issued by EPS $\mathcal{P}$ and o_i is the effective precipitation occurrence for this day ($o_i = 1$ for a wet day, $= 0$ otherwise).

The ability of EPS $\mathcal{P}$ to estimate the precipitation amount is evaluated with the Continuous Ranked Probability Score (CRPS, Brown, 1974; Matheson and Winkler, 1976). When estimated over M prediction days, the mean CRPS reads:

$$\overline{\text{CRPS}} = \frac{1}{M} \sum_{i=1}^{M} \int_{-\infty}^{+\infty} [F_i(x) - H_{y_i}(x)]^2 dx, \tag{8}$$
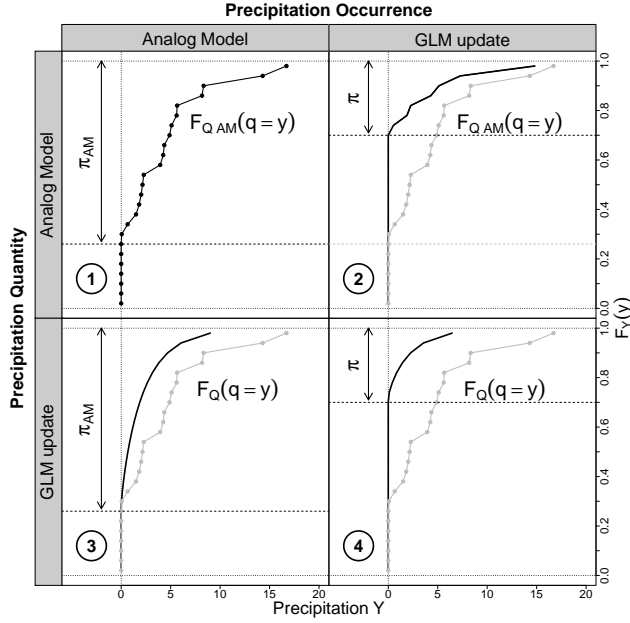
Figure 2. Illustrations of the four cases met for the issue of $F_Y(y)$ by the two-stage model. Case 1: None of the occurrence and amount (quantity) GLMs could be retained during the regression stage: AM_{25} is used to predict the whole precipitation distribution. Case 2: Only the occurrence GLM could be retained. It gives the estimated occurrence probability. The distribution of non-zero precipitation comes from AM_{25}. Case 3: Only the amount (quantity) GLM could be retained. It gives the distribution of non-zero precipitation. The occurrence probability is the empirical occurrence probability from AM_{25}. Case 4: Both occurrence and amount (quantity) GLMs could be estimated: they respectively give the occurrence probability and the distribution of non-zero precipitation, to be further combined for the full distribution of precipitation.

where, for a given prediction day i, H_{y_i} and F_i denote respectively the cdf of the observation y_i and the cdf derived from EPS $\mathcal{P}$. x denotes the predictand quantiles of the cdfs. Note that H_{y_i} corresponds to the Heaviside function where $H_{y_i} = 1$ if $x \geq y_i$ and $H_{y_i} = 0$ otherwise.

For this evaluation, the probabilistic prediction of the predictand y is here described, for each prediction day, with a dis-
5 cretized cdf composed of N values, with $N = 25$. When AM_{25} is used as backup model, the N values are the precipitation observations of the 25th analog days. When the prediction is issued with SCAMP , the N values are those of the 25 percentiles $(k-0.5)/25$, k in $[1, N]$, of the predicted cdf F_Y.

In the following, we discuss the prediction skill for precipitation occurrence and amount with the Brier Skill Score (BSS) and the Continuous Ranked Probability Skill Score (CRPSS) respectively. Both scores normalize the prediction skill of EPS
10 $\mathcal{P}$ with that obtained with a reference EPS $\mathcal{P}_\varphi$. $\mathcal{P}_\varphi$ is here a climatological EPS based on a calendar climatology defined for each prediction day by the precipitation distribution of all days belonging to a seasonal window ($\pm$ 30 days) centered on the

corresponding calendar day. In this context, the BSS and CRPSS respectively read:

$$\text{BSS} = 1 - \frac{\overline{\text{BS}}}{\overline{\text{BS}}_\varphi}, \tag{9}$$

and

$$\text{CRPSS} = 1 - \frac{\overline{\text{CRPS}}}{\overline{\text{CRPS}}_{\mathcal{P}_\varphi}}, \tag{10}$$

where $\overline{\text{BS}}_\varphi$ and $\overline{\text{CRPS}}_{\mathcal{P}_\varphi}$ correspond to the mean BS and the mean CRPS obtained with EPS $\mathcal{P}_\varphi$. For both scores, a negative value indicates that the prediction obtained with EPS $\mathcal{P}$ is worse than the prediction obtained with the climatological EPS $\mathcal{P}_\varphi$. A score of 1 conversely denotes a perfect EPS $\mathcal{P}$.

In the following, to assess the added value of the two-stage SCAMP model when compared to the benchmark AM_{25} analog model, we additionally estimate the gain in prediction skill as: $\Delta S = S_{\text{SCAMP}} - S_{\text{AM}_{25}}$ where S corresponds either to the BSS or the CRPSS.

4 Results

The two-stage model is used for the probabilistic prediction of small-scale precipitation over the continental French territory for each day of the 1982-2001 period. We here present the prediction skill obtained for occurrence and amount with the two predictors sets presented in section 2. As discussed later in section 5, the four predictors of each set are not necessarily all used; the predictors which have some predictive power for the considered predictand vary from one day to the other.

4.1 Performance of SCAMP

Fig. 3a presents the BSS skill score of SCAMP for precipitation occurrence prediction. The highest BSS values – up to 0.5 – are found in the western part of the Massif Central, in the Alps and along the Atlantic coast. Lower skill (BSS from 0.45 to 0.5) is obtained in northern and western lowlands. The lowest skill (0.35) is obtained for few cells located along the Mediterranean coast.

The BSS gain obtained with SCAMP over AM_{25} is rather important (up to 0.1 BSS points) and presents a high space variability (Fig. 3b). The gain (between 0.05 and 0.1 BSS points) is high in the mountainous areas (Pyrenees, Massif Central, Alpes, Vosges). The highest gains are found along the Mediterranean coast and in the Southern Alps where the BSS of SCAMP was the lowest. This highlights the weakness of the AM_{25} in these regions – characterized by more frequent convective precipitation and thus a weaker link with large-scale atmospheric circulation – and the interest for thermodynamic and more local predictors. Conversely, lower gains are observed in the western part of France characterized by more frontal precipitation and thus a stronger link with large-scale circulation. Note also that the spatial distribution of ΔBSS is very close (even if it has higher values) to the one obtained by SCAMP with R_{700} as unique predictor (not shown here).
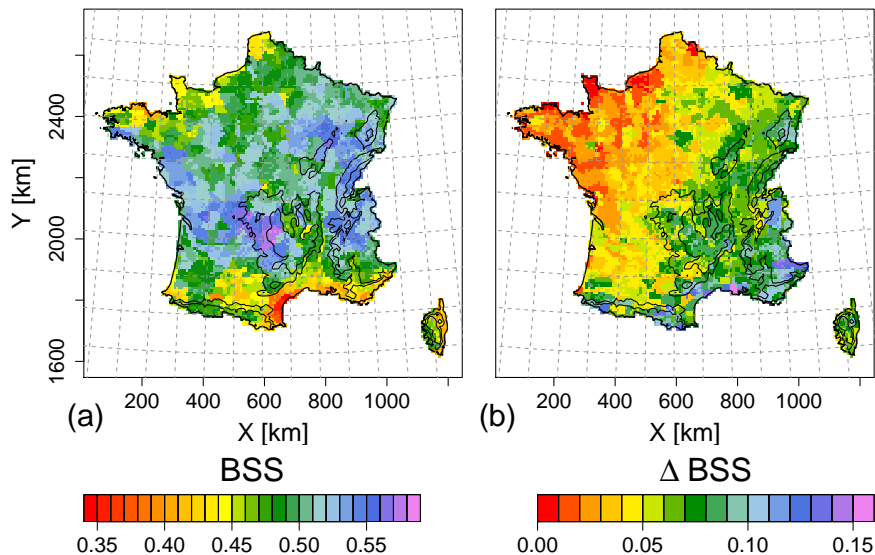
Figure 3. (a) BSS obtained with SCAMP (best possible value = 1). (b) BSS gain obtained with SCAMP compared to AM_{25}. Black solid lines correspond to the French borders and the contours around mountainous regions (400- and 800-m elevation) while the dashed lines show the ERA-40 grid mesh.

Fig. 4a shows the CRPSS obtained with SCAMP. The CRPSS values also depend on topography. The highest values, up to 0.45, are obtained in the western part of the Massif Central, the northern Alps, the Jura and the Vosges massifs. Lower values, between 0.32 and 0.45, are obtained in lowlands. The lowest skill (below 0.30) is again obtained along the Mediterranean coast.

The CRPSS gain obtained over AM_{25} is significant for most grid cells, with the highest value (up to 0.10 CRPSS points) obtained in the Rhone valley and in north-eastern France (Fig. 4b). Similarly to the BSS gain, a lower CRPSS gain is here also obtained in lowlands and western France. The spatial distribution of ΔCRPSS is here also very close (even if it has higher values) to the one obtained by SCAMP with W_{700} as unique predictor for amount (not shown here).

Despite the large dependency on regional features such as topography or proximity to the sea, adding local and thermodynamic information in SCAMP greatly improves the prediction skill over that of AM_{25}, for both precipitation occurrence and amount.

4.2 Characterisation of SCAMPs behaviour

As described in section 3.3, the regression stage of SCAMP is equivalent to update the empirical distribution obtained from the atmospheric analogs directly. For some prediction days, the regression stage can be however only partly activated, for either occurrence or amount. It can be even not activated at all. In these cases, the prediction is fully or partly obtained from the backup model AM_{25}.
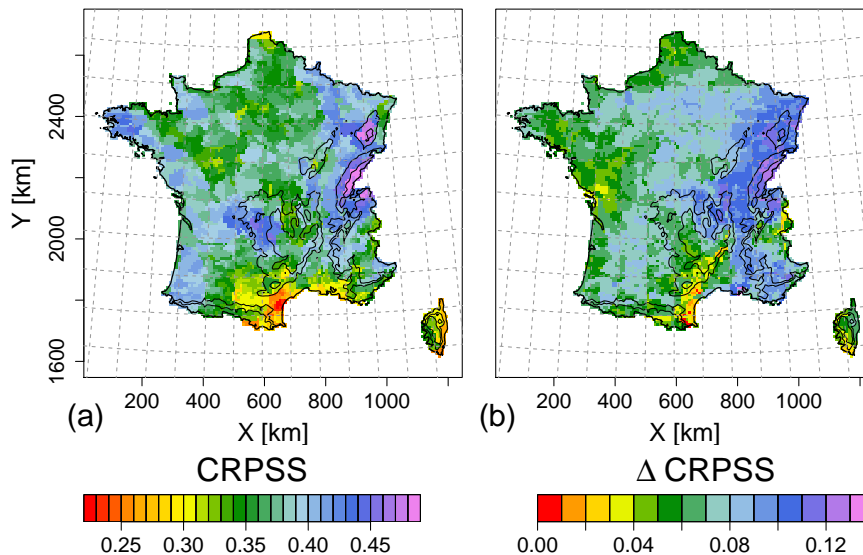
Figure 4. (a) CRPSS obtained with SCAMP (best possible value=1). (b) CRPSS gain obtained with SCAMP compared to AM$_{25}$.

The frequency each activation case (case 1 to 4) is obtained over the simulation period is given in Fig. 5. The situation where both precipitation occurrence and amount GLMs are activated (case 4) is very frequently observed. It corresponds to more than 85% of the days except in south-eastern France where only 60% of the days are concerned. All in all, the regression stage of SCAMP is very often activated (more than 97% of the days) to predict the occurrence probability (case 2 + 4). In the failing full-updating cases, AM$_{25}$ is usually applied to backup the precipitation amount prediction (case 1+2). Case 1, where the whole prediction is backup with AM$_{25}$ is finally very rare. For a large majority of the grid cells, it occurs less than 35 times in the 20-year period considered (corresponding to around 5‰).

Fig. 6 presents the mean precipitation anomaly for each of the previous cases, i.e. the ratio between the mean amount obtained for all days belonging to the considered case and the overall mean precipitation amount. An anomaly greater (resp. lower) than 1 indicates days that are rainier (resp. drier) than usual. The cases 1 to 4 correspond clearly to different precipitation configurations. The mean precipitation amount of days in case 4 is close to the overall mean. Days in cases 1 and 2 are very dry. Days in case 3 are very wet with a mean precipitation three times larger than the overall mean.

For a given prediction day, the precipitation state of its analog days is actually expected to be roughly similar to that of the day. This thus explains the SCAMP's behavior described above. In case 1 and 2, analog days of the prediction day are likely very dry. The number of humid analog days is thus likely small to very small, and likely too small to allow for a robust estimation of the precipitation amount GLM. Analog days are conversely likely humid in case 4 or even very humid in case 3. The number of humid days in those cases is thus likely large enough to allow for a robust estimation of the precipitation amount GLM. The very humid configuration of case 3 suggests that predictions days are characterized by a very large number
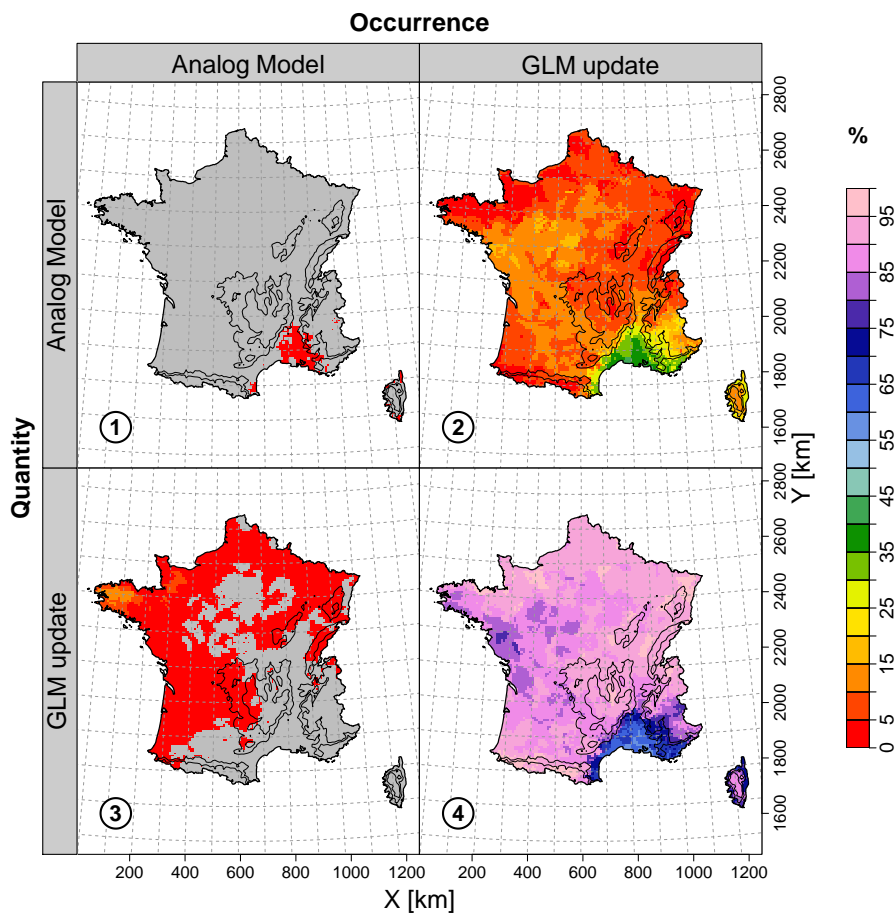
Figure 5. Percentage of days where 1) no updates are applied, 2) only the precipitation occurrence is updated, 3) only the precipitation amount is updated and 4) the occurrence and the precipitation amount are updated. Grids with gray colors correspond to grid cells where the corresponding case has been met less than 35 times over the 20-year evaluation period.

of humid analog days, which can in turn prevent for a robust estimation of the occurrence GLM (e.g. the occurrence GLM cannot be estimated in configurations where all days are wet).

This can also explain the specific results obtained in Southeast. Case 2 is indeed activated much more often in this region (increase of 30% percentage points) than elsewhere and, in a symmetric way, Case 4 is activated much less often in this region

5 (decrease of 30% percentage points). The reason underlying this result is to be related to the much higher proportion of dry days in Southeast (see Fig.S1 in Supplementary Material). In this region, the number of wet analog days is thus likely small for a large number of predictions days. As suggested above, this is obviously not a difficulty for the estimation of the occurrence GLM. This is conversely likely one for the estimation of the amount GLM. A small number of wet analogs likely prevents for a robust estimation of the precipitation amount GLM. This likely explains the much lower (resp. higher) frequency of Case 4

10 (resp. Case 2) in Southeast.
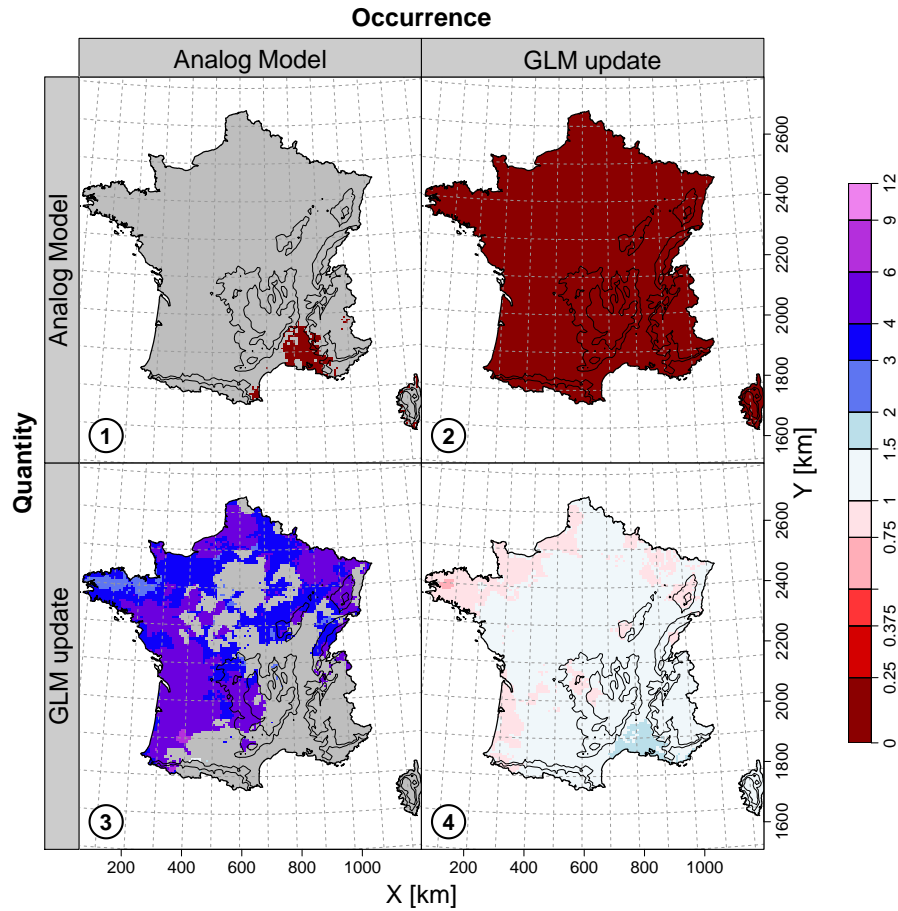
Figure 6. Ratio between the mean amount obtained for all days belonging to a given case and the overall mean precipitation amount. The 4 cases and gray grids: same as in Fig. 5.

The CRPSS gain achieved with SCAMP's results from the updated prediction of both precipitation occurrence and amount. To assess the relative effects of these updates on the gain, we further compared the four following prediction experiments:

Exp. 1: The prediction of both the occurrence and the amount is achieved with AM_{25} for all prediction days. This corresponds to the results given by Chardon et al. (2014, cf. Fig. 3).

5 **Exp. 2:** When possible, the precipitation occurrence probability is updated with the occurrence GLM. The non-zero precipitation amount is always predicted with AM_{25}.

Exp. 3: When possible, the precipitation amount is updated with the amount GLM. The precipitation occurrence probability is always predicted with AM_{25}.
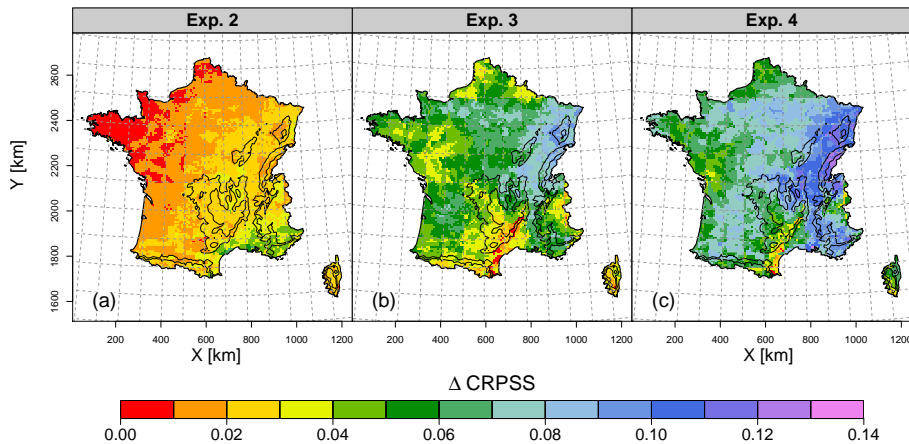
Figure 7. Gain in CRPSS for different prediction experiments (see section 4.2 for details) compared to the performance of AM_{25}. (a) Exp.2: only the precipitation occurrence probability is updated (when possible), (b) Exp.3: only the precipitation amount is updated (when possible), (c) Exp.4: both occurrence probability and precipitation amount are updated (when possible).

Exp. 4: When possible, both precipitation occurrence probability and amount are updated with the occurrence and amount GLMs. This corresponds to the two-stage configuration already evaluated previously.

The CRPSS gain obtained between Exp.1 and 2, between Exp. 1 and 3 and between Exp.1 and 4 are presented in Fig. 7 (the results for Exp.4, already presented in Fig. 4, are presented again for the ease of comparison).

5   For a large majority of grid cells, the CRPSS gain obtained with an updated prediction of the occurrence probability (from 0 to 0.05 CRPSS points) is significantly lower than that obtained with an updated prediction of amount (from 0.03 to 0.1 CRPSS points). The CRPSS gain obtained in the latter case is additionally close to that obtained with the full two-stage model. The CRPSS gain obtained by SCAMP in Fig. 7c is thus explained in most cases by the updated prediction of precipitation amount. The scheme is somehow different in the south of France along the Mediterranean coast and in the Cevennes-Vivarais moun-

10   tains. In those regions, the CRPSS gain obtained by SCAMP is mostly explained by the updated prediction of the occurrence probability. Updating only the precipitation amount leads to fairly no CRPSS gain.

5   Discussion

The sets of potential predictors used in SCAMP for the prediction of precipitation occurrence and amount have been listed in section 2. For each variable, the number of potential predictors is here equal to four. All four predictors are not necessary

15   retained for the GLM. For a given prediction day, a GLM with a single predictor or a combination of several predictors among the four can be selected. Fifteen regressive structures plus the backup AM_{25} model are possible in our context (Table 2).

For a given prediction day, the regressive structure selected by SCAMP for precipitation occurrence or for precipitation amount are supposed to include the best information for the prediction. In the following, we assess how often each structure

has been selected. This allows for some insight in the atmospheric information really used for the regression stage and how this information varies in time.

Fig. 8 and Fig. 9 present the percentage of times that the 15 regressive structures and the backup AM_{25} are used for the prediction of precipitation occurrence and amount, respectively. As in Fig. 5, gray cells indicate that the regression structure has been retained less than 35 times over the 20-year evaluation period. For both occurrence and amount, the selection frequency of the structures is also rather region dependent and strongly influenced by topography.

For occurrence (Fig. 8), the most often selected structure is Str. n°1, which is only based on R_{700} (more than 25% of the time for the whole of France). R_{700} was actually found to give the highest predictive power when used in a single predictor configuration. Another structure which is also often selected (more than 15% for a high number of grids) is Str. n°7 which combines R_{700} with $Occ - 1$. Secondary structures – as for example Str. n°6 and n°13 combining R_{700} to W_{700} and $Occ - 1$ – can be selected more than 10% of the days for some given regions. Other structures are seldom selected and some of them (Str. n°8, n°11, n°14 and n°15) are almost never selected.

For precipitation amount, the most often selected structures are Str. n°3 and Str. n°1, both based on one single predictor, W_{700} or R_{700} respectively. These structures are selected more than 25% of the time (resp. more than 15%) for a high number of grids. The secondary structures (Str. n°6, n°8 and n°13) are selected from 5% up to 20% depending on the region. They always include W_{700} in combination with some other predictor (R_{700}, H or $R_{700} + T_{700}$). Str. n°9, n°12, n°14 and n°15 are almost never selected. The others are selected less than 10 to 5% of the time.

Note that for the selection of the best regression structure for a given prediction day, all these 15 regressive structures have been in turn tested. The results above suggest that this systematic test is not necessary and that it could be reasonable to consider only the few structures which are frequently retained or which are retained a "reasonable" fraction of the days. However, the selection frequency of a given structure actually varies with the seasons and/or the encountered synoptic situation and some secondary regressive structures can be retained frequently for specific situations. This is illustrated in Fig. 10 for a cell located in north-western France. The figure presents how the selection frequency of each regression structure differs in different seasons and weather patterns (WP, defined in Table 3, Garavaglia et al., 2010) from the selection frequency obtained for the all-days situation.

For precipitation occurrence (Fig. 10a), the selection of the main regressive structures (i.e. Str. n°1 and n°7 respectively based on R_{700} and $R_{700} + Occ - 1$) is up to 15% more frequent (resp. less frequent) for WP3 (resp. WP5) compared to the all-days situation. For precipitation amount (Fig. 10b), the selection frequency of the main regressive structures (Str. n°1 and n°3 based on R_{700} and W_{700} respectively) can similarly change up to +/-10%. The reduced selection of a main regressive structure for a given season or WP can lead to preferentially retain some secondary regressive structure. For instance, the regressive Str. n°8 based on $W_{700} + H$ is selected 10% more frequently for WP2 than for the all-days situation (Fig. 10b).

The preferential (or conversely reduced) selection of some regression structures for given WTs or seasons was estimated for all grid cells of France. In most cases, the preferential (or reduced) selection was found to present a noticeable spatial coherency. Different configurations are observed as illustrated in Fig. 11 and discussed below.
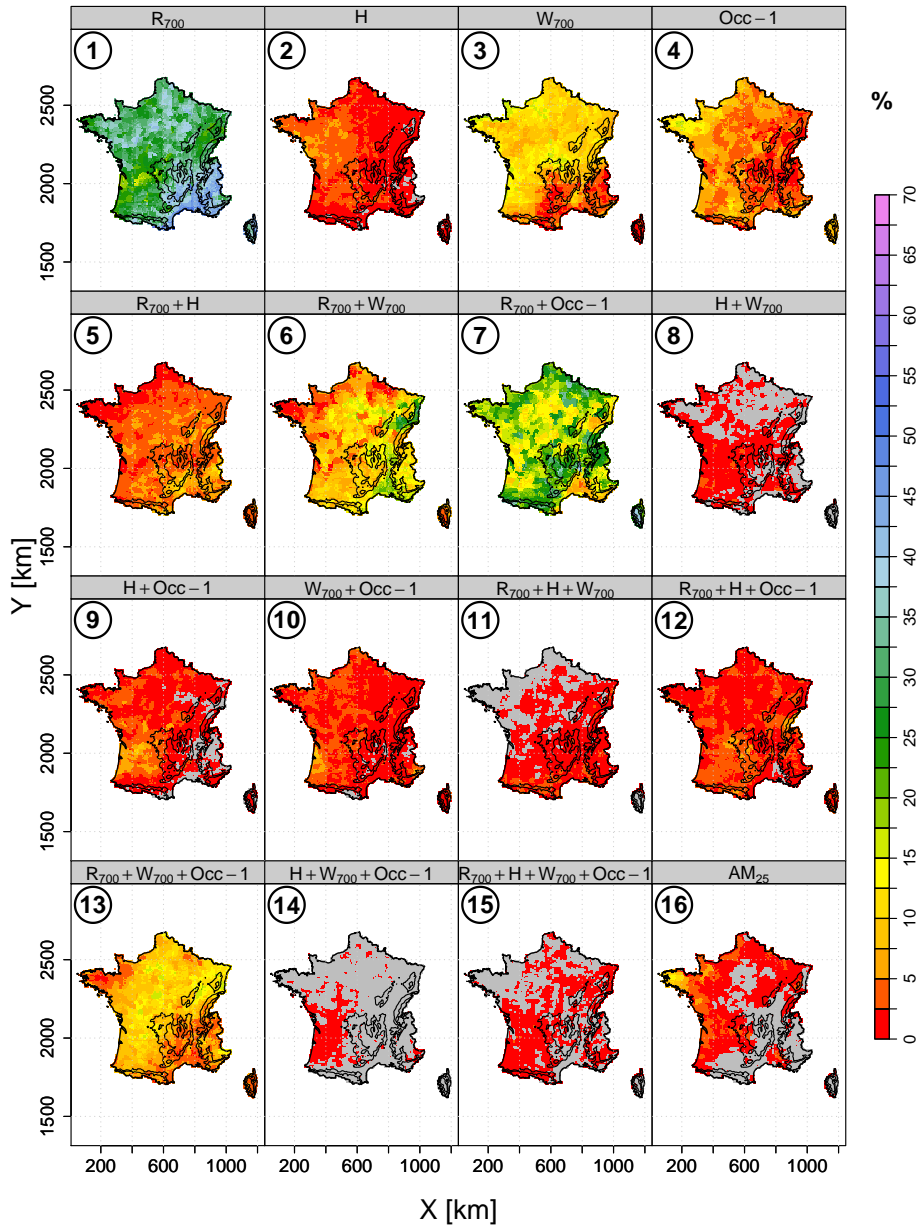
Figure 8. Prediction of occurrence probability: selection frequencies (%) of the 15 regression structures and of the backup model AM_{25}. Predictors involved are indicated in graphs headers and index of the regressive structure in top left corners. Gray grids: same as in Fig. 5. The selection frequency of AM_{25} corresponds to the sum of those obtained for cases 1 and 3 in Fig. 5.

The preferential selection of some regression structures can first be observed over large to very large regions. As an example, the preferential selection of Str. n°3 for the prediction of precipitation amount for days in WP7 (more than +15% compared
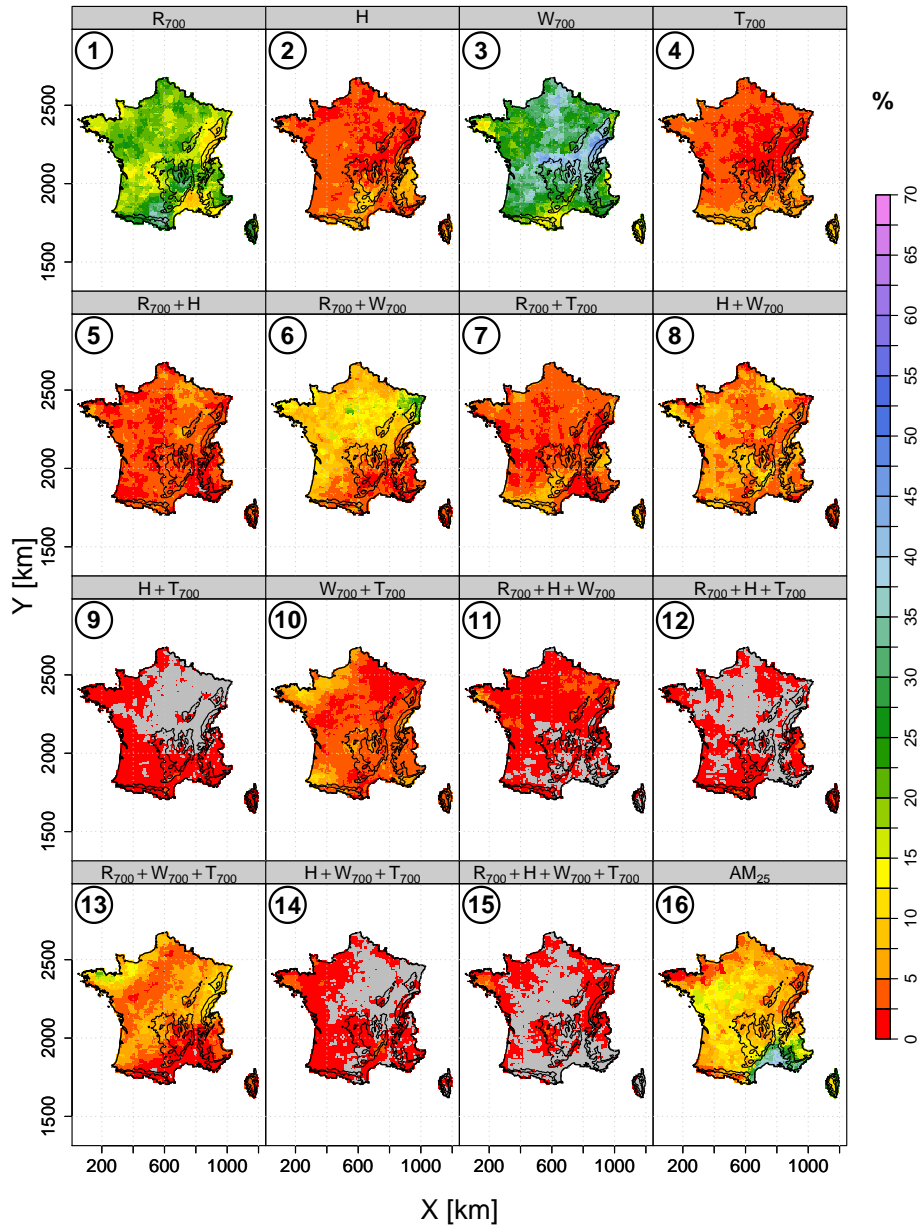
Figure 9. Same as Fig. 8 for the probabilistic prediction of precipitation amount. The selection frequency of AM_{25} corresponds to the sum of those obtained for cases 1 and 2 in Fig. 5.

to usual) is obtained for all grid cells in France. Whatever the location, the vertical velocity W_{700} seems thus required in this specific weather pattern. Another example is that of WP8 which corresponds to an Anticyclonic situation. Whatever the
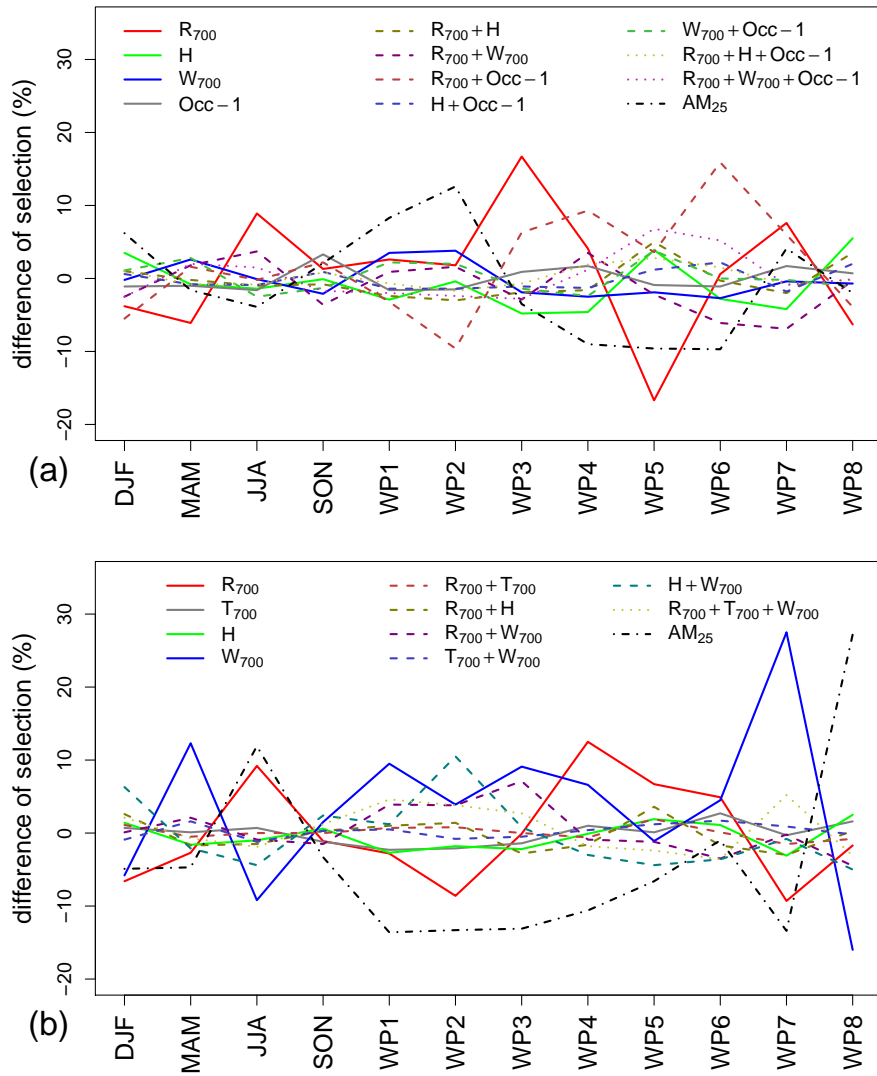
Figure 10. For each season and weather type, difference (%) in selection frequency with the all-days case for different regression structures. Results for the prediction of (a) occurrence and (b) amount. A positive difference indicates that the considered regressive structure is selected more often than for the all-day situation. Results are displayed for a grid cell located in the north-west of France. For a clearer illustration, the three or four regressive structures that are almost never selected are not displayed.

location, no precipitation is really expected for this configuration. No predictor is thus required in addition to geopotential heights used in the analog stage. This configuration logically leads to a large preferential selection of the backup AM_{25} model.

For a given weather pattern, the preferential selection of a regressive structure can also vary from one region to the other. For WP2 for instance, the structures based on W_{700} or on W_{700} and H are selected much more often along the Atlantic coast

and in the north of France. The backup AM_{25} model is conversely more selected in the South-East, in the Mediterranean coast especially. For this weather regime, the South-east is actually protected by the Massif Central mountain and does thus usually not receive precipitation (cf. Fig. 3 of Garavaglia et al., 2010).

The preferential selection of a regressive structure can be also obtained for rather small and specific regions. In Fig. 11b, the regressive Str. n°8 based on $W_{700} + H$ is more frequently selected for WP7 (around +15%) in the Cevennes-Vivarais regions (south-eastern part of the Massif Central) and in the pre-alpine mountains (western part of the Alps). The combination of W_{700} and H seems thus to be very informative in those configurations for this really rare WP (4% of the 20-year period).

Whatever the configuration, the preferential selection of regression structures presents some spatial coherency, at small or large regional scales. This obviously also suggests the spatial robustness of the informative predictors to be retained for given large scale weather configurations.

6   Conclusions

The relevance of a two-stage analog/regression model has been explored in this study for the probabilistic prediction of precipitation over France. Atmospheric analogs of the prediction day are identified to estimate the parameters of a two-part regression model further applied for the prediction. The regression model consists of a logistic GLM for the prediction of precipitation occurrence and a logarithmic GLM for the prediction of precipitation amount. The prediction obtained with this two-stage approach updates the predictive distribution that would have been achieved directly from a one stage analog model based on atmospheric circulation analogs. The two-stage approach makes the downscaling model adaptive: as the analog days are identified for each prediction day, the predictors and regression coefficients of the regression models can vary from one day to the other.

The regression stage allows a non-negligible prediction skill gain compared to the reference analog model (gain up to 0.1 skill score points for both the BSS and the CRPSS). The CRPSS gain is mainly achieved due to the regression model estimated for the precipitation amount. The introduction of local scale predictors such as relative humidity is obviously crucial there. The adaptive nature of the model and thus the possibility to tailor the downscaling relationship (both predictors and regression coefficients) to the current prediction day seems to be decisive as well. The CRPSS gain obtained with the two-stage approach is actually two times larger than the one obtained by Chardon et al. (2014) with a two level analog model where a unique and same second level analogy variable (namely humidity) is considered for all days.

The prediction skill and adaptability of this two-stage approach was illustrated for the prediction of both the precipitation occurrence and amount in a simplified configuration where four predictors, selected in a preliminary analysis from a large ensemble of potential predictors, are used in the regression stage. The predictors used for precipitation occurrence are the relative humidity and vertical velocity at 700 hPa, the helicity integrated from 1000 hPa to 500 hPa and the occurrence of the previous day. A similar set of predictors is used for the precipitation amount (the occurrence of the previous day is replaced by the 700 hPa temperature). Most of the time, the final regression model only includes one or two predictors. It also very often includes the relative humidity R_{700} for precipitation occurrence and R_{700} or the vertical velocity W_{700} for precipitation
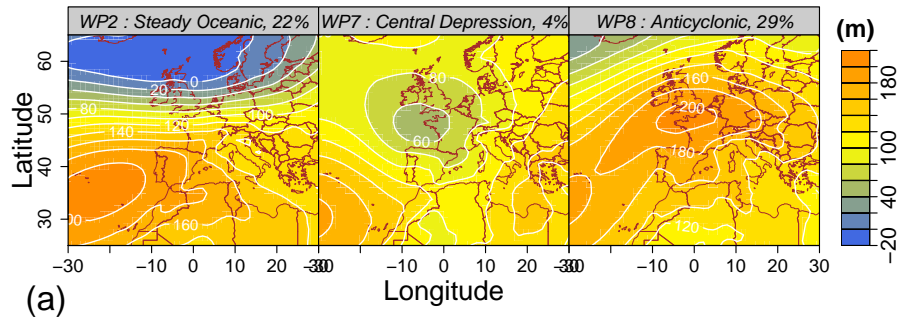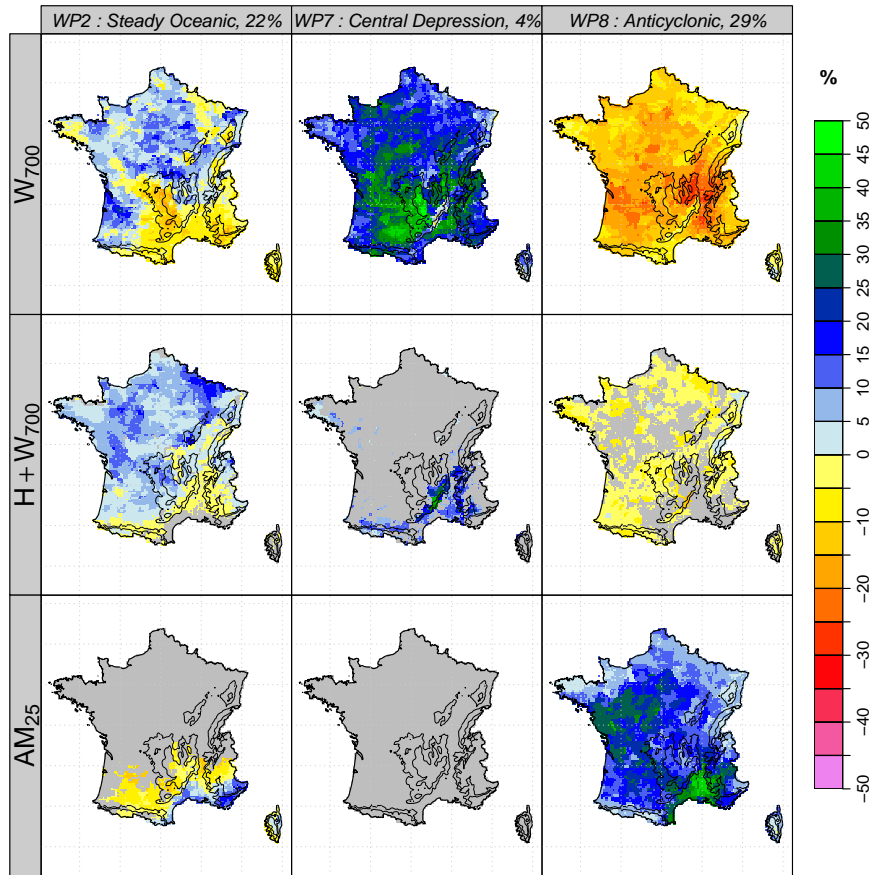
Figure 11. (a) Mean geopotential height at 1000 hPa for three WPs (Garavaglia et al., 2010). (b) For each WP, difference (%) in selection frequency with the all-days case. Results for two regression structures (W_{700} and $W_{700} + H$) and for AM$_{25}$. Predictand is precipitation amount. A positive (resp. negative) difference indicates an extra-selection (resp. reduced selection). Gray grids: same as in Fig. 5.

amount. Some combinations of predictors, almost never used in general, appear to be more frequently retained for some specific weather patterns and/or locations in France, revealing their potential interest for these situations.

For the sake of simplicity and to limit the degrees of freedom in our analysis, we considered a unique set of four potential predictors for all SAFRAN grid cells. This obviously leads to a sub-optimal prediction configuration. The main meteorological processes driving precipitation in France obviously differ from one region to the other. The most informative predictors are thus expected to be region-dependent and the set of predictors to be considered in the regression stage could be refined on a regional basis. This is expected to improve the skill of the prediction. The same would apply for an application of SCAMP to other regions worldwide.

A number of atmospheric variables have been considered as potential predictors in similar downscaling studies. The predictors found to be of interest are most often few. They are roughly the same than those considered in the preliminary analysis of the present work. However, as in the present work, the analyses usually carried out to identify these informative predictors are potentially misleading. The selection of a variable is indeed often based on its predictive power, estimated with some prediction skill score in an all-days evaluation framework. As highlighted in the present work however, some predictors are likely to be informative for very few meteorological situations. An all-days evaluation is expected to reveal robust predictors. It however very likely misses important situation-specific predictors. The two-stage approach here estimates the statistical downscaling link from a homogeneous set of days, with respect to their large-scale atmospheric circulation configuration. Those days are moreover atmospheric analogs to the prediction day. This two-stage approach has thus the potential to reveal the predictive power of very specific predictors, suited for very specific meteorological configurations. It leaves very likely room for significant improvements of the prediction skill for such unusual configurations. It gives likely also the opportunity to better understand the atmospheric factors under play in a number of non-frequent and atypical meteorological situations. Notwithstanding the technical limitations that may hamper such analyses, a broader exploration of a much larger diversity of predictors, possibly non-conventional ones, would be thus definitively worth in this context.

Both the predictors and the regression coefficients were shown in our work to depend on the analog days identified in the analog stage. This is the reason of the adaptability of the downscaling discussed above. Besides the adaptability, we ideally expect that for a given prediction day the predictor selection and the associated regression coefficients are robust. Further analyses should explore this issue. An interesting work would be for instance to check that the predictors and their related coefficients do not significantly change when the set of analog days considered for the estimation is modified as a result of a different setup of the Analog Model (e.g. when one changes the archive period or the archive length).

Results of our work depend on a number of choices and assumptions. They for instance likely depend on the database used for the large scale atmospheric predictors. The day-to-day behavior of such an analog/regression approach (and the skill of the prediction) likely depends on the database and especially on the quality of the predictors. An atmospheric reanalysis with a higher spatial resolution would for instance likely allow for a better description of the shapes of geopotential fields and for a more relevant simulation of regional / local thermodynamic processes. It would likely lead in turn to have higher quality variables for some atmospheric parameters such as air instability. This may allow for a better identification of the daily specificity in the downscaling relationship and of the most informative predictors to be used each day. The reverse may occur

when using lower quality predictors, for instance lower quality data from reanalyses available for the 20th century or lower quality data from climate or numerical weather forecasting models. The quality of the predictors is thus obviously also an important issue to be further considered. It may lead to reveal different informative predictors depending on the intended use of the model (forecast, simulation or climate impact studies).

5   SCAMP was here used for the prediction of small scale precipitation at individual grid cells. The prediction of precipitation fields, obviously required for a number of impact studies, is also a challenging issue (e.g. Clark et al., 2004; Yang et al., 2005; Thober et al., 2014). Different adaptations of SCAMP would be worth investigating in this context. SCAMP could be for instance applied for the prediction of mean areal precipitation over the whole targeted spatial domain and some spatial disaggregation process could be further used to generate the required fields (e.g. Mezghani and Hingray, 2009; Rupp et al.,

10   2012). As highlighted by Chardon et al. (2016), the prediction skill of SCAMP is expected to increase with the size of the spatial domain targeted for the prediction, which makes such an approach rather appealing. Another possible strategy for spatial predictions would be to rely on the advantages introduced by the analog stage of SCAMP. Chardon et al. (2014) indeed showed that for a given prediction day a same set of analog days can be used over rather large domains (up to a few 100's of kilometers) for a quasi-optimal prediction of local scale precipitation. The precipitation field of each analog day (which is thus

15   spatially coherent because already observed) could thus be used as a first guess precipitation field scenario for the considered region. The field could be next updated at each location with day- and location specific coefficients obtained from the regression stage of SCAMP. This spatial prediction issue will be considered in future works.

Author contributions.   This study is part of J. Chardon's PhD thesis. B. Hingray and A-C. Favre supervised the PhD. All authors contributed to the designed experiments to the writing of the document. J. Chardon developed the model code and performed the simulations.

20   *Competing interests.*   The authors declare that they have no conflict of interest.

Acknowledgements.   The authors especially thank Charles Obled and Isabella Zin for fruitful discussions on the analog method. The authors thank also the Grenoble University High Performance Computing centre CIMENT (https://ciment.ujf-grenoble.fr/wiki-pub/index.php/Welcome_to_the_CIMENT_site!) for their help and the large computing resource they provide. We would especially like to thank the two anonymous reviewers for their relevant comments and suggestions that allowed us to improve this study and broaden its scope.

25   .

Appendix A: Acronyms

AM	Analog Model
AM_{25}	Analog Model (based on the 25 nearest atmospheric analogs) used as benchmark or backup prediction model.
BS	Brier Score
BSS	Brier Skill Score
cdf	cumulative distribution function
CRPS	Continuous Ranked Probability Score
CRPSS	Continuous Ranked Probability Skill Score
EPS	Ensemble Prediction System
GLM	Generalized Linear Model
SAFRAN	8x8km precipitation reanalysis for France from MeteoFrance
SCAMP	two-stage analog / regression model
SDM	Statistical Downscaling Model
TWS	Teweless-Wobus Score
WP	Weather Pattern

References

Akaike, H.: A new look at the statistical model identification, IEEE T. Automat. Contr., 19, 716–723, doi:10.1109/TAC.1974.1100705, 1974.

Asong, Z. E., Khaliq, M. N., and Wheater, H. S.: Projected changes in precipitation and temperature over the Canadian Prairie Provinces using the Generalized Linear Model statistical downscaling approach, J. Hydrol., 539, 429–446, doi:10.1016/j.jhydrol.2016.05.044, 2016.

Ben Daoud, A., Sauquet, E., Bontron, G., Obled, C., and Lang, M.: Daily quantitative precipitation forecasts based on the analogue method: Improvements and application to a French large river basin, Atmos. Res., 169, Part A, 147 – 159, doi:http://dx.doi.org/10.1016/j.atmosres.2015.09.015, 2016.

Boé, J., Terray, L., Habets, F., and Martin, E.: Statistical and dynamical downscaling of the Seine basin climate for hydro-meteorological studies, Int. J. Climatol., 27, 1643–1655, doi:10.1002/joc.1602, 2007.

Bontron, G. and Obled, C.: A probabilistic adaptation of meteorological model outputs to hydrological forecasting, Houille Blanche, pp. 23–28, 2005.

Brier, G. W.: Verification of forecasts expressed in terms of probability, Mon. Wea. Rev., 78, doi:10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2, 1950.

Brown, T. A.: Admissible scoring systems for continuous distributions, Manuscript P-5235, The Rand Corporation, The Rand Corporation, Santa Monica, CA, available from The Rand Corporation, 1700 Main St., Santa Monica, CA 90407-2138, 1974.

Caillouet, L., Vidal, J.-P., Sauquet, E., and Graff, B.: Probabilistic precipitation and temperature downscaling of the Twentieth Century Reanalysis over France, Climate of the Past, 12, 635–662, doi:10.5194/cp-12-635-2016, 2016.

Campozano, L., Tenelanda, D., Sanchez, E., Samaniego, E., and Feyen, J.: Comparison of Statistical Downscaling Methods for Monthly Total Precipitation: Case Study for the Paute River Basin in Southern Ecuador, Adv. Meteor., doi:10.1155/2016/6526341, 2016.

Chandler, R. E. and Wheater, H. S.: Analysis of rainfall variability using generalized linear models: A case study from the west of Ireland, Wat. Resour. Res., 38, 10–1–10–11, doi:10.1029/2001WR000906, 2002.

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

Chardon, J., Favre, A.-C., and Hingray, B.: Effects of Spatial Aggregation on the Accuracy of Statistically Downscaled Precipitation Predictions, J. Hydromet., 17, 1561–1578, doi:10.1175/JHM-D-15-0031.1, 2016.

Clark, M., Gangopadhyay, S., Hay, L., Rajagopalan, B., and Wilby, R.: The Schaake shuffle: A method for reconstructing space-time variability in forecasted precipitation and temperature fields, J. Hydromet., 5, 243–262, doi:10.1175/1525-7541(2004)005<0243:TSSAMF>2.0.CO;2, 2004.

Coe, R. and Stern, . D.: Fitting Models to Daily Rainfall Data, J. Appl. Meteor., 21, 1024–1031, doi:10.1175/1520-0450(1982)021<1024:FMTDRD>2.0.CO;2, 1982.

Dayon, G., Boe, J., and Martin, E.: Transferability in the future climate of a statistical downscaling method for precipitation in France, J. Geophys. Res. - Atm., 120, 1023–1043, doi:10.1002/2014JD022236, 2015.

Enke, W. and Spegat, A.: Downscaling climate model outputs into local and regional weather elements by classification and regression, Climate Res., 08, 195–207, doi:10.3354/cr008195, 1997.

Gaitan, C. F., Hsieh, W. W., and Cannon, A. J.: Comparison of statistically downscaled precipitation in terms of future climate indices and daily variability for southern Ontario and Quebec, Canada, Climate Dyn., 43, 3201–3217, doi:10.1007/s00382-014-2098-4, 2014.

Gangopadhyay, S., Clark, M., and Rajagopalan, B.: Statistical downscaling using K-nearest neighbors, Wat. Resour. Res., 41, 2005.

Garavaglia, F., Gailhard, J., Paquet, E., Lang, M., Garçon, R., and Bernardara, P.: Introducing a rainfall compound distribution model based on weather patterns sub-sampling, Hydrol. Earth Syst. Sci., 14, 951–964, doi:10.5194/hess-14-951-2010, 2010.

Guilbaud, S. and Obled, C.: L'approche par analogues en prévision météorologique., La Météorologie, 8, 21–35, 1998.

Hanssen-Bauer, I., Achberger, C., Benestad, R., Chen, D., and Forland, E.: Statistical downscaling of climate scenarios over Scandinavia, Climate Res., 29, 255–268, doi:10.3354/cr029255, 2005.

Holton, J. R. and Hakim, G. J.: An introduction to dynamic meteorology, vol. 88, Academic press, 2012.

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

Ibarra-Berastegi, G., Saenz, J., Ezcurra, A., Elias, A., Diaz Argandona, J., and Errasti, I.: Downscaling of surface moisture flux and precipitation in the Ebro Valley (Spain) using analogues and analogues followed by random forests and multiple linear regression, Hydrol. Earth Syst. Sci., 15, 1895–1907, doi:10.5194/hess-15-1895-2011, 2011.

Kuentz, A., Mathevet, T., Gailhard, J., and Hingray, B.: Building long-term and high spatio-temporal resolution precipitation and air temperature reanalyses by mixing local observations and global atmospheric reanalyses: the ANATEM model, Hydrol. Earth Syst. Sci., 19, 2717–2736, doi:10.5194/hess-19-2717-2015, 2015.

Lafaysse, M., Hingray, B., Mezghani, A., Gailhard, J., and Terray, L.: Internal variability and model uncertainty components in a multireplicate multimodel ensemble of hydrometeorological projections : the Alpine Durance basin, Wat. Resour. Res., 50, 3317–3341, 2014.

Maraun, D. and al.: Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user, Rev. Geophys., 48,RG3003, doi:201010.1029/2009RG000314, 2010.

Marty, R., Zin, I., Obled, C., Bontron, G., and Djerboua, A.: Toward Real-Time Daily PQPF by an Analog Sorting Approach: Application to Flash-Flood Catchments, J. Appl. Meteorol. Clim., 51, 505–520, doi:10.1175/JAMC-D-11-011.1, 2012.

Matheson, J. E. and Winkler, R. L.: Scoring Rules for Continuous Probability Distributions, Management Science, 22, 1087 –1096, doi:10.1287/mnsc.22.10.1087, http://mansci.journal.informs.org/content/22/10/1087.abstract, 1976.

McCullagh, P. and Nelder, J. A.: Generalized Linear Models, Second Edition, Chapman and Hall/CRC, Boca Raton, 2 edition edn., 1989.

Mezghani, A. and Hingray, B.: A combined downscaling-disaggregation weather generator for stochastic generation of multisite hourly weather variables over complex terrain: development and multi-scale validation for the Upper Rhone River basin, J. Hydrol., 377, 245–260, 2009.

Murphy, A. H.: A new vector partition of the probability score, J. Appl. Meteor. Climatol., 12, 595–600, doi:10.1175/1520-0450(1973)012<0595:ANVPOT>2.0.CO;2, 1973.

Nasseri, M., Tavakol-Davani, H., and Zahraie, B.: Performance assessment of different data mining methods in statistical downscaling of daily precipitation, J. Hydrol., 492, 1–14, doi:10.1016/j.jhydrol.2013.04.017, 2013.

Nelder, J. A. and Wedderburn, R. W. M.: Generalized Linear Models, J. Roy. Stat. Soc. A Sta., 135, 370–384, 1972.

Obled, C., Bontron, G., and Garçon, R.: Quantitative precipitation forecasts: a statistical adaptation of model outputs through an analogues sorting approach, Atmos. Res., 63, 303–324, doi:10.1016/S0169-8095(02)00038-8, 2002.

Quintana-Segui, P. and al.: Analysis of near-surface atmospheric variables: Validation of the SAFRAN analysis over France, J. Appl. Meteorol. Clim., 47, 92–107, doi:10.1175/2007JAMC1636.1, 2008.

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

Raynaud, D., Hingray, B., Zin, I., Anquetin, S., Debionne, S., and Vautard, R.: Atmospheric analogs for physically consistent scenarii of surface weather in Europe and Maghreb, Int. J. Climatology, doi:doi :10.1002/joc.4844, 2016.

Ribalaygua, J., Torres, L., Pórtoles, J., Monjo, R., Gaitán, E., and Pino, M. R.: Description and validation of a two-step analogue/regression downscaling method, Theor. Appl. Clim., 114, 253–269, doi:10.1007/s00704-013-0836-x, 2013.

5 Rupp, D. E., Licznar, P., Adamowski, W., and Lesniewski, M.: Multiplicative cascade models for fine spatial downscaling of rainfall: parameterization with rain gauge data, Hydrol. Earth Syst. Sci., 16, 671–684, doi:10.5194/hess-16-671-2012, 2012.

Schwarz, G.: Estimating the Dimension of a Model, Ann. Stat., 6, 461–464, doi:10.1214/aos/1176344136, 1978.

Stern, R. D. and Coe, R.: A Model Fitting Analysis of Daily Rainfall Data, J. Roy. Stat. Soc. A Sta., 147, 1, doi:10.2307/2981736, 1984.

Teweless, J. and Wobus, H.: Verification of prognosis charts, Bull. Amer. Meteor. Soc., 35, 2599–2617, 1954.

10 Thober, S., Mai, J., Zink, M., and Samaniego, L.: Stochastic temporal disaggregation of monthly precipitation for regional gridded data sets, Wat. Resour. Res., 50, 8714–8735, doi:10.1002/2014WR015930, 2014.

Uppala, S. M. and al.: The ERA-40 re-analysis, Q. J. Roy. Meteor. Soc., 131, 2961–3012, doi:10.1256/qj.04.176, 2005.

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

15 Wetterhall, F., Bardossy, A., Chen, D., Halldin, S., and Xu, C.-Y.: Statistical downscaling of daily precipitation over Sweden using GCM output, Theor. Appl. Clim., 96, 95–103, doi:10.1007/s00704-008-0038-0, 2009.

Wilby, R. L., Hay, L. E., and Leavesley, G. H.: A comparison of downscaled and raw GCM output: implications for climate change scenarios in the San Juan River basin, Colorado, J. Hydrol., 225, 67–91, doi:10.1016/S0022-1694(99)00136-5, 1999.

Woodcock, F.: On the Use of Analogues to Improve Regression Forecasts, Mon Weather Rev., 108, 292–297, 1980.

20 Yang, C., Chandler, R., Isham, V., and Wheater, H.: Spatial-temporal rainfall simulation using generalised linear models, Wat. Resour. Res., 41, doi:10.1029/2004WR0037395, 2005.

Table 1. Large-scale potential variables considered in the work. Stars: predictors obtained from the best GLMs identified for the 12 test SAFRAN grid cells (section 2). Double stars: predictors used for the analog stage. Bold text: predictors retained for the SCAMP version presented and evaluated in this work. See Holton and Hakim (2012) for the definition of the variables.

Acronym	Predictor description
R_{850}	*Relative humidity at 850 hPa
R_{700}	***Relative humidity at 700 hPa**
R_{500}	Relative humidity at 500 hPa
TCW	Total Column Water
$R_{850}TCW$	*Product of R_{850} and TCW
T_{700}	***Air temperature at 700 hPa**
B_{700}	*Baroclinity at 700 hPa
Δ_z	700hPa - 1000hPa thickness of the air column
Z_{1000}	****Geopotential height at 1000 hPa**
Z_{700}	Geopotential height at 700 hPa
Z_{500}	****Geopotential height at 500 hPa**
F_{700}	Wind speed at 700 hPa
U_{700}	West component of wind speed at 700 hPa
V_{700}	South component of wind speed at 700 hPa
W_{700}	***Vertical velocity (vertical component of wind speed) at 700 hPa**
H	***Helicity of horizontal wind integrated from 1000 to 500 hPa**
PV_{400}	*Potential vorticity of the atmosphere at 400 hPa
$\Delta\theta$	*Potential temperature gradient between 925 and 700 hPa
FR_{700}	*Humidity flux at 700 hPa
FU_{700}	West component of humidity flux at 700 hPa
FV_{700}	South component of humidity flux at 700 hPa
∇FR_{700}	*Divergence of FR_{700}
$Occ-1$	***Precipitation occurrence of the day before the prediction day**

Table 2. Possible regressive structures (i.e. combination of predictors) for the modeling of precipitation occurrence and amount.

Structure index	Precipitation occurrence	Precipitation amount
Str. n°1	R_{700}	R_{700}
Str. n°2	H	H
Str. n°3	W_{700}	W_{700}
Str. n°4	$Occ\text{-}1$	T_{700}
Str. n°5	$R_{700} + H$	$R_{700} + H$
Str. n°6	$R_{700} + W_{700}$	$R_{700} + W_{700}$
Str. n°7	$R_{700} + Occ\text{-}1$	$R_{700} + T_{700}$
Str. n°8	$H + W_{700}$	$H + W_{700}$
Str. n°9	$H + Occ\text{-}1$	$H + T_{700}$
Str. n°10	$W_{700} + Occ\text{-}1$	$W_{700} + T_{700}$
Str. n°11	$R_{700} + H + W_{700}$	$R_{700} + H + W_{700}$
Str. n°12	$R_{700} + H + Occ\text{-}1$	$R_{700} + H + T_{700}$
Str. n°13	$R_{700} + W_{700} + Occ\text{-}1$	$R_{700} + W_{700} + T_{700}$
Str. n°14	$H + W_{700} + Occ\text{-}1$	$H + W_{700} + T_{700}$
Str. n°15	$R_{700} + H + W_{700} + Occ\text{-}1$	$R_{700} + H + W_{700} + T_{700}$

Table 3. Names of the weather patterns (WP) defined in Garavaglia et al. (2010) and related frequency for 1982-2001 period.

Index	Denomination	Annual frequency (%)
WP1	Atlantic Wave	8
WP2	Steady Oceanic	22
WP3	Southwest Circulation	8
WP4	South Circulation	17
WP5	Northeast Circulation	6
WP6	East Return	6
WP7	Central Depression	4
WP8	Anticyclonic	29