



HAL
open science

Explaining Rainfall Accumulations over Several Days in the French Alps Using Low-Dimensional Atmospheric Predictors Based on Analogy

Juliette Blanchet, Jean-Dominique Creutin

► **To cite this version:**

Juliette Blanchet, Jean-Dominique Creutin. Explaining Rainfall Accumulations over Several Days in the French Alps Using Low-Dimensional Atmospheric Predictors Based on Analogy. *Journal of Applied Meteorology and Climatology*, 2020, 59 (2), pp.237-250. 10.1175/JAMC-D-19-0112.1 . hal-03087661

HAL Id: hal-03087661

<https://hal.science/hal-03087661>

Submitted on 16 Sep 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Explaining Rainfall Accumulations over Several Days in the French Alps Using Low-Dimensional Atmospheric Predictors Based on Analogy

JULIETTE BLANCHET AND JEAN-DOMINIQUE CREUTIN

Université Grenoble Alpes, CNRS, IGE, Grenoble, France

(Manuscript received 14 May 2019, in final form 23 October 2019)

ABSTRACT

We propose a new approach to explain multiday rainfall accumulation over a French Alpine watershed using large-scale atmospheric predictors based on analogy. The classical analogy framework associates a rainfall cumulative distribution function (CDF) with a given atmospheric situation from the precipitation accumulations yielded by the closest situations. The analogy may apply to single-day or multiday sequences of pressure fields. The proposed approach represents a paradigm shift in analogy. It relies on the similarity of the local topology mapping the pressure field sequences, somehow forgetting the pressure fields per se. This topology is summarized by the way the sequences of pressure fields resemble their neighbors (dimensional predictors) and how fast they evolve in time (dynamical predictors). Although some information—and hence predictability—is expected to be lost when compared with classical analogy, this approach provides new insight on the atmospheric features generating rainfall CDFs. We apply both approaches to geopotential heights over western Europe in view of assessing 3-day rainfall accumulations over the Isère River catchment at Grenoble, France. Results show that dimensional predictors are the most skillful features for predicting 3-day rainfall—bringing alone 60% of the predictability of the classical analogy approach—whereas the dynamical predictors are less explicative. These results open new directions of research that the classical analogy approach cannot handle. They show, for instance, that both dry sequences and strong rainfall sequences are associated with singular 500-hPa geopotential shapes acting as local attractors—a way of explaining the change in rainfall CDFs in a changing climate.

1. Introduction

Relying on the hypothesis that similar synoptic situations are likely to result in similar local effects despite unexplained variability (Lorenz 1969), the analog method has been developed in the 1980s for forecasting daily rainfall depths based on analogy in geopotential heights (Duband 1981). Since then the method has been widely applied, mainly for daily and intradaily precipitation, using various predictors ranging from pressure fields to wind and rain fields over domains of various size and configurations (Chardon et al. 2014; Dayon et al. 2015; Horton and Brönnimann 2019). It has been used either in forecasting (Ben Daoud et al. 2016; Duband 1981) and, more recently, in nowcasting (Atencia and Zawadzki 2015; Panziera et al. 2011) or in downscaling larger-scale outputs (Dayon et al. 2015; Raynaud et al. 2017; Zorita and von Storch 1999).

Application of the analog method to *sequences* of synoptic situation is much less developed. Matulla et al.

(2008) and Zhou and Zhai (2016) downscale precipitation at a given target day from the state of the atmosphere on that day and the preceding 7 days. The main motivation for looking at sequences is that the local-scale variable in a given day may also result from the atmospheric states of the antecedent days. They evidence some benefit in accounting for antecedent days for estimating persistent ordinary features such as dry and wet spells in California and Austria (Matulla et al. 2008) or persistent large precipitation (exceeding the 95% quantile) in China (Zhou and Zhai 2016). Zhou et al. (2018) also consider the analog method accounting for precedent days to forecast two persistent extreme precipitation events that caused severe flooding in the Yangtze–Huai River valley (China) in summer 2016. They show the method outperforms the ensemble mean and about half the ensemble members of a numerical weather model for lead times longer than 3 days.

To the best of our knowledge, the only study that addresses the frequency of occurrence of multiday precipitation accumulation—and particularly the extremes—on the basis of analogy of atmospheric sequences is

Corresponding author: Juliette Blanchet, juliette.blanchet@univ-grenoble-alpes.fr

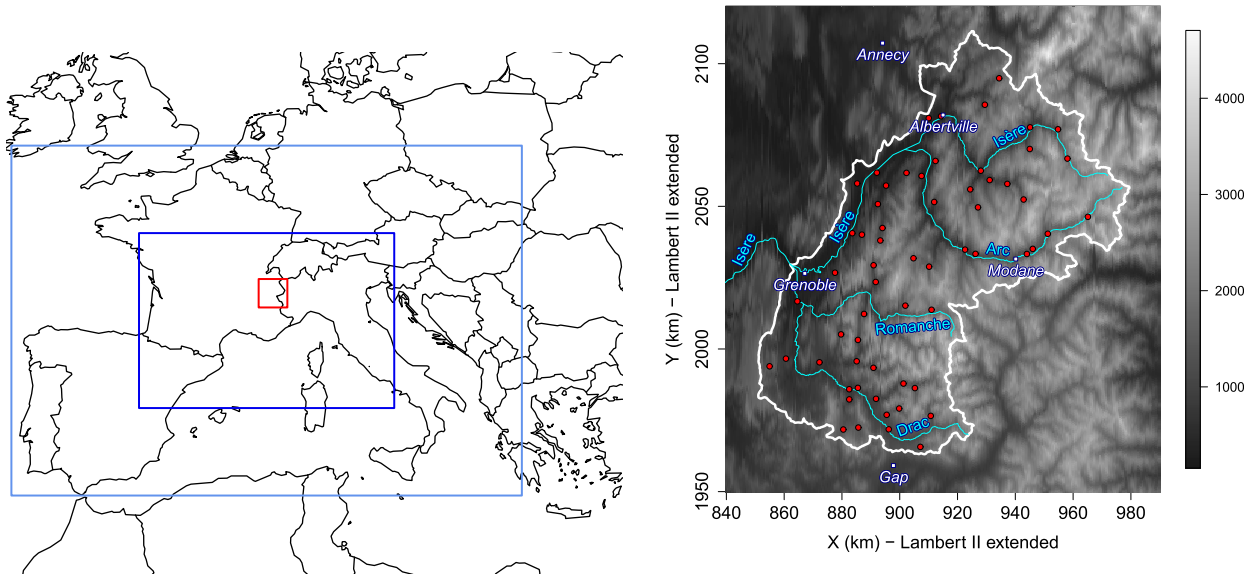


FIG. 1. (left) Map of western Europe with the analogy windows of geopotential heights at 500 (light blue) and 1000 (blue) hPa and the region of interest (red) outlined. (right) Elevation map of the region (m), with the main cities and rivers shown. The 61 rain gauges are indicated in red. The coordinate projection is according to the Lambert II extended system.

Blanchet et al. (2018). The study shows that the average singularity of geopotential shapes—defined as the mean distance to the nearest analogs—together with the celerity of deformation of the geopotential states remarkably stratify the CDF of multiday precipitation extremes. However, the ability of the celerity–singularity space to predict the frequency of occurrence of multiday extremes is not quantitatively evaluated.

This article aims at assessing the degree of predictability of multiday rainfall CDF based on analogy of low-dimensional atmospheric predictors. The considered predictors intend to summarize the topology of the space defined by the geopotential field sequences. They characterize how the sequences of pressure fields cluster locally (dimensional predictors) and how fast they evolve (dynamical predictors). They include the celerity and singularity of Blanchet et al. (2018), together with the two predictors used in Faranda et al. (2016, 2017), Messori et al. (2017), and Rodrigues et al. (2018) describing the local clustering of atmospheric states—the local dimension and the local stability—that are considered here for atmospheric sequences. The phase spaces associated with different combinations of these predictors are compared in terms of their ability to predict multiday rainfall CDF in the French Alps for the dry, wet, and extremely wet sequences. Comparison is made with the more common analog framework predicting multiday rainfall CDF directly from the most similar atmospheric sequences. Section 2 presents the precipitation and atmospheric data. Section 3 develops the classical analogy approach

and the low-dimensional predictor analogy approach. Section 4 compares results of the two approaches for predicting multiday rainfall CDF in a cross-validation framework. Section 5 concludes.

2. Data

The data and the study region are the same as in Blanchet et al. (2018). We consider the Isère River catchment at Grenoble, France, in the northern French Alps (Fig. 1). The altitude of the catchment ranges from about 200 m in the Isère River valley to more than 4100 m in the Ecrins national park. The median altitude is about 1500 m. The region experiences cold winters and warm summers. Precipitation falls mainly as snow above 1500 m between December and March and originates mainly from the Atlantic Ocean, whereas air flows originating from both the Atlantic and the Mediterranean Sea trigger precipitation in summer.

Since the Isère River catchment has a response time of a few days, we consider 3-day rainfall accumulations at the catchment scale as a good flood predictor. Catchment daily series are obtained by interpolating daily rain gauge data. We use 61 stations provided by MétéoFrance, featuring less than 25% of missing data in the period 1950–2011. Daily data are interpolated on a $1 \times 1 \text{ km}^2$ grid using a thin-plate spline function (Duchon 1977; Wahba and Wendelberger 1980). Catchment series are obtained by summing these gridded data.

With a view toward modeling the CDF of 3-day rainfall accumulation in the Isère River catchment, we consider

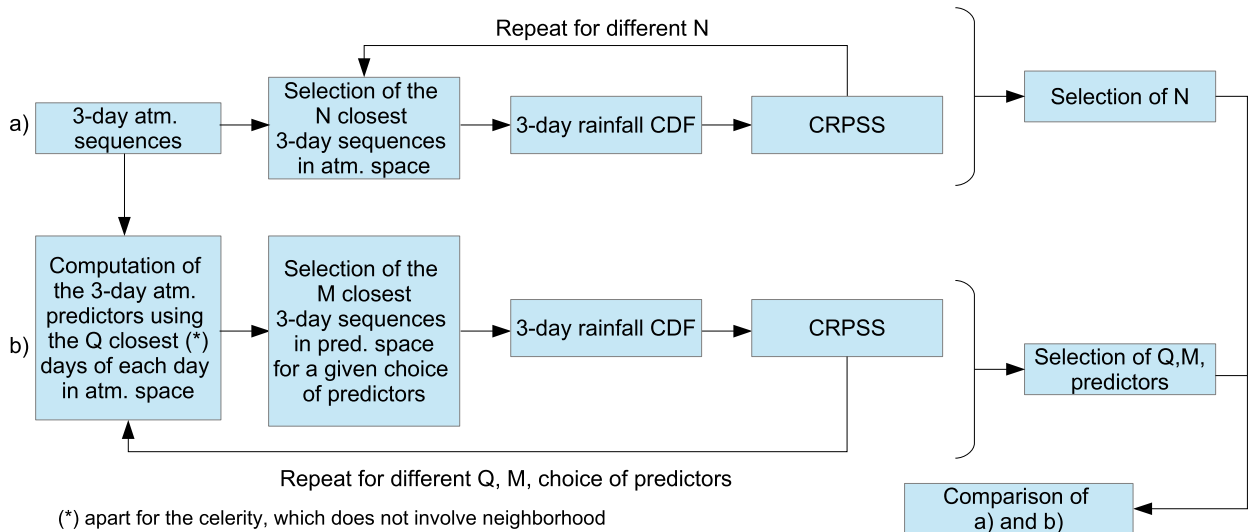


FIG. 2. Cross-validation workflow for prediction analyses: (a) classical analogy approach and (b) low-dimensional analogy approach based on atmospheric predictors of dynamics and dimensionality.

large-scale geopotential height fields over western Europe at 500 and 1000 hPa. These predictors have been found to be the most informative large-scale analogy predictors of daily precipitation for France (Obled et al. 2002; Radanovics et al. 2013), as well as for different regions in Europe with contrasted meteorological regimes (Raynaud et al. 2017). Daily means of geopotential heights are extracted from NOAA–CIRES Twentieth Century Reanalysis, version 2c (V2c; Compo et al. 2011), using arbitrarily the first member of the reanalyses. However, a separate analysis on the second and mean members of the reanalyses, as well as on the ECMWF twentieth century reanalysis (ERA-20C; Poli et al. 2016), shows remarkably similar results, so we think the choice of the reanalysis is actually of little importance for this study. The considered reanalysis has a spatial resolution of 2° in longitude and latitude. The geopotential windows are chosen to be similar to the optimal windows found in Raynaud et al. (2017) in the region. Both windows are centered on 6°E longitude and 44°N latitude (see Fig. 1). The 500-hPa window is 32° in longitude and 16° in latitude, and the 1000-hPa window is one-half as big in both longitude and latitude. Although the reanalysis data are available back to 1851, in this study only the 1950–2011 period is used so as to match the rainfall measurement period.

3. Method

We present two distinct methods for computing multiday rainfall CDFs in an analogy framework. Figure 2 provides a schematic illustration of these methods in a cross-validation framework.

a. Classical analogy approach

Various versions of the analog method as a way of predicting daily rainfall CDF at local scale have been developed since the pioneering work of Duband (1981). For the great majority of the time, analogy is based on atmospheric states (Chardon et al. 2014; Dayon et al. 2015; Horton and Brönnimann 2019; Raynaud et al. 2017; Zorita and von Storch 1999), although analogy between atmospheric sequences is considered in Matulla et al. (2008) and Zhou and Zhai (2016) for constructing daily rainfall CDFs and in Yiou et al. (2013, 2014) for reconstructing daily sea level pressure, temperature, and wind fields above the North Atlantic. Here we present the analogy approach for downscaling multiday rainfall CDFs based on analogy of sequences of situations. To the best of our knowledge, the multiday-multiday case (i.e., downscaling multiday precipitation based on analogy of multiday atmospheric sequences) has never been developed so far.

Following previous works on probabilistic precipitation forecasting in France (Bontron and Obled 2005; Chardon et al. 2014; Raynaud et al. 2017), we consider analogy in shape of geopotential fields rather than in absolute value, using the Teweles–Wobus score (Teweles and Wobus 1954). Denoting z_{jk} as the height at a given pressure level, grid point s_j , and observation time t_k , the Teweles–Wobus score for days t_k and $t_{k'}$ is given by

$$\Delta_{k,k'} = \frac{\sum_{(j,j') \in \mathcal{A} \setminus dj} |(z_{jk} - z_{j'k}) - (z_{jk'} - z_{j'k'})|}{\sum_{(j,j') \in \mathcal{A} \setminus dj} \max(|z_{jk} - z_{j'k}|, |z_{jk'} - z_{j'k'}|)}, \quad (1)$$

where \mathcal{A}_{dj} is the set of adjacent grid points in horizontal and vertical directions in the analogy window of geopotential height at the considered pressure level.

Let S_k be the 3-day sequence (t_k, t_{k+1}, t_{k+2}) . The Teweles–Wobus score for sequences S_k and $S_{k'}$ at pressure p is defined as

$$\Delta_{S_k, S_{k'}} = \sum_{r=0,1,2} \Delta_{k+r, k'+r}. \quad (2)$$

The smaller $\Delta_{S_k, S_{k'}}$ is, the more similar the sequences of geopotential shapes will be. The N analog sequences of the sequence S_k at pressure p are defined as those $S_{k'}$ corresponding to the N smallest values of $\Delta_{S_k, S_{k'}}$. The choice of the neighborhood size N will be investigated in section 4a.

Let the random variable R_{S_k} be the precipitation accumulation during the sequence S_k and let r_{S_k} be the observed value. Relying on the hypothesis that precipitation during sequences analog to S_k are plausible alternatives to the observation r_{S_k} , the CDF of R_{S_k} is estimated as the CDF associated with its N analog sequences.

b. Low-dimensional predictor analogy approach

1) ATMOSPHERIC PREDICTORS OF DYNAMICS AND DIMENSIONALITY

The analog sequences of the classical approach of section 3a are determined by similarity in a high-dimensional space—the space of geopotential fields, hereinafter referred to as the “atmospheric space.” It has dimension $D = 16 \times 8$ for the 500-hPa heights and $D = 8 \times 4$ for the 1000-hPa heights. This section intends to define analogy in a low-dimensional space (basically 1, 2, or 3) using interpretable predictors. Wetterhall et al. (2005) consider analogs in a low-dimensional space of leading principal components (PCs) of daily pressure fields. They interpret the obtained PCs as different classes of circulation. However, this approach needs further developments to be extended to the multi-day case that requires PCs for sequences of pressure fields. Here we take another approach by constructing a low-dimensional space summarizing somehow the local topology of the atmospheric space in terms of dimensionality and dynamics. The dimensionality summarizes how similar the most similar sequences of geopotential fields are. In other words, it describes whether we are along a highly reproducible atmospheric trajectory or not, and this whatever the shape of the geopotential field. The dynamics summarizes how fast a sequence of geopotential fields or its surrounding sequences evolves in time.

Following Blanchet et al. (2018), Faranda et al. (2016), Faranda et al. (2017), and Messori et al. (2017), there are

two considered predictors of dimensionality of a given day t_k . The first is the singularity (sing) defined as the mean Teweles–Wobus score between t_k and its Q closest days (Blanchet et al. 2018):

$$\text{sing}_k = \frac{1}{Q} \sum_{q=1}^Q \Delta_{k,(q)}, \quad (3)$$

where $\Delta_{k,(1)}, \dots, \Delta_{k,(Q)}$ are the Q smallest values, in ascending order, of $\Delta_{k,1}, \Delta_{k,2}, \dots$. The sing describes the reproducibility in shape of the atmospheric field. The lower it is, the more the geopotential shape at day t_k is similar to that at some other days. The second predictor of dimensionality is the relative singularity (rsing), defined as the singularity divided by $\Delta_{k,(Q)}$:

$$\text{rsing}_k = \frac{\text{sing}_k}{\Delta_{k,(Q)}}. \quad (4)$$

The rsing lies between 0 and 1. It characterizes the relative usualness (in shape) of the atmospheric field at day t_k : the smaller rsing is, the closer it is to its neighbors relative to the surrounding density, that is, the more likely the shape at day t_k relative to the variability of the surrounding shapes. Considering the D -dimensional space of atmospheric fields ($D = 128$ for 500 hPa and $D = 32$ for 1000 hPa), rsing characterizes the local degree of clustering around the point associated with day t_k in this space. Noting that the average value of rsing for uniformly distributed points in a D -dimensional space is $D/(D + 1)$, the cases of local “attraction” (clustering) will start with rsing values lower than 0.99 for the 500-hPa case and 0.97 for the 1000-hPa case.

The relative singularity defined by Eq. (4) is very closely related to the local dimension of Faranda et al. (2016, 2017) and Messori et al. (2017). However, therein a logarithmic transformation of $\Delta_{k,k'}$ is used, which makes its estimation maybe more robust (using extreme value theory) but its interpretation less straightforward.

Also following Blanchet et al. (2018), Faranda et al. (2016), Faranda et al. (2017), and Messori et al. (2017), there are two considered dynamical predictors. The first is the celerity (cel) defined as the Teweles–Wobus score between t_{k-1} and t_k (Blanchet et al. 2018):

$$\text{cel}_k = \Delta_{k-1,k}. \quad (5)$$

The cel characterizes the celerity of deformation of the atmospheric field from one day to another. The second dynamical predictor is the local persistence (pers), defined as the probability that any analog field of the

geopotential field at day t_k remains an analog on the following day:

$$\text{pers}_k = \Pr(\Delta_{k,(q)+1} \leq \Delta_{k,(Q)} | \Delta_{k,(q)} \leq \Delta_{k,(Q)}), \quad (6)$$

where $(q) + 1$ is the following day of the q th-most analog day to t_k , for $q = 1, \dots, Q$. According to the geometric distribution, $1/\text{pers}$ is the mean residence time of neighbors of t_k in its neighborhood. The pers is thus the inverse of the θ parameter of Faranda et al. (2016, 2017) and Messori et al. (2017). The local persistence depends on the mean speed of deformation of the geopotential fields similar in shape to day t_k . It is thus related to the dynamics of deformation, but it differs from the celerity of Eq. (5) on two main points: first, it involves a neighborhood (i.e., the most similar shapes), with the underlying idea that the speed of deformation of my neighbors informs one about my own deformation; second, since neighborhoods can have different sizes, it describes the relative celerity of deformation, accounting for the variability of the surrounding shapes.

Averaging the above daily predictors over the days of a given sequence S_k gives the predictors cel_{S_k} , sing_{S_k} , rsing_{S_k} and pers_{S_k} . Note that, of the four above predictors the celerity is the only one not involving neighboring sequences since the computation of sing , rsing , and pers involves the Q closest geopotential fields (the choice of the neighborhood size Q will be discussed in section 4b). In a separate analysis, we also considered as predictor for t_k the mean celerity of the neighbors of t_k , but this choice did not improve the results. For the sake of conciseness and to be consistent with Blanchet et al. (2018), we do not report these results here.

2) TWO-STAGE ANALOGY

Analogy is constructed in two stages (Fig. 2). The first stage is an analogy in the pressure field space. It relies on the Teweles–Wobus score, allowing us to define the predictors of section 3b(1) based on the Q -nearest sequences (apart from the celerity, which does not involve neighboring sequences). The second stage considers a set of predictors (e.g., 1–3 in section 4b). It defines analogy in a “predictor space” by determining the M analog sequences of dynamics and dimensionality based on the Euclidean distance between the standardized predictors (i.e., the predictors divided by their standard deviation since they have very different ranges). Relying on the hypothesis that precipitation during analog sequences of S_k in the predictor space is a plausible alternative to the observation r_{S_k} , the CDF of R_{S_k} is estimated as the CDF associated with its M analog sequences.

Note that analogy in the predictor space represents a paradigm shift compared with the classical analogy

approach. Whereas the classical approach relies on the similarity of the geopotential field sequences, the predictor approach relies on the similarity of the local topology, characterized by the way the geopotential sequences cluster locally (sing and rsing) and how fast they evolve (cel and pers), somehow forgetting their shapes.

c. Cross-validation workflow

To compare the two analogy approaches, we consider a onefold cross-validation framework. As illustrated in Fig. 2, the framework allows us 1) to select N , the optimal number of neighbors to consider in the classical analogy approach; 2) to select M and Q , the optimal number of neighbors to consider in the two-stage predictor analogy approach; 3) to compare performance of different subsets of predictors; and 4) to compare performance of the two analogy approaches.

The cross-validation framework is constructed as follows. Let S_k be a given sequence. We compute the CDF for S_k based on the 3-day rainfalls for the analog sequences of S_k , excluding S_k , using either analogy approach. Since rainfall accumulations can be either null or positive, the CDF for any S_k , $F_{S_k}(r)$, can be written as

$$F_{S_k}(r) = p_{S_k}^0 + (1 - p_{S_k}^0)F_{S_k}^+(r), \quad (7)$$

where $p_{S_k}^0$ is the probability of zero rainfall accumulation and $F_{S_k}^+$ is the CDF of nonzero rainfall accumulation. We estimate $p_{S_k}^0$ empirically. We tested different parametric distribution for modeling $F_{S_k}^+$ [including the gamma distribution and the more flexible extended generalized Pareto distribution of Naveau et al. (2016)] but for the sake of conciseness, to avoid discussing estimation method and choice of distribution, and since the results presented here barely change, we present below only the results when $F_{S_k}^+$ is estimated empirically.

As an objective way of assessing prediction performance, we compute the continuous ranked probability score (CRPS) for S_k . It allows us to evaluate the predicted cumulative distribution functions $F_{S_k}(r)$, in comparison with the single observed value r_{S_k} for sequence S_k :

$$\text{CRPS}_{S_k} = \int_0^{+\infty} F_{S_k}(r) - H(r - r_{S_k})^2 dr, \quad (8)$$

where $H(r - r_{S_k})$ is the Heaviside function that is null when $r - r_{S_k} < 0$ and has the value 1 otherwise. The better the prediction for S_k is, the lower the CRPS score (CRPSS) will be. As a way of comparing the overall performance over all sequences, we consider its skill score expression, with the climatological distribution of precipitation as the reference:

$$\text{CRPSS} = 1 - \frac{\sum_k \text{CRPS}_{S_k}}{\sum_k \text{CRPS}_{S_k}^0}, \quad (9)$$

where $\text{CRPS}_{S_k}^0$ is computed using Eq. (8) with $F_{S_k}(r)$ replaced by $F^0(r)$, the empirical 3-day rainfall CDF computed over all sequences. The better the prediction is, the larger the CRPSS will be. A CRPSS of 0 means that the analogy approach does not improve prediction when compared with the climatological distribution. A CRPSS of, for example, 0.2, corresponds to a decrease of 20% of the average CRPS provided by analogy. To focus on different parts of the distribution, we compute the CRPSS over 1) all sequences, 2) all wet sequences, 3) the 62 largest rainfall accumulations (roughly corresponding to the annual maxima), and 4) the indicator values of rain/no rain [in the latter case, F_{S_k} in Eq. (8) is the Bernoulli distribution with parameter $1 - p_{S_k}^0$].

4. Results

a. Classical analogy approach

Unless said otherwise, we consider analogy in geopotential heights at 500 hPa. Comparison with geopotentials at 1000 hPa, and combination of the two of them, will be provided in section 4b. Figure 3 compares the CRPSS of the classical analogy approach depending on the number of neighbors, here given as a percentage of the total number of sequences, that is, in the form $N = n\alpha_N$, with $n = 22\,643$ denoting the number of 3-day sequences. In addition to the 3-day case, we temporally consider the daily case for a comparison with the most common use of the method. Note that the CRPSS values in the daily and 3-day cases cannot be compared one-to-one since they involve different normalizations [i.e., different CRPS^0 in Eq. (9)]. A larger value of CRPSS in, for example, the 3-day case, means that improvement in CRPS in comparison with the climatology is larger than in the 1-day case, but this does not necessarily imply a better CRPS.

According to the CRPSS values, analogy in the 3-day case improves overall prediction performance over the daily case (larger CRPSS values). However, prediction of the rain/no rain occurrence is more improved in the daily case mainly because there are more very small rainfall values at 3-day scale than at daily scale, which makes harder the distinction between rain and no rain. Comparison of the CRPSS values when the number of neighbors changes shows that the best prediction performance overall is obtained when approximately 45–110 neighbors are considered (0.2%–0.5% of the days or sequences). Decrease in prediction

performance when the number of neighbors increases is particularly marked for the largest rainfalls, for which the better performance is achieved with 22–45 neighbors (0.1%–0.2%). As a compromise and for the sake of robustness, we select at 3-day scale $\alpha_N = 0.2\%$, although smaller neighborhoods seem to be even slightly better for the extremes. However, 1) it would not make sense, in a prediction framework, to select the neighborhood depending on the rainfall accumulations and 2) very small neighborhoods lead to nonrobust estimates of the CDFs. There is obviously a balance to find between high degree of analogy and decent number of analogs. Considering 0.2% of analogs seems to be a good compromise.

b. Low-dimensional predictor analogy approach

We consider nine subsets of predictors, from 1 to 3 dimensions, as shown in the x labels of Fig. 4. The other subsets involving dynamical predictors only are not shown for the sake of conciseness and because they give worse results. We tested for each subset different sizes of neighborhoods, that is, different values of Q —used in the computation of predictors—and M —determining analog sequences—with α_Q and α_M varying from 0.2% to 2%. For clarity reasons, we report in Fig. 4 the CRPSS values for α_Q and α_M being either 0.2% or 0.5%, which are the best-performing values. Figure 4 shows that differences in CRPSS values with respect to α_Q and α_M (and up to 2%; not shown) for a given set of predictors are lower than differences among predictors. In other words, the predicting performance comes at the first order from the choice of predictors, rather than from the size of the neighborhoods of analogy.

Let us focus first on the single use of the singularity versus the relative singularity. Figure 4 shows that, overall, the relative singularity clearly underperforms the singularity. However, its predicting performance greatly increases for the largest rainfall accumulations. This means that, although the degree of clustering in the atmospheric space (rsing) is not a good predictor when compared with the density of points (sing) for usual rainfalls and intermittency, the largest rainfall accumulation is triggered by atmospheric shapes that are unusually clustered within a high local density. It thus seems that the shapes responsible for extremes are very specific. The dynamical predictors (cel and pers) bring negligible to minor improvement except for the intermittency. However, we note that by considering daily means of geopotentials at 0000, 0600, 1200, and 1800 UTC (see section 2) we smooth out instantaneous fields, which may be disadvantaging the dynamical predictors. The use of the relative singularity together with the singularity slightly improves the prediction

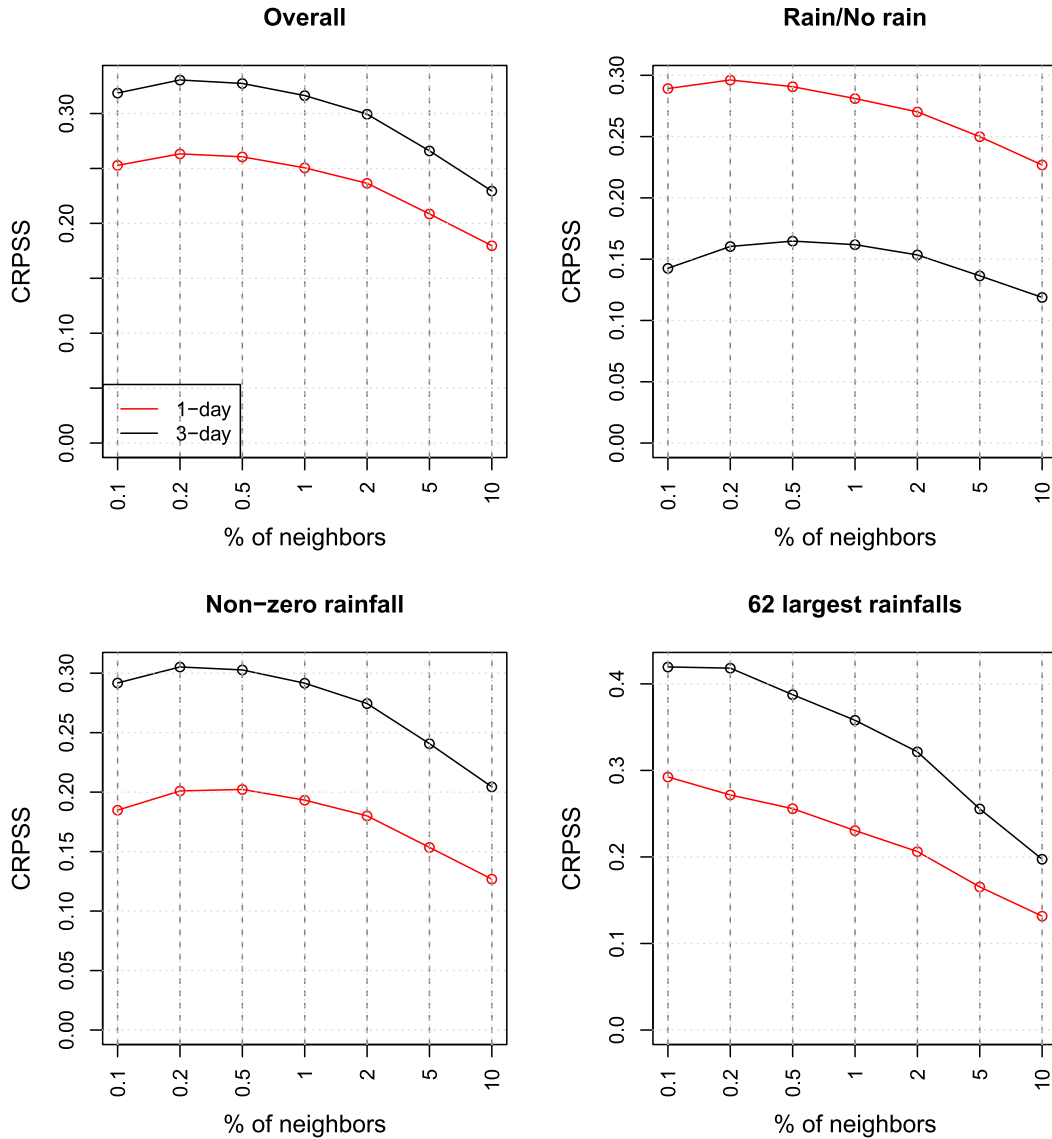


FIG. 3. Classical analogy approach applied to geopotential heights at 500 hPa: comparison of CRPSS for 1- (red) and 3-day (black) rainfall, depending on the number of neighbors, here indicated as a percentage of the total number of sequences or days (e.g., in the 3-day case, “0.5” means $N = 22\,643 \times 0.5\%$).

performance compared with the singularity alone, for the whole range of rainfall accumulations. For all cases, whether zero, common, or large rainfall accumulations, the best performance is achieved with the triplet celerity–singularity–relative singularity, but the greatest part of the CRPSS improvement comes from the singularity at first order and from the relative singularity at second order (mainly for the extremes). With regard to the choice of the size of the neighborhoods, the best CRPSS values are obtained whatever set of predictors for $\alpha_Q = 0.5\%$ and $\alpha_M = 0.5\%$ (red dotted lines). Although differences are small, we see that the optimal neighborhood for predicting rainfall—and particularly

the extremes—in the classical analogy approach (i.e., $\alpha_N = 0.2\%$ according to Fig. 3) is not the optimal neighborhood for computing predictors (i.e., $Q \neq N$).

A comparison of Figs. 3 and 4 shows that the best CRPSS values obtained with low-dimensional predictors are around 60% those of the classical analogy approach. As expected, there is a significant loss in using the mere topology of the surrounding atmospheric space instead of the full list of analog atmospheric sequences. However, a major advantage of the predictor approach is to drive toward a different and synthetic interpretation of the atmospheric features leading to zero-, small-, or large-rainfall accumulations. Figure 5 shows

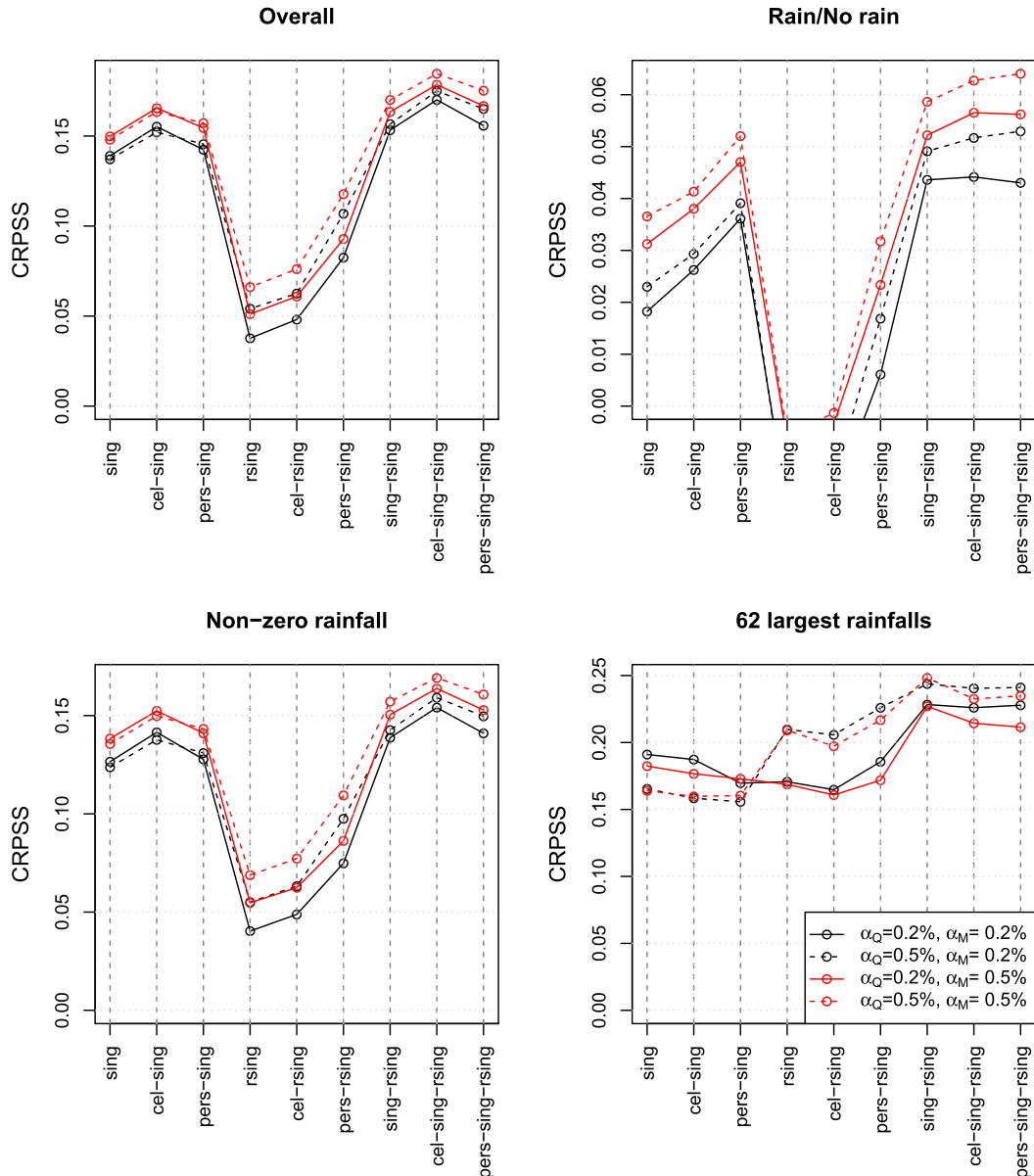


FIG. 4. Predictor analogy approach applied to 3-day geopotential heights at 500 hPa: comparison of CRPSS for different values of α_Q , which is used in the computation of the predictors, and α_M , which is used to determine analog sequences.

the two-dimensional predictor space for the five pairs of predictors. A look at the spatial distribution of points (Fig. 5a) reveals a clear nonuniform spreading. Almost one-half of each predictor space is empty. Even in the half of predictor space that is not empty, points are clearly nonuniformly distributed. The densest parts of each scatterplot are about 1500 times as dense as the least dense parts.

We colored in Figs. 5b and 5c the points of the predictor space with respect to the proportion of dry 3-day sequences (Fig. 5b) and the average nonzero 3-day rainfalls (Fig. 5c) in the neighborhood analogy of each point (using $\alpha_Q = \alpha_M = 0.5\%$). A comparison of the

two scatterplots of a given pair of predictors (Fig. 5b vs 5c) shows, in almost all cases, a clear duality. The parts of the predictor space with the largest probability of dry sequence and with the largest rainfall accumulation are almost opposite one another. The values themselves are well contrasted: the probability of dry sequences lies between 0 and about 40%, which is considerable given that the average proportion of dry sequences is 8%. The average nonzero 3-day rainfall varies from around 0 to more than 10 mm day^{-1} , the latter corresponding overall to the 90% quantile of nonzero rainfall. Each pair of predictors is able to continuously stratify the rain/no rain probability

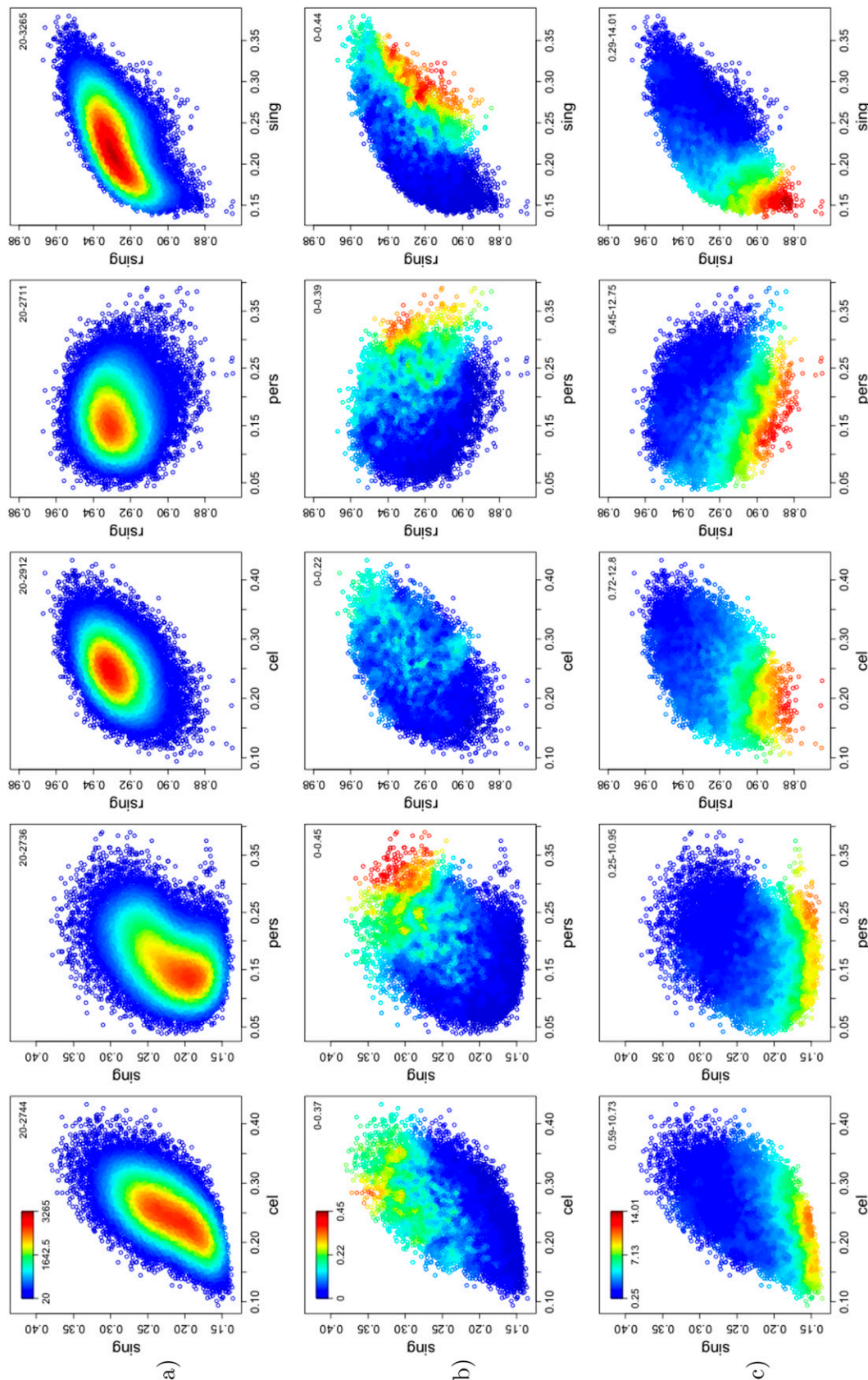


FIG. 5. Two-dimensional predictor space of (left) celerity–singularity, (left center) persistence–singularity, (center) celerity–relative singularity, (right center) persistence–relative singularity, and (right) singularity–relative singularity, corresponding to 500-hPa geopotential heights. The points of each row are colored with respect to (a) the density of points (number of points at a distance lower than one-half of the standard deviation), (b) the probability of zero rainfall, or (c) the mean of nonzero rainfall (mm day^{-1}). The same color scale is used for the five panels of a given row. The top-right numbers indicate the range of values actually taken in the corresponding panel.

and the rainfall accumulations. Of course, some pairs perform better than others. For example, as already seen in Fig. 4, the celerity–relative singularity provides the fuzziest stratification of the rain/no rain probability, although it sharply stratifies the rainfall accumulations. Most of the rainfall stratification is actually provided by the singularity and the relative singularity, while the other predictors improve relatively little, in coherence with the CRPSS scores of Fig. 4. However, note that the largest rainfall accumulations tend to occur for relatively small celerities, corresponding to situations of greater steadiness of the atmospheric shapes, as also noted in Blanchet et al. (2018). The best stratification of the rainfall accumulations is obtained when the singularity and relative singularity are used jointly but, in coherence with Fig. 4, the influence of the relative singularity is only significant for the largest accumulations, which are associated with the lowest values of both predictors. In particular, the geopotential sequences triggering more than 10 mm day^{-1} have a relative singularity of 0.86–0.90, which is extremely unlikely for uniformly distributed points in the atmospheric space (p value lower than 10^{-6}). This means that the geopotential sequences leading to the largest rainfall accumulations have reproducible shapes (low singularity) but, using Lorenz' terminology (Lorenz 1969), each shape is an attractor of the atmospheric space, inducing a higher degree of clustering in its neighborhood (low relative singularity). Dry sequences tend to show medium to large singularity but are still among the lowest relative singularity for the range of singularity. The geopotential shapes of the dry sequences are also attractors. We see then two poles of attractors: one corresponding to usual shapes and associated with large rainfall accumulations, and the other one corresponding to less common shapes and associated with dry sequences. Note that the intermittency is also remarkably associated with persistent atmospheric shapes, showing that attraction acts in time.

Until now the analysis considered the 500-hPa geopotential heights. As a way of comparing the prediction skills depending of the geopotential, we show in Fig. 6 the scatterplots corresponding to 1000 hPa, to be compared with Figs. 5b and 5c. These new diagrams differ from many respects. In general, the scatterplots at 1000 hPa are noisier and less contrasted than at 500 hPa (note that the color scales differ between Figs. 5 and 6). They also differently stratify the mean rainfall. In particular the dynamical predictors (celerity and persistence) stratify more at 1000 hPa than at 500 hPa (cf., e.g., the y -oriented stratification of the celerity–relative singularity scatterplot in Fig. 5c with the xy -oriented stratification in Fig. 6). With regard to the dimensional predictors, the singularity–relative singularity plane at 1000 hPa does not stratify the probability of

dry sequence. The largest values of the probability of dry sequence correspond to medium values of the relative singularity—meaning that geopotential shapes of the dry sequences are not attractors—but among the largest values of persistence. Moreover, the scatterplot in the persistence–singularity plane varies notably in shape when going from 1000 to 500 hPa. In particular, persistence and singularity are never large at the same time, whereas this corresponds to the largest probability of dry sequence at 500 hPa (see Fig. 5b). Notably, for both geopotentials, persistence explains intermittency, whereas relative singularity (local clustering) explains accumulations, although the explicative power of the 1000-hPa shape is considerably lower.

As a way of quantifying the qualitative differences commented above, we show in Fig. 7 the CRPSS scores obtained for 500 hPa, 1000 hPa, and the combination of them. In the latter case, the Teweles–Wobus score used for the computation of predictors is the average Teweles–Wobus score for both heights (Blanchet et al. 2018). The scores at 1000 hPa are almost always lower than at 500 hPa, in accordance with Fig. 5 versus Fig. 6. At whatever height, the relative singularity is useless for intermittency, while it is indispensable for the large rainfall accumulations at 1000-hPa height. For the large rainfalls, the choice of predictors (among those compared) is much less discriminant than the choice of geopotential. The scores when using the combination of geopotentials are almost the average of the individual scores, apart for the large rainfall accumulations. Surprisingly, the classical analogy approach shows inverse results. First, geopotential shapes at 500 hPa are less discriminant than at 1000 hPa. This is in accordance with the relevance maps found in Horton et al. (2012) to explain rainfall readings at Marécottes, a Swiss gauging station close to our study area and exposed to comparable weather situations. Second, the two heights seem to bring complementary information since combining them improves the scores rather than averaging them. Obled et al. (2002) also finds that combining geopotential fields improves predictive skills of basin precipitation in France. We note that the best score in the predictor analogy approach—corresponding to the celerity–singularity–relative singularity at 500 hPa—is only 0.5 times the best score of the classical analogy approach, corresponding to the combination of geopotentials at 500 and 1000 hPa.

5. Conclusions and discussion

We proposed in this article a new two-stage analogy approach for explaining 3-day rainfall CDF based on atmospheric predictors of dynamics and dimensionality. The method differs from the classical analogy approach in that it is based on the mere topology of the surrounding

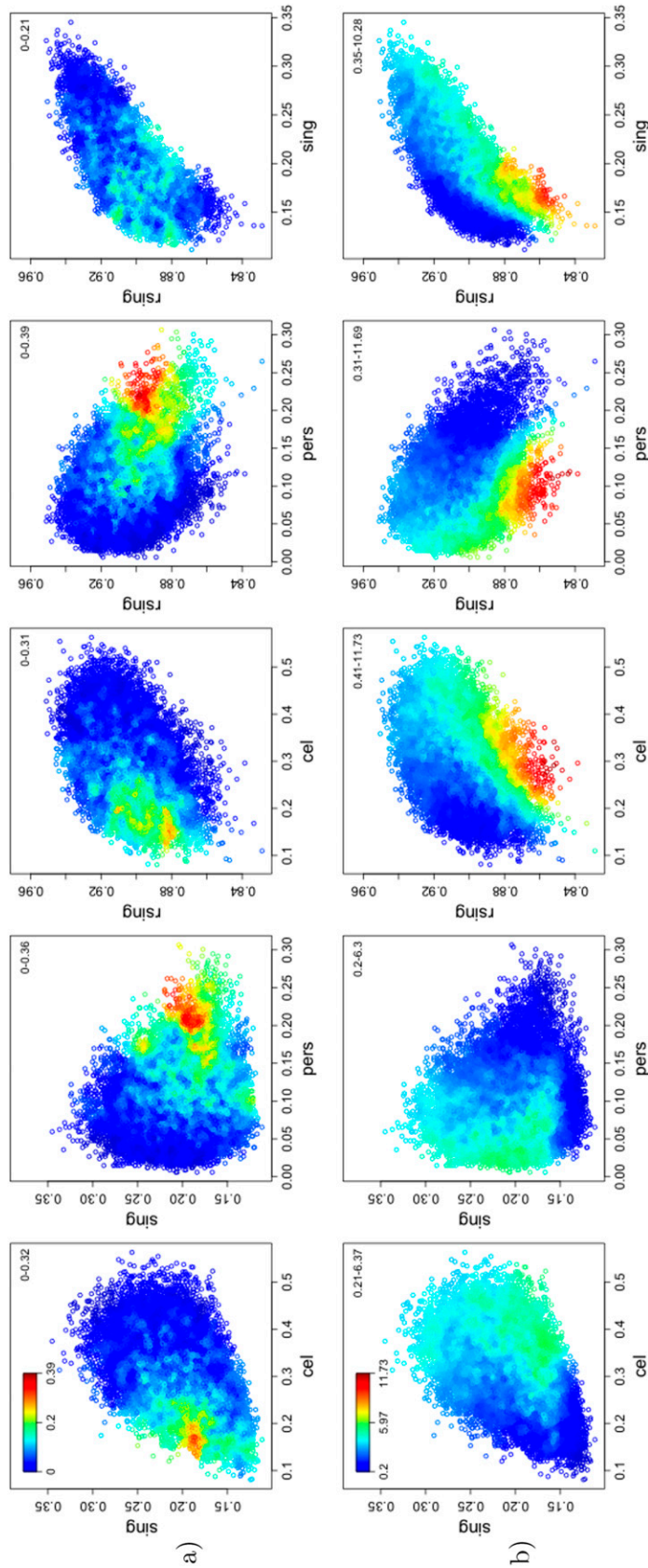


FIG. 6. As in Fig. 5, but for 1000-hPa geopotential heights. The points of each row are colored with respect to (a) the probability of zero rainfall or (b) the mean of nonzero rainfall (mm day⁻¹).

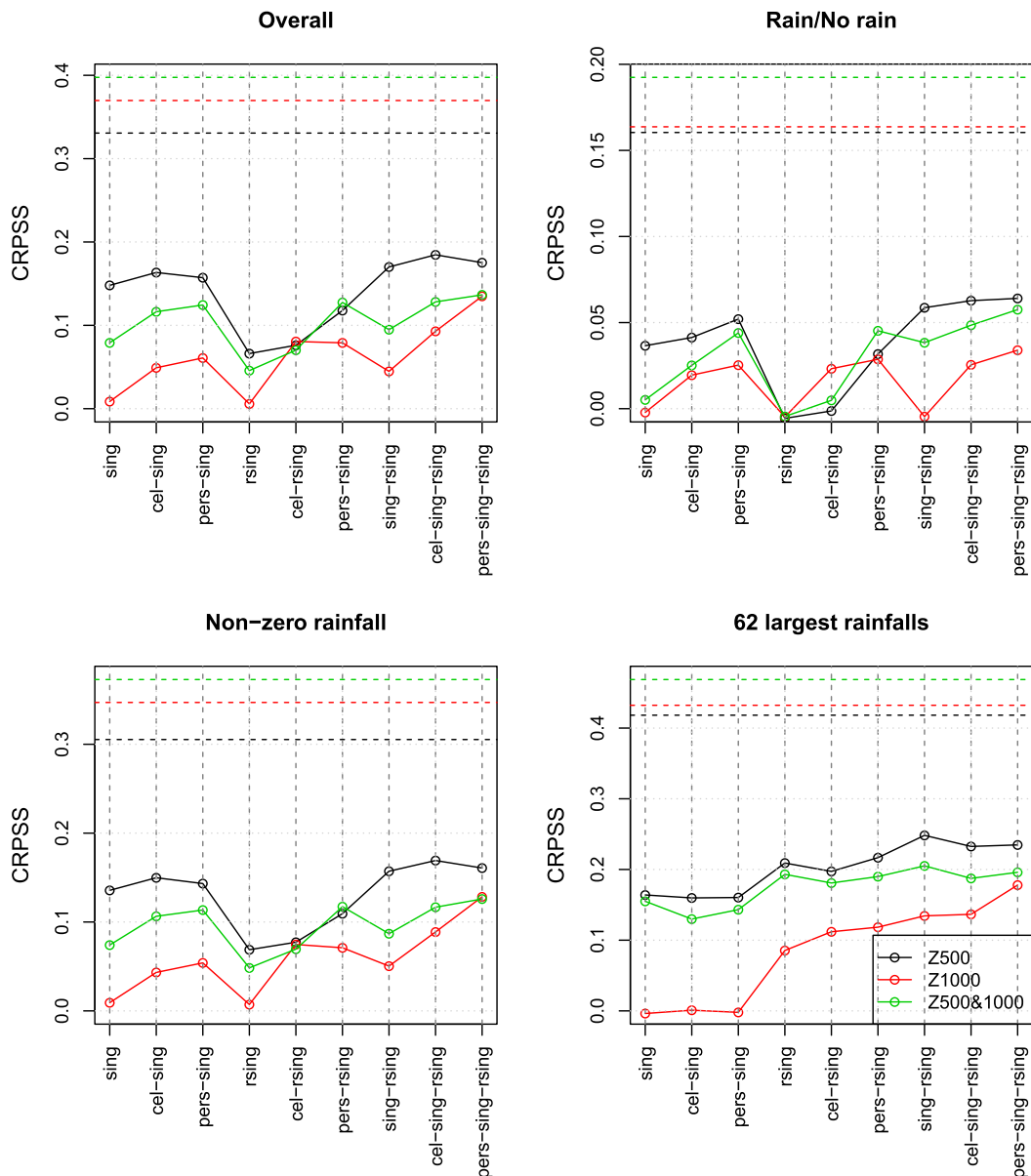


FIG. 7. Predictor analogy approach applied to 3-day rainfall: comparison of CRPSS when using geopotential heights at 500 hPa, 1000 hPa, or both of them. The dotted horizontal lines show the CRPSS of the classical analogy approach.

atmospheric space instead of the full list of analog atmospheric sequences. Our results confirm previous works showing the link between rainfall accumulations and atmospheric mesoscale circulation in western Europe. For France, the studies of [Boé and Terray \(2008\)](#) and [Garavaglia et al. \(2010\)](#), for example, aimed at segmenting the atmospheric space of pressure fields over time into a set of weather types governing precipitation properties over target regions. With the two-stage analogy we propose to carry out this segmentation through an intermediate low-dimensional space mapped with predictors summarizing the topology of the atmospheric space. The

position of each sequence in the low-dimensional space appears to govern corresponding rainfall characteristics and opens the way to a “continuous” examination of the atmospheric space according to the local topology.

Examining the predictive power of topological predictors sheds a new light on the governing meteorological situations. We revealed, for instance, that both dry sequences and strong rainfall sequences are associated with geopotential shapes acting as local attractors, the 500-hPa field being much more significant than the 1000-hPa one. The fact that these attractors concern neighborhoods of the atmospheric space that display respectively low

(dry sequences) and high (strong rainfall sequences) densities deserve some care to interpret. Strong rainfall sequences come from geopotential fields that reproduce very narrowly the same shape (very low singularity and relative singularity) and that are very seldom (low density area in the singularity–relative singularity plane). A low singularity thus does not mean the corresponding pressure fields are common, this means that the way they resemble some others is stronger and this topological situation is rare.

These results could be obviously improved in various ways, in particular by considering new atmospheric predictors—for example, differences in celerities between 500 and 1000 hPa, or complementing the geopotentials by other atmospheric predictors such as moisture index, vertical velocity, helicity, or temperature (Marty et al. 2012; Chardon et al. 2014, 2018; Ben Daoud et al. 2016), or by using measures other than Teweles–Wobus score or intermediate geopotentials such as the 700- or 850-hPa height (Horton et al. 2018). Even more interesting is that the new analogy approach based on low-dimensional predictors opens new directions of research that the classical analogy approach cannot handle, such as 1) atmospheric predictors as a first-order measure of a climate model's ability to reproduce rainfall statistics, 2) the change in predictors as a way of explaining the change in rainfall CDF in future climate without looking explicitly to the CDF, and 3) the predictor space as a way of probabilizing atmospheric sequences.

Acknowledgments. This study was initiated by Simon Stalla at IGE Grenoble. It is part of a collaboration between the University Grenoble Alpes and Grenoble Alpes Métropole, the metropolitan authority of the Grenoble conurbation (deliberation 12 of the Metropolitan Council of 27 May 2016). Twentieth Century Reanalysis data are provided by the NOAA/OAR/ESRL PSD (https://www.esrl.noaa.gov/psd/data/gridded/data.20thC_ReanV2c.pressure.html). Support for the Twentieth Century Reanalysis Project, version 2c, dataset is provided by the U.S. Department of Energy Office of Science Biological and Environmental Research (BER) and by the National Oceanic and Atmospheric Administration Climate Program Office. The rainfall data used in this paper are maintained by Météo-France and are available online (<https://publitheque.meteo.fr>; login requested).

REFERENCES

- Atencia, A., and I. Zawadzki, 2015: A comparison of two techniques for generating nowcasting ensembles. Part II: Analogs selection and comparison of techniques. *Mon. Wea. Rev.*, **143**, 2890–2908, <https://doi.org/10.1175/MWR-D-14-00342.1>.
- Ben Daoud, A., E. Sauquet, G. Bontron, C. Obled, and M. Lang, 2016: Daily quantitative precipitation forecasts based on the analogue method: Improvements and application to a French large river basin. *Atmos. Res.*, **169**, 147–159, <https://doi.org/10.1016/j.atmosres.2015.09.015>.
- Blanchet, J., S. Stalla, and J.-D. Creutin, 2018: Analogy of multiday sequences of atmospheric circulation favoring large rainfall accumulation over the French Alps. *Atmos. Sci. Lett.*, **19**, e809, <https://doi.org/10.1002/asl.809>.
- Boé, J., and L. Terray, 2008: A weather-type approach to analyzing winter precipitation in France: Twentieth-century trends and the role of anthropogenic forcing. *J. Climate*, **21**, 3118–3133, <https://doi.org/10.1175/2007JCLI1796.1>.
- Bontron, G., and C. Obled, 2005: A probabilistic adaptation of meteorological model outputs to hydrological forecasting. *La Houille Blanche*, **1**, 23–28, <https://doi.org/10.1051/lhb:200501002>.
- Chardon, J., B. Hingray, A.-C. Favre, P. Autin, J. Gailhard, I. Zin, and C. Obled, 2014: Spatial similarity and transferability of analog dates for precipitation downscaling over France. *J. Climate*, **27**, 5056–5074, <https://doi.org/10.1175/JCLI-D-13-00464.1>.
- , —, and —, 2018: An adaptive two-stage analog/regression model for probabilistic prediction of small-scale precipitation in France. *Hydrol. Earth Syst. Sci.*, **22**, 265–286, <https://doi.org/10.5194/hess-22-265-2018>.
- Compo, G. P., and Coauthors, 2011: The Twentieth Century Reanalysis Project. *Quart. J. Roy. Meteor. Soc.*, **137**, 1–28, <https://doi.org/10.1002/qj.776>.
- Dayon, G., J. Boé, and E. Martin, 2015: Transferability in the future climate of a statistical downscaling method for precipitation in France. *J. Geophys. Res. Atmos.*, **120**, 1023–1043, <https://doi.org/10.1002/2014JD022236>.
- Duband, D., 1981: Prévision spatiale des hauteurs de précipitations journalières. *La Houille Blanche*, **7/8**, 497–512, <https://www.shf-lhb.org/articles/lhb/pdf/1981/05/lhb1981046.pdf>.
- Duchon, J., 1977: Splines minimizing rotation-invariant seminorms in Sobolev spaces. *Constructive Theory of Functions of Several Variables*, W. Schempp and K. Zeller, Eds., Springer, 85–100.
- Faranda, D., G. Messori, and P. Yiou, 2016: Dynamical proxies of North Atlantic predictability and extremes. *Sci. Rep.*, **7**, 41278, <https://doi.org/10.1038/SREP41278>.
- , —, M. C. Alvarez-Castro, and P. Yiou, 2017: Dynamical properties and extremes of Northern Hemisphere climate fields over the past 60 years. *Nonlinear Processes Geophys.*, **24**, 713–725, <https://doi.org/10.5194/npg-24-713-2017>.
- Garavaglia, F., J. Gailhard, E. Paquet, M. Lang, R. Garçon, and P. Bernardara, 2010: Introducing a rainfall compound distribution model based on weather patterns sub-sampling. *Hydrol. Earth Syst. Sci.*, **14**, 951–964, <https://doi.org/10.5194/hess-14-951-2010>.
- Horton, P., and S. Brönnimann, 2019: Impact of global atmospheric reanalyses on statistical precipitation downscaling. *Climate Dyn.*, **52**, 5189–5211, <https://doi.org/10.1007/S00382-018-4442-6>.
- , M. Jaboyedoff, R. Metzger, C. Obled, and R. Marty, 2012: Spatial relationship between the atmospheric circulation and the precipitation measured in the western Swiss Alps by means of the analogue method. *Nat. Hazards Earth Syst. Sci.*, **12**, 777–784, <https://doi.org/10.5194/nhess-12-777-2012>.
- , —, and C. Obled, 2018: Using genetic algorithms to optimize the analogue method for precipitation prediction in the

- Swiss Alps. *J. Hydrol.*, **556**, 1220–1231, <https://doi.org/10.1016/j.jhydrol.2017.04.017>.
- Lorenz, E. N., 1969: Atmospheric predictability as revealed by naturally occurring analogues. *J. Atmos. Sci.*, **26**, 636–646, [https://doi.org/10.1175/1520-0469\(1969\)26<636:APARBN>2.0.CO;2](https://doi.org/10.1175/1520-0469(1969)26<636:APARBN>2.0.CO;2).
- Marty, R., I. Zin, C. Obled, G. Bontron, and A. Djerboua, 2012: Toward real-time daily pcpf by an analog sorting approach: Application to flash-flood catchments. *J. Appl. Meteor. Climatol.*, **51**, 505–520, <https://doi.org/10.1175/JAMC-D-11-011.1>.
- Matulla, C., X. Zhang, X. L. Wang, J. Wang, E. Zorita, S. Wagner, and H. von Storch, 2008: Influence of similarity measures on the performance of the analog method for downscaling daily precipitation. *Climate Dyn.*, **30**, 133–144, <https://doi.org/10.1007/S00382-007-0277-2>.
- Messori, G., R. Caballero, and D. Faranda, 2017: A dynamical systems approach to studying midlatitude weather extremes. *Geophys. Res. Lett.*, **44**, 3346–3354, <https://doi.org/10.1002/2017GL072879>.
- Naveau, P., R. Huser, P. Ribereau, and A. Hannart, 2016: Modeling jointly low, moderate, and heavy rainfall intensities without a threshold selection. *Water Resour. Res.*, **52**, 2753–2769, <https://doi.org/10.1002/2015WR018552>.
- Obled, C., G. Bontron, and R. Garçon, 2002: Quantitative precipitation forecasts: A statistical adaptation of model outputs through an analogues sorting approach. *Atmos. Res.*, **63**, 303–324, [https://doi.org/10.1016/S0169-8095\(02\)00038-8](https://doi.org/10.1016/S0169-8095(02)00038-8).
- Panziera, L., U. Germann, M. Gabella, and P. V. Mandapaka, 2011: Nora–nowcasting of orographic rainfall by means of analogues. *Quart. J. Roy. Meteor. Soc.*, **137**, 2106–2123, <https://doi.org/10.1002/qj.878>.
- Poli, P., and Coauthors, 2016: ERA-20C: An atmospheric reanalysis of the twentieth century. *J. Climate*, **29**, 4083–4097, <https://doi.org/10.1175/JCLI-D-15-0556.1>.
- Radanovics, S., J.-P. Vidal, E. Sauquet, A. Ben Daoud, and G. Bontron, 2013: Optimising predictor domains for spatially coherent precipitation downscaling. *Hydrol. Earth Syst. Sci.*, **17**, 4189–4208, <https://doi.org/10.5194/hess-17-4189-2013>.
- Raynaud, D., B. Hingray, I. Zin, S. Anquetin, S. Debonne, and R. Vautard, 2017: Atmospheric analogues for physically consistent scenarios of surface weather in Europe and Maghreb. *Int. J. Climatol.*, **37**, 2160–2176, <https://doi.org/10.1002/JOC.4844>.
- Rodrigues, D., M. C. Alvarez-Castro, G. Messori, P. Yiou, Y. Robin, and D. Faranda, 2018: Dynamical properties of the North Atlantic atmospheric circulation in the past 150 years in CMIP5 models and the 20CrV2c reanalysis. *J. Climate*, **31**, 6097–6111, <https://doi.org/10.1175/JCLI-D-17-0176.1>.
- Teweles, S., and H. Wobus, 1954: Verification of prognostic charts. *Bull. Amer. Meteor. Soc.*, **35**, 455–463, <https://doi.org/10.1175/1520-0477-35.10.455>.
- Wahba, G., and J. Wendelberger, 1980: Some new mathematical methods for variational objective analysis using splines and cross validation. *Mon. Wea. Rev.*, **108**, 1122–1143, [https://doi.org/10.1175/1520-0493\(1980\)108<1122:SNMMFV>2.0.CO;2](https://doi.org/10.1175/1520-0493(1980)108<1122:SNMMFV>2.0.CO;2).
- Wetterhall, F., S. Halldin, and C.-Y. Xu, 2005: Statistical precipitation downscaling in central Sweden with the analogue method. *J. Hydrol.*, **306**, 174–190, <https://doi.org/10.1016/J.JHYDROL.2004.09.008>.
- Yiou, P., T. Salameh, P. Drobinski, L. Menut, R. Vautard, and M. Vrac, 2013: Ensemble reconstruction of the atmospheric column from surface pressure using analogues. *Climate Dyn.*, **41**, 1333–1344, <https://doi.org/10.1007/S00382-012-1626-3>.
- , M. Boichu, R. Vautard, M. Vrac, S. Jourdain, E. Garnier, F. Fluteau, and L. Menut, 2014: Ensemble meteorological reconstruction using circulation analogues of 1781–1785. *Climate Past*, **10**, 797–809, <https://doi.org/10.5194/cp-10-797-2014>.
- Zhou, B., and P. Zhai, 2016: A new forecast model based on the analog method for persistent extreme precipitation. *Wea. Forecasting*, **31**, 1325–1341, <https://doi.org/10.1175/WAF-D-15-0174.1>.
- , —, and R. Niu, 2018: Comparative assessment of two objective forecast models for cases of persistent extreme precipitation events in the Yangtze–Huai River valley in summer 2016. *Wea. Forecasting*, **33**, 221–238, <https://doi.org/10.1175/WAF-D-17-0039.1>.
- Zorita, E., and H. von Storch, 1999: The analog method as a simple statistical downscaling technique: Comparison with more complicated methods. *J. Climate*, **12**, 2474–2489, [https://doi.org/10.1175/1520-0442\(1999\)012<2474:TAMAAS>2.0.CO;2](https://doi.org/10.1175/1520-0442(1999)012<2474:TAMAAS>2.0.CO;2).