

Precipitation Downscaling with Gibbs Sampling: An Improved Method for Producing Realistic, Weather-Dependent, and Anisotropic Fields

JOSEPH BELLIER AND MICHAEL SCHEUERER

Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, and NOAA/Physical Sciences Laboratory, Boulder, Colorado

THOMAS M. HAMILL

NOAA/Physical Sciences Laboratory, Boulder, Colorado

(Manuscript received 12 March 2020, in final form 2 July 2020)

ABSTRACT: Downscaling precipitation fields is a necessary step in a number of applications, especially in hydrological modeling where the meteorological forcings are frequently available at too coarse resolution. In this article, we review the Gibbs sampling disaggregation model (GSDM), a stochastic downscaling technique originally proposed by Gagnon et al. The method is capable of introducing realistic, weather-dependent, and possibly anisotropic fine-scale details, while preserving the mean rain rate over the coarse-scale pixels. The main developments compared to the former version are (i) an adapted Gibbs sampling algorithm that enforces the downscaled fields to have a similar texture to that of the analysis fields, (ii) an extensive test of various meteorological predictors for controlling specific aspects of the texture such as the anisotropy and the spatial variability, and (iii) a review of the regression equations used in the model for defining the conditional distributions. A perfect-model experiment is conducted over a domain in the southeastern United States. The metrics used for verification are based on the concept of gridded, stratified variogram, which is introduced as an effective way of reproducing the abilities of human eyes for detecting differences in the field texture. Results indicate that the best overall performances are obtained with the most sophisticated, predictor-based GSDM variant. The 600-hPa wind is found to be the best year-round predictor for controlling the anisotropy. For the spatial variability, kinematic predictors such as wind shear are found to be best during the convective periods, while instability indices are more informative elsewhere.

KEYWORDS: Precipitation; Statistical techniques; Forecast verification/skill; Stochastic models; Subgrid-scale processes

1. Introduction

Thanks to scientific and computational progress, it is nowadays possible to run high-resolution hydrological models on a continental scale (e.g., the National Water Model; Salas et al. 2018; Viterbo et al. 2020), using fine-scale meteorological analyses for calibration. In a forecasting context though, meteorological inputs at such resolution can only be produced by short-term, limited-area numerical weather prediction (NWP) models. For applications like medium-range forecasting or climate prediction, the meteorological forcings are most often available at a coarser resolution, thus calling for a procedure referred to as *downscaling*. When applied to the variable precipitation, the term *disaggregation* is sometimes used, with the idea of fragmenting the volume of rainfall from coarse to fine-scale pixels. Besides its nonlinear behavior, precipitation exhibits a high spatial variability that makes it one of the most challenging weather variables to downscale (Maraun et al. 2010). Meanwhile, this spatial variability has been shown to play a crucial role in the hydrologic response of a catchment. Several studies, when applying small-scale spatial perturbations to rainfall fields while keeping the total volume

unchanged, have reported large effects not only on peak flow (Rebora et al. 2006a; Younger et al. 2009) but also on the soil moisture and groundwater level (Schuurmans and Bierkens 2007). Precipitation downscaling is thus a critical step for hydrological modeling if the model resolution mismatches that of the meteorological forcing available.

Another motivation for the research presented in this article is related to ensemble forecast postprocessing. The U.S. National Weather Service (NWS) has launched the National Blend of Models (NBM) program, with the objective of providing to regional and national forecast centers a seamless and spatially consistent digital starting point over the entire United States. The NBM generates high-resolution, calibrated, probabilistic precipitation forecasts by statistically postprocessing multimodel ensembles (Hamill and Scheuerer 2018). However, for now these postprocessed distributions are available for every pixel and lead time independently, meaning that a forecaster cannot derive a probabilistic forecast of the precipitation accumulated over a larger area or a longer period of time. This would require the recreation of synthetic ensembles that integrate the spatial and temporal dependence of precipitation. The ensemble copula coupling (ECC; Schefzik et al. 2013) is an attractive method for that purpose, as it imposes the space–time structure of the raw (i.e., unprocessed) ensembles to the synthetic ensembles. However, ECC is not immediately applicable in the context of the NBM, because of a mismatch between the spatial resolution of the postprocessed distributions and that of the raw ensembles. The strategy envisaged to

Supplemental information related to this paper is available at the Journals Online website: <https://doi.org/10.1175/JHM-D-20-0069.s1>.

Corresponding author: Joseph Bellier, joseph.bellier@noaa.gov

DOI: 10.1175/JHM-D-20-0069.1

© 2020 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](#) (www.ametsoc.org/PUBSReuseLicenses).

address that issue is to downscale the raw ensemble precipitation fields prior to applying ECC. The present article focuses on the downscaling process only, as the outcomes may be useful in a wider range of applications. Its specific use within the NBM context will be the subject of a following paper.

Precipitation downscaling has been the subject of active research (for reviews see Fowler et al. 2007; Maraun et al. 2010). We restrict here to methods that are statistical (as opposed to dynamical, whose computational cost for running high-resolution atmospheric models can be prohibitively high), and which generate a gridded fine-scale field (as opposed to techniques that downscale to individual sites). Ideally, such methods should be able to (i) preserve the mean rain rate over each coarse-scale pixel (assuming that coarse-scale fields are bias corrected, which is beyond the scope of this paper), (ii) avoid any form of “blockiness” (i.e., visual discontinuities at the edges of the coarse-scale pixels), (iii) reproduce subgrid anisotropy, and (iv) account for repeatable local effects, such as orography.

The review proposed in Table 1, which does not pretend to be exhaustive, attempts to summarize existing methods regarding which properties (i)–(iv) they honor, and to categorize them by their physical assumption(s) the downscaling process is based on. A first group gathers methods that assume an invariance of some statistical characteristics of precipitation over a range of scales, a property referred to as *scaling*. This can, for instance, manifest as a log–log linearity of the power spectrum through scales (Rebora et al. 2006b), or a simple-scaling behavior of the rainfall fluctuations (Perica and Foufoula-Georgiou 1996b). Most often, the downscaling process is a multiplicative cascade, i.e., an iterative process that fragments the rainfall volume of the coarse-scale pixels into smaller and smaller pixels, using a random generator that respects a given scale-invariance property. While most scaling-based techniques are able to conserve the mean rain rate, for many it is at the expense of a strong blockiness (e.g., Ahrens 2003; Rebora et al. 2006a; Guan et al. 2009). To avoid this artifact, other methods try to characterize the *spatial* structure of precipitation directly at the fine scale of interest, by modeling the conditional distribution of precipitation at each fine-scale pixel based on the distribution at neighboring pixels, usually assuming Markov properties. Once such conditional distribution model is defined, full downscaled fields can be generated using sampling algorithms such as the Gibbs sampling (Allcroft and Glasbey 2003; Onibon et al. 2004; Gagnon et al. 2012). Alternatively, Ebtehaj et al. (2012) and Foufoula-Georgiou et al. (2014) express the desired spatial properties in a suitable space (e.g., wavelets, derivative) where they exhibit some form of sparsity, which can then be used to regularize (and solve) the inverse problem into which the authors recast downscaling. Overall, the main strength of spatial-based method is to generate spatially coherent fields that are free of unrealistic discontinuities. However, as such, not many can reproduce a subgrid anisotropy, the authors generally seeking to limit the complexity to an acceptable level (e.g., Allcroft and Glasbey 2003; Rebora et al. 2006b). Finally, we refer to the third main physical assumption as *explainability*, which assumes that the spatial variability of precipitation can be related to fine-scale

predictors (e.g., topographic, geographic, or atmospheric), by means of traditional regression methods (Guan et al. 2009; Park 2013) or more complex machine learning algorithms (Xu et al. 2015; He et al. 2016). The ability to account for repeatable local effects such as orography is their main strength, although a number of studies have shown that it is possible to modify scaling or spatial-based methods to also fulfill that need (Badas et al. 2006; Gagnon et al. 2013; Terzago et al. 2018).

Among the existing downscaling methods listed in Table 1, the Gibbs sampling disaggregation model (GSDM), introduced by Gagnon et al. (2012) and refined in Gagnon and Rousseau (2014) to include wind predictors, looks particularly appealing with regard to properties (i)–(iv). The approach consists in modeling the spatial structure of the rainfall fields through a conditional model of the precipitation at one pixel given the value of the neighboring pixels (and eventually meteorological predictors), without modeling the full joint distribution. The Gibbs sampling algorithm is then used to sample realizations of the full field, ensuring after each iteration that the mean rain rate at the coarse scale is preserved. While the original version does not account for local effects, Gagnon et al. (2013) has shown that it possible to introduce orographic effects. In this paper though, we will focus on the simplified case without orography, keeping this aspect for a future study.

Despite its capabilities, the GSDM as described in the above references has several limitations. In particular, it relies on the assumption that the spatial variability of rainfall can be fully described by the conditional distribution model, which in practice is hard to achieve while keeping an acceptable level of parsimony. As a consequence, the Gibbs sampling may converge toward a joint distribution whose spatial structure is unrealistic. In this paper, we address this issue with a so-called “incomplete” Gibbs sampling: instead of waiting for the algorithm to converge, we predefine a much smaller number of iterations, and calibrate the conditional distribution model such that the final downscaled fields have the desired *texture*. The benefice is twofold: a parsimonious model can be used without the risk of converging to unrealistic fields, and the number of iterations is drastically reduced, which makes the GSDM computationally affordable for use over very large areas. By modifying the algorithm though, the original regression models (for defining the conditional distributions) but also the meteorological predictors may not be optimal anymore. This study therefore tests a large variety of models and predictors, discuss their impacts on the downscaled fields, and performs a verification using different quantitative metrics of the rainfall texture.

The notion of texture, which is central in this revised GSDM, refers to the extent to which precipitation changes with distance and potentially with direction. To quantify differences in texture between two fields, we introduce new metrics based on the concept of gridded, stratified variograms. *Gridded* means measuring the variogram for field values separated by specific distances and directions, which permits taking into account a potential anisotropy, while *stratified* means discriminating the texture between low and high precipitation values.

TABLE 1. A summary of precipitation downscaling methods that have been proposed in the literature, restricted to those that are statistical, and which generate a gridded fine-scale field. This list does not pretend to be exhaustive. Columns (i)–(iv) refer to the desirable properties of a downscaling method (see text).

Name of method (if any) and reference	Main physical assumption(s)	Downscaling mechanism(s)	(i) Conserve coarse-scale volume	(ii) Free of blockiness	(iii) Reproduce subgrid anisotropy seamlessly	(iv) Account for repeatable local effects
Perica and Foufoula-Georgiou (1996b); Ahrens (2003)	Scaling	Multiplicative cascade	Yes	No ^a	No	No
STRAIN (Deidda 2000)	Scaling	Multiplicative cascade	In expectation	No	No	No ^b
RainFARM (Rebora et al. 2006b)	Scaling	Transformation of a GRF	Yes	No ^c	No	No ^d
HiDRUS (Raut et al. 2019)	Scaling	Multiplicative cascade + advection	No	Yes	Yes	Yes
Allcroft and Glasbey (2003)	Spatial	Block Gibbs sampling	Approximately	Yes	No	No
Onibon et al. (2004)	Spatial	Gibbs sampling	Only the entire domain average	Yes	No	No
VarD (Foufoula-Georgiou et al. 2014); SPaD (Ebtehaj et al. 2012)	Spatial	Regularization of the inverse problem	Yes	Yes	No	No
GSDM (Gagnon et al. 2012)	Spatial, explainability	Gibbs sampling	Yes	Yes	Yes	No ^e
ASCAR (Guan et al. 2009); Xu et al. (2015)	Explainability, scaling	Multivariate regression or ANN + multiplicative cascade	Yes	No	No	Yes
Prec-DWARF (He et al. 2016)	Explainability, spatial	Random forest	No	Yes	Yes	Yes
Park (2013)	Explainability, spatial	Multivariate regression + area-to-point kriging	Yes	Yes	No	Yes

^a Extension in Harris and Foufoula-Georgiou (2001) to reduce blockiness.

^b Extension in Badas et al. (2006) to account for orographic effects.

^c Extension in Terzago et al. (2018) to reduce blockiness.

^d Extension in Terzago et al. (2018) to account for orographic effects.

^e Extension in Gagnon et al. (2013) to account for orographic effects.

The remainder of this article is organized as follows. Section 2 presents the case study and data. Section 3 describes the GSDM algorithm, and details the different combinations of regression models and predictors. The gridded, stratified variogram for quantifying texture is then introduced in section 4. Section 5 provides results and discusses them, while section 6 concludes.

2. Case study and data

The area considered in this downscaling experiment is a $4^\circ \times 4^\circ$ square delineated by 86° – 90° W longitude and 31° – 35° N latitude, spanning the states of Mississippi and Alabama in the southeastern United States. Winter precipitation is mostly stratiform, with large-scale, steady rainfall features, whereas

summer precipitation is mostly convective, although the rainfall features can take place at various scales depending on the degree of convective organization. Spring and fall months see both stratiform and convective precipitation, depending on the synoptic situation. The region has been chosen for its flat topography, such that precipitation fields can be assumed free of orographic effects. The integration of these local effects within our reviewed GSDM is therefore not included in this paper. Finally, the domain size has been chosen as a trade-off between the requirement of having enough rainfall features in the downscaled fields for a robust assessment of texture, while being small enough for the assumption of stationarity to be valid.

The fine-scale precipitation data comes from the climatology calibrated precipitation analysis (CCPA; Hou et al. 2014), a calibrated version of the Stage IV gridded dataset (Lin and Mitchell 2005) that covers the contiguous United States at 4.7-km resolution and utilizes both gauge and radar data. For this study, CCPA has been upscaled on a regular grid at the resolution of $1/16^\circ$ (~ 7 km), which will be our “fine-scale resolution.” We consider the same accumulation period as in the NBM operational postprocessing, i.e., 6 h, which we express in local time (1800–0000, 0000–0600, 0600–1200, and 1200–1800 CST) for ease of physical interpretation.

The meteorological predictors used by the downscaling model come from the ERA5 reanalysis (Hersbach et al. 2020), produced by the European Centre for Medium-Range Weather Forecasts (ECMWF) and available at $1/4^\circ$ (~ 28 km) resolution. This will be our “coarse-scale resolution.” The list of all predictors tested will be discussed in section 3d. Those that are not directly available in ERA5 are computed with the Python module MetPy (May et al. 2019; UCAR 2019), using the temperature, geopotential, specific humidity, and U – V wind variables at all the pressure levels available in ERA5. The reported predictor values are the averages of the instantaneous values at the beginning and end of each 6-h period.

In the present downscaling experiment, referred to as a “perfect model” experiment, the 64×64 (pixels) precipitation analysis fields at $1/16^\circ$ resolution are aggregated to the $1/4^\circ$ resolution to form 16×16 coarse-scale fields, before downscaling them back to $1/16^\circ$ (scaling ratio of 4). It is important to note that this setup happens to be dyadic regarding both the domain size ($64 = 2^6$) and the scaling ratio ($4 = 2^2$), but this is by no means mandatory for applying the GSDM, nor for computing the verification metrics presented in this article.

The period considered is 2004–17, divided into two periods of 7 years each: the even years are used for calibration, and the odd years for validation. The model is calibrated separately for each month and each 6-h period, to account for the seasonal and diurnal variability of the precipitation process.

3. The Gibbs sampling disaggregation model

a. Notation

Let $(x, y) \in \mathcal{D}_{XY}$ and $(i, j) \in \mathcal{D}_{IJ}$ denote the locations in Cartesian coordinates of the coarse-scale and fine-scale pixels, respectively, where $\mathcal{D}_{XY} = \{(1, 1), \dots, (X, Y)\}$ and $\mathcal{D}_{IJ} = \{(1, 1), \dots, (I, J)\}$ represent the coarse- and fine-scale grids, respectively. In our case study previously described, $X = Y = 16$ while $I = J = 64$. We define the coarse-scale precipitation random field $\mathbf{C} = \{C_{(x,y)} : (x, y) \in \mathcal{D}_{XY}\}$, from which a realization is available from a NWP or a climate prediction model, and the fine-scale random field $\mathbf{R} = \{R_{(i,j)} : (i, j) \in \mathcal{D}_{IJ}\}$ from which a realization (i.e., a downscaled field) is desired. Let us further consider a fine-to-coarse coordinate mapping function $\mathcal{F} : \mathcal{D}_{IJ} \rightarrow \mathcal{D}_{XY}$, defined as $\mathcal{F}(i, j) = (x, y)$, which maps every fine-scale location to the coarse-scale location it belongs to. This function is to be defined by every user, depending on their own downscaling setup (scaling ratio,

overlapping of the fine- and coarse-scale grids, etc.). We also set $G_{(x,y)}$ for $(x, y) \in \mathcal{D}_{XY}$ as the number of locations (i, j) that satisfy $\mathcal{F}(i, j) = (x, y)$. Except maybe at boundaries of the downscaling domain, $G_{(x,y)}$ is likely to be constant for all (x, y) .

b. Markov assumption and Gibbs sampling

The downscaling problem can be expressed as finding a realization of \mathbf{R} that satisfies the constraint of conservation of the mean rain rate:

$$\frac{1}{G_{(x,y)}} \sum_{(i,j):\mathcal{F}(i,j)=(x,y)} R_{(i,j)} = C_{(x,y)}, \quad (1)$$

for every coarse-scale location $(x, y) \in \mathcal{D}_{XY}$. However, the problem as such is ill-defined, as there is an infinite ways of defining a joint distribution $p(\mathbf{R})$ that satisfies the constraint (1). The idea of the GSDM is to assume some form of Markov properties in the field \mathbf{R} , such that $p(R_{(i,j)} | \mathbf{R}_{-R_{(i,j)}})$, the conditional distribution of the pixel $R_{(i,j)}$ given all other pixels, can be approximated by $p(R_{(i,j)} | \mathbf{N}_{(i,j)})$, where $\mathbf{N}_{(i,j)}$ is the subset of \mathbf{R} containing only the neighboring pixels of $R_{(i,j)}$ (but not $R_{(i,j)}$ itself). The definition of the conditional distribution $p(R_{(i,j)} | \mathbf{N}_{(i,j)})$, which is the core of the GSDM, will be described in details in sections 3c and 3d. For now, assume that a statistical model for $p(R_{(i,j)} | \mathbf{N}_{(i,j)})$ exists, and let $\Gamma_{(i,j)}$ be the corresponding cumulative distribution function (CDF).

The procedure for obtaining a realization of the complete field \mathbf{R} is based on the Gibbs sampling, an iterative algorithm that allows to sample a realization of a process whose joint distribution is unknown but the conditional distributions of every element are known [for a good introduction, see Casella and George (1992)]. Of course, none of the elements in $\mathbf{N}_{(i,j)}$ are known initially, but the Gibbs sampling iteratively updates every pixel of the field (leaving the others unchanged) such that, when it is time to update a given pixel $R_{(i,j)}$, its neighborhood $\mathbf{N}_{(i,j)}$ is “realistic.” The algorithm, as proposed by Gagnon et al. (2012), can be described as follows:

Step 0. Iterate over all locations $(i, j) \in \mathcal{D}_{IJ}$ and initialize $R_{(i,j)}$ with the value of the coarse-scale pixel it belongs to:

$$R_{(i,j)} \leftarrow C_{\mathcal{F}(i,j)}.$$

Step 1. Iterate over all locations $(i, j) \in \mathcal{D}_{IJ}$ and update $R_{(i,j)}$ by sampling a value from its conditional distribution:

$$R_{(i,j)} \leftarrow \Gamma_{(i,j)}^{-1}(u)$$

where u is randomly drawn from $\mathcal{U}(0, 1)$.

Step 2. Iterate over all locations $(x, y) \in \mathcal{D}_{XY}$ and compute the coarse-scale multiplicative coefficients $\alpha_{(x,y)}$ defined as

$$\alpha_{(x,y)} = \left(\frac{1}{G_{(x,y)}} \sum_{(i,j):\mathcal{F}(i,j)=(x,y)} R_{(i,j)} \right) / C_{(x,y)}.$$

Once the $X \times Y$ coefficients $\alpha_{(x,y)}$ are stored, iterate over all locations $(i, j) \in \mathcal{D}_{IJ}$ and adjust $R_{(i,j)}$ such that

the constraint of mean rain rate conservation expressed in (1) is satisfied:

$$R_{(i,j)} \leftarrow \alpha_{\mathcal{F}(i,j)} R_{(i,j)}.$$

Step 3. Repeat N_{iter} times the steps 1–2. After the last iteration, extract the final downscaled field $\mathbf{R} = \{R_{(i,j)}; (i, j) \in \mathcal{D}_D\}$, replacing by zero all $R_{(i,j)}$ that are smaller than a certain threshold, e.g., 0.1 mm per 6 h.

An important point in the above algorithm is, for a downscaling purpose, the possibility to introduce the adjustment for conservation of the mean rain rate (step 2) after every Gibbs sampling iteration. Thanks to this recurrence, and because the distributions $p(R_{(i,j)}|\mathbf{N}_{(i,j)})$ are on average centered around the mean of $\mathbf{N}_{(i,j)}$ (see section 3c), the multiplicative coefficients $\alpha_{(x,y)}$ are typically close to 1, such that the effect on the spatial structure is negligible. In numerous downscaling methods proposed in the literature, this volume adjustment can only be performed at the very end of the process, often with $\alpha_{(x,y)}$ substantially different from 1, which causes a noticeable blockiness.

According to the theory of the Gibbs sampling, if N_{iter} is large enough, and assuming that the effect of step 2 is negligible on the spatial structure, the final field \mathbf{R} should be a statistical sample of the joint distribution $p(\mathbf{R})$. However, tracking the convergence of the Gibbs sampler is not straightforward. Although Gagnon and Rousseau (2014) recommend $N_{\text{iter}} = 300$, there appears to be no good “rule of thumb” for fixing a number of iterations that is appropriate to every context (Raftery and Lewis 1992). A second difficulty lies in defining and calibrating a model for $p(R_{(i,j)}|\mathbf{N}_{(i,j)})$ that accounts for all sources of spatial variability such that the implicit $p(\mathbf{R})$, to which the Gibbs sampling converges, has the desired spatial properties. Gagnon et al. (2012) recommend to fit $p(R_{(i,j)}|\mathbf{N}_{(i,j)})$ to every nonzero pixel $R_{(i,j)}$ in the set of fine-scale training analyses, but considering the neighboring pixels $\mathbf{N}_{(i,j)}$ as known (i.e., no Gibbs sampling). However, any misrepresentation of a spatial characteristic in this conditional distribution model, albeit small, can then be amplified to the point that the final field \mathbf{R} becomes strongly unrealistic. An example is given in Fig. S1 in the online supplemental material, where the Gibbs sampling converges to a field exhibiting anisotropic “stripes” that are unrealistically sharp and spaced, although the model for $p(R_{(i,j)}|\mathbf{N}_{(i,j)})$ has been properly fitted to the training data.

To circumvent these issues, we propose a novel use of the algorithm described above, where instead of waiting for the Gibbs sampling to converge, we predefine a relatively small value for N_{iter} , and calibrate the model for $p(R_{(i,j)}|\mathbf{N}_{(i,j)})$ such that the field \mathbf{R} after N_{iter} iterations has the desired texture. We acknowledge that proceeding this way is not fully in line with the Gibbs sampling theory, in that convergence to the implicitly defined joint distribution model is no longer the main objective. However, our more pragmatic approach, of using the Gibbs sampling simply to generate subgridscale variability that is informed by meteorological predictors and matches observed spatial structures, has two major benefits. First, the risk of generating strongly unrealistic fields is

greatly reduced, as the parameters of the model control the texture of the final fields (i.e., after the Gibbs sampling). Second, the computational cost for running the algorithm, which is directly proportional to N_{iter} , is drastically reduced. For this study, satisfying results were found with $N_{\text{iter}} = 10$, which is at least an order of magnitude smaller than what the algorithm would have required to converge. The performance will eventually deteriorate for smaller N_{iter} , however it was not found to improve much for higher N_{iter} , whereas the computational cost increases. Note that users with a different downscaling setup, e.g., a higher scaling ratio, may require more iterations for a realistic spatial structure to establish, hence we recommend a trial-and-error process for fixing a suitable N_{iter} value.

The calibration procedure of this revised GSDM thus consists in matching the full downscaled fields (obtained after N_{iter} iterations of the Gibbs sampling) with their verifying analysis, and search for the set of parameters of the model for $p(R_{(i,j)}|\mathbf{N}_{(i,j)})$ that minimizes the averaged difference in texture between the two. Denoting the sets of downscaled and analysis fields by $\{\mathbf{R}_1, \dots, \mathbf{R}_M\}$ and $\{\mathbf{R}_1^{\text{obs}}, \dots, \mathbf{R}_M^{\text{obs}}\}$, respectively, with M the number of fields in the calibration period, the actual quantity being minimized, referred to as the *cost function*, can be defined as

$$\overline{\mathcal{L}} = \frac{1}{M} \sum_{m=1}^M \mathcal{L}(\mathbf{R}_m, \mathbf{R}_m^{\text{obs}}), \quad (2)$$

where \mathcal{L} is a *loss function* that quantifies the difference in texture between two rainfall fields. This loss function can be specific to each user, depending on which aspects of the texture they are most interested in reproducing. We propose one in the section 4 of this article, based on the concept of gridded, stratified variogram. It is important to note that, because of the stochasticity of the Gibbs sampling, $\overline{\mathcal{L}}$ is itself a random quantity, which requires specific precautions for its minimization. These are discussed in the appendix. Moreover, since the entire set $\{\mathbf{R}_1, \dots, \mathbf{R}_M\}$ is required for a single evaluation of $\overline{\mathcal{L}}$, it is strongly recommended, for the calibration procedure to remain affordable, to optimize the Gibbs sampling code for fast computation. For this study, the algorithm was coded in Python but optimized using the just-in-time compiler Numba (Lam et al. 2015).

c. Standard model for $p(R_{(i,j)}|\mathbf{N}_{(i,j)})$

Now that the Gibbs sampling algorithm has been described, we detail the methodology for modeling the conditional distribution $p(R_{(i,j)}|\mathbf{N}_{(i,j)})$. We largely follow the convenient statistical framework proposed by Gagnon et al. (2012), although a number of modifications are suggested to address some limitations in the original model.

The general idea is to assume that the nonzero values of $R_{(i,j)}$ follow a parametric distribution with expectation $E_{(i,j)}$ and standard deviation $\text{SD}_{(i,j)}$, which will be expressed as a function of the neighboring pixels and/or meteorological predictors via two regression models. The lognormal distribution is assumed here, and its parameters $\mu_{(i,j)}$ and $\sigma_{(i,j)}$ (i.e., the mean and standard deviation of the variable’s logarithm) are related to $E_{(i,j)}$ and $\text{SD}_{(i,j)}$ via

$$\mu_{(ij)} = -\frac{1}{2} \ln \left(\frac{E_{(ij)}^2 + \text{SD}_{(ij)}^2}{E_{(ij)}^4} \right), \quad (3)$$

$$\sigma_{(ij)} = \sqrt{\ln \left(1 + \frac{\text{SD}_{(ij)}^2}{E_{(ij)}^2} \right)}. \quad (4)$$

Note that while the lognormal distribution is frequently used for modeling nonzero precipitation data, the Gamma distribution is a relevant alternative too, which has been found by Gagnon et al. (2012) to give similar results. If the latter (or any other distribution) is to be chosen, Eqs. (3) and (4) must be adapted accordingly, but the rest of the methodology remains unchanged. The main step then consists in defining regression models for $E_{(ij)}$ and $\text{SD}_{(ij)}$. This section describes a “standard” version, where regression models do not involve any meteorological predictors yet.

Based on the assumption of Markov properties in the rainfall field, the expectation $E_{(ij)}$ can be expressed as a linear combination of the neighboring pixels. As a trade-off between parsimony and ability to account for anisotropy, the neighborhood is restricted to the eight nearest pixels. Since north-to-south and south-to-north anisotropies are indistinguishable when looking at a rainfall field, and likewise for the other directions, the information in the eight neighboring pixels can be reduced to four directional averages defined as $A_{(ij)}^{\uparrow} = (R_{(i,j-1)} + R_{(i,j+1)})/2$, $A_{(ij)}^{\downarrow} = (R_{(i-1,j)} + R_{(i+1,j)})/2$, $A_{(ij)}^{\leftarrow} = (R_{(i-1,j-1)} + R_{(i+1,j+1)})/2$, and $A_{(ij)}^{\rightarrow} = (R_{(i-1,j+1)} + R_{(i+1,j-1)})/2$, with $\bar{A}_{(ij)}$ the average of the four. Cases where (i, j) is at the boundary of the domain are dealt with field mirroring. The authors have then proposed a formulation that expresses $E_{(ij)}$ as a linear combination of these quantities, in a way that the different terms are easily interpretable:

$$E_{(ij)} = \bar{A}_{(ij)} + \beta_d \left(\frac{A_{(ij)}^{\uparrow} + A_{(ij)}^{\downarrow}}{2} - \frac{A_{(ij)}^{\leftarrow} + A_{(ij)}^{\rightarrow}}{2} \right) + \beta_{\times} (A_{(ij)}^{\leftarrow} - A_{(ij)}^{\rightarrow}) + \beta_{+} (A_{(ij)}^{\uparrow} - A_{(ij)}^{\downarrow}). \quad (5)$$

The first term $\bar{A}_{(ij)}$ has a coefficient fixed to 1, such that $E_{(ij)}$ stochastically tends to the neighborhood mean and thus ensures that multiplicative coefficients $\alpha_{(x,y)}$ in the Gibbs sampling remain close to 1 (see section 3b). The second term, referred to as the distance term, aims at increasing the correlation between $R_{(ij)}$ and its 4 nearest neighbors (left, right, above, and below) with respect to the correlation with the four others (in the diagonals). The coefficient β_d is thus expected to have a positive value. Finally, the third and fourth terms account for potential anisotropy. Assuming (here and throughout the article) that 0° points to the east, the coefficient β_{\times} (β_{+}) should be positive if the anisotropy is stronger in the 45° (90°) than in the -45° (0°) direction. The combination of the two parameters β_{\times} and β_{+} thus controls simultaneously the direction and strength of the anisotropy.

A second regression model must then be defined for the standard deviation $\text{SD}_{(ij)}$, which is tightly related to the spatial variability of the downscaled field. If $\text{SD}_{(ij)}$ is small, the probability of sampling a new value of $R_{(ij)}$ that is much different from

$E_{(ij)}$ will be small, as the conditional distribution $p(R_{(ij)}|\mathbf{N}_{(ij)})$ is tight around $E_{(ij)}$. This will enforce a strong spatial dependence, typically seen in winter. In summer, $\text{SD}_{(ij)}$ is expected to be larger, increasing the spatial variability. Apart from the seasonal variability, it is further assumed that $\text{SD}_{(ij)}$ increases with $E_{(ij)}$. We here propose the two-parameter linear model

$$\text{SD}_{(ij)} = \beta_{s_1} + \beta_{s_2} E_{(ij)}, \quad (6)$$

which has been found to perform slightly better than the nonlinear relationship $\text{SD}_{(ij)} = \beta_{s_1} E_{(ij)}^{\beta_{s_2}}$ suggested by the original authors.

The regression model for $E_{(ij)}$ formulated in Eq. (5) is referred to as E30, where the first digit indicates the number of free parameters to estimate (β_d , β_{\times} , and β_{+} here), while the second digit indicates the number of meteorological predictors (none here). Similarly, the model for $\text{SD}_{(ij)}$ in Eq. (6) can be referred to as S20. Alternative models will be proposed in the next section, keeping the same terminology. The association of two regression models, one for $E_{(ij)}$ and one for $\text{SD}_{(ij)}$, will be referred hereinafter to as a GSDM variant. The one described above, namely, E30–S20, has been found to be the most skillful variant that does not rely on any meteorological predictors.

d. Alternative models for $E_{(ij)}$ and $\text{SD}_{(ij)}$

The equations of the different alternative models of E30 and S20 are given in Tables 2 and 3, respectively. For a better grasp of how the models differ from one another, Fig. 1 illustrates, for a given validation date, the downscaled fields obtained with all possible combinations of models.

For the expectation $E_{(ij)}$, models E00 and E10 are introduced as direct simplifications of E30, with the objective of assessing the importance of the distance and the anisotropy terms. In Fig. 1, by comparing E10 to E00 the impact of the distance term is limited but yet detectable, with rainfall features being slightly more structured. However, by comparing E30 to E10 the effect of the anisotropy terms is striking. The fields now show distinct anisotropic features oriented north-east, which is the preferential direction in this area for that period of the year. With model E30, it is indeed expected that all downscaled fields in validation reproduce the dominant anisotropy shape observed in calibration (in other words, the “climatological anisotropy”). However, the date selected in Fig. 1 is an example where the anisotropy departs from its climatology, with features in the original analysis field that are oriented northwest. This misrepresentation could have dramatic effects on hydrological modeling, for instance if a catchment is oriented in the same direction—or not—as the rainfall features.

To address that issue, models E21 and E32 introduce meteorological predictors. First, it should be noted that anisotropic features in a precipitation field can have two origins: the spatial structure of instantaneous rainfall cells (e.g., a squall line), and their advection through time. For the 6-h accumulation period considered in this study, the advection prevails, making any horizontal atmosphere motion vector a good predictor candidate, with its direction controlling the anisotropy

TABLE 2. List of all regression models that are tested for the expectation $E_{(i,j)}$.

Name	Regression model
E00	$E_{(i,j)} = \bar{A}_{(i,j)}$
E10	$E_{(i,j)} = \bar{A}_{(i,j)} + \beta_d \left[\left(\frac{A_{(i,j)}^{\downarrow} + A_{(i,j)}^{\uparrow}}{2} \right) - \left(\frac{A_{(i,j)}^{\downarrow}}{2} + \frac{A_{(i,j)}^{\uparrow}}{2} \right) \right]$
E30	$E_{(i,j)} = \bar{A}_{(i,j)} + \beta_d \left[\left(\frac{A_{(i,j)}^{\downarrow} + A_{(i,j)}^{\uparrow}}{2} \right) - \left(\frac{A_{(i,j)}^{\downarrow}}{2} + \frac{A_{(i,j)}^{\uparrow}}{2} \right) \right] + \beta_{\times}(A_{(i,j)}^{\downarrow} - A_{(i,j)}^{\uparrow}) + \beta_{+}(A_{(i,j)}^{\downarrow} - A_{(i,j)}^{\uparrow})$
E21	$E_{(i,j)} = \bar{A}_{(i,j)} + \beta_d \left[\left(\frac{A_{(i,j)}^{\downarrow} + A_{(i,j)}^{\uparrow}}{2} \right) - \left(\frac{A_{(i,j)}^{\downarrow}}{2} + \frac{A_{(i,j)}^{\uparrow}}{2} \right) \right] + \beta_a (\cos[2(P_{(i,j)}^{\text{AD}} - 45^\circ)](A_{(i,j)}^{\downarrow} - A_{(i,j)}^{\uparrow}) + \cos[2(P_{(i,j)}^{\text{AD}} - 90^\circ)](A_{(i,j)}^{\downarrow} - A_{(i,j)}^{\uparrow}))$
E32	$E_{(i,j)} = \bar{A}_{(i,j)} + \beta_d \left[\left(\frac{A_{(i,j)}^{\downarrow} + A_{(i,j)}^{\uparrow}}{2} \right) - \left(\frac{A_{(i,j)}^{\downarrow}}{2} + \frac{A_{(i,j)}^{\uparrow}}{2} \right) \right] + (\beta_{a_1} + \beta_{a_2} P_{(i,j)}^{\text{AS}})(\cos[2(P_{(i,j)}^{\text{AD}} - 45^\circ)](A_{(i,j)}^{\downarrow} - A_{(i,j)}^{\uparrow}) + \cos[2(P_{(i,j)}^{\text{AD}} - 90^\circ)](A_{(i,j)}^{\downarrow} - A_{(i,j)}^{\uparrow}))$

direction while its magnitude controls the anisotropy strength. Gagnon and Rousseau (2014) have used the wind vector at the 700-hPa pressure level. Here, we consider as candidates the wind at 100, 200, 300, 400, 500, 600, 700, 775, 850, and 925 hPa, the average wind in the lowest 6-km layer above ground level (AGL), and finally the storm motion (Bunkers et al. 2000), assuming right-movers only. These predictor fields are bilinearly interpolated to the fine scale. In the equations in Table 2, $P_{(i,j)}^{\text{AD}}$ and $P_{(i,j)}^{\text{AS}}$ denote respectively the direction and the magnitude, at location (i, j) , of the predictor vector being used. The model E21 includes $P_{(i,j)}^{\text{AD}}$ for controlling the anisotropy direction, but relies on a free parameter β_a for the strength. The most complex model, E32, uses both $P_{(i,j)}^{\text{AD}}$ and $P_{(i,j)}^{\text{AS}}$. It is very similar to the model proposed in Gagnon and Rousseau (2014), except that an additional coefficient β_{a_1} has been introduced such that the model can reduce to E21 in case $P_{(i,j)}^{\text{AS}}$ is not informative. For the specific date shown in Fig. 1, the usefulness of E21 compared to E30 is obvious, with anisotropic features much more similar to the observed ones. However, the difference between E21 and E32 is hard to notice on this isolated example, as it requires a longer period to assess the informativeness of $P_{(i,j)}^{\text{AS}}$.

Concerning the standard deviation $\text{SD}_{(i,j)}$, the alternative model S10 is a direct simplification of S20, with the objective of assessing the importance of the dependence of $\text{SD}_{(i,j)}$ to $E_{(i,j)}$. In Fig. 1, this dependence term seems to render more realistic the graininess of the fields. The more complex model S31 then introduces a predictor, denoted $P_{(i,j)}^{\text{SV}}$, for controlling the spatial variability. The convective available potential energy (CAPE), introduced in the GSDM by Gagnon et al. (2012), may at first glance seem a pertinent predictor. Perica and Fofoula-Georgiou (1996a) also found a strong relationship between CAPE and the scaling parameters of their statistical model for describing the rainfall spatial variability. However, their relation was established only for a sample of “pre-storm” values of CAPE (i.e., that are not contaminated yet by convection), although CAPE is susceptible to drop rapidly after convection has started. In our downscaling

context, ensuring that all CAPE values used as predictor reflect the environment prior to storms seems unrealistic; hence a series of alternative predictors are tested, which are categorized into two groups.

The first group contains thermodynamic indices of the atmosphere instability: CAPE and convective inhibition (CIN), as computed by the ECMWF model (it is actually the most unstable CAPE, and the minimum CIN; see Tsonevsky et al. 2018), but also the Showalter index (SI; Showalter 1947), lifted index (LI; Galway 1956), and total totals index (TT; Miller 1975). The hypothesis is that the atmosphere instability is associated with a risk of convection and then a higher spatial variability of rainfall, hence we assume a positive relationship between $\text{SD}_{(i,j)}$ and the above indices, except SI and LI for which a negative relationship is assumed (lower values are associated with higher instability). The second group gathers kinematic predictors: storm relative helicity (SRH; Markowski and Richardson 2011) in the lowest 1- and 3-km layers AGL; the bulk Richardson number (BRN) shear; the vertical wind shear in the following layers: surface (sfc)–925, sfc–850, sfc–700, sfc–500, sfc–300, sfc–200, 850–700, 850–500, 850–300, 850–200, 700–500, 700–300, and 700–200 hPa; and finally, the wind speed predictors already used as candidates for the anisotropy strength. For all these kinematic predictors, a negative relationship with $\text{SD}_{(i,j)}$ is assumed. For instance, a strongly sheared environment may be associated with

TABLE 3. List of all regression models that are tested for the standard deviation $\text{SD}_{(i,j)}$.

Name	Regression model
S10	$\text{SD}_{(i,j)} = \beta_s$
S20	$\text{SD}_{(i,j)} = \beta_{s_1} + \beta_{s_2} E_{(i,j)}$
S31	$\text{SD}_{(i,j)} = \beta_{s_1} + \beta_{s_2} P_{(i,j)}^{\text{SV}} + \beta_{s_3} E_{(i,j)}$ (positive ^a)
	$\text{SD}_{(i,j)} = \beta_{s_1} \exp\left(-\frac{\beta_{s_2} P_{(i,j)}^{\text{SV}}}{\beta_{s_1}}\right) + \beta_{s_3} E_{(i,j)}$ (negative ^a)

^a Refers to the assumed relationship between $\text{SD}_{(i,j)}$ and $P_{(i,j)}^{\text{SV}}$; see text.

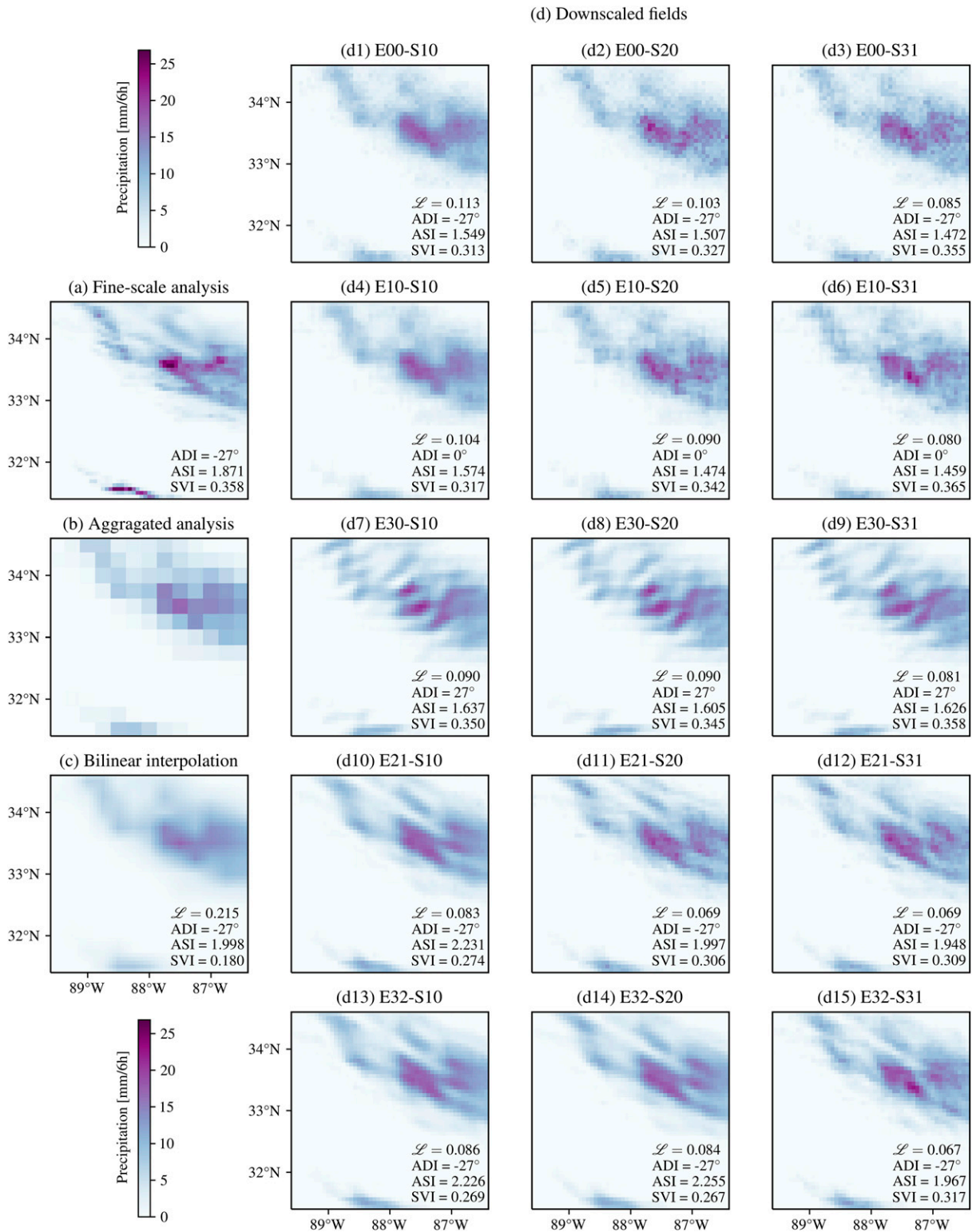


FIG. 1. (a) Example of a precipitation analysis field (1 Apr 2015, 1200–1800 CST). (b) Corresponding coarse-scale field after aggregation. (c) Result of the bilinear interpolation. (d1)–(d15) Results of the downscaling with all possible variants of the GSDM. The different regression models for $E^{(i,j)}$ are in rows, while the different regression models for $SD^{(i,j)}$ are in columns. For all downscaled fields, the difference \mathcal{L} in texture with the analysis field is reported, as well as the values of the indices ADI, ASI, and SVI.

stratiform precipitation within a frontal activity, but also with strongly organized convection. In both cases, the spatial variability of rainfall is expected to be lower, compared to weakly sheared environments where thunderstorms can develop but die out quickly (the falling rain killing off the vertical updraft), leading to spatially scattered rainfall features.

Therefore, the variety of candidate predictors for $P_{(i,j)}^{SV}$ requires two forms of model S31 (Table 3). The “positive” model assumes an increasing linear relationship (nonlinear models were not found to perform better), while the “negative” model is constructed such that $SD_{(i,j)}$ decreases asymptotically with $P_{(i,j)}^{SV}$ but remains positive. Note that Fig. 1 does not show much on the difference between S31 and S20, as a longer period is required to assess the informativeness of $P_{(i,j)}^{SV}$.

4. Quantifying the texture of a rainfall field

We now describe the concept of a gridded, stratified variogram, as a way of quantifying the texture of a rainfall field. This will be used for defining the loss function \mathcal{L} , but also indices of specific aspects of the texture.

a. The choice of the mathematical tool

With stochastic downscaling methods, like the GSDM, the spatial position of the fine-scale details is somewhat random, meaning that a pixel-by-pixel comparison is inadequate when confronting the downscaled fields to their verifying analysis. Instead, it is expected that the downscaled fields reproduce the probabilistic spatial structure of these details, a notion that we can refer to as texture. In the field of spatial verification (Gilleland et al. 2010), different approaches have been employed to quantify the texture of a precipitation field. The most intuitive mathematical tool, the spatial autocorrelation function, has been used in the context of downscaling by Ferraris et al. (2003) and Raut et al. (2018), among others. Over a finite domain, the autocorrelation can be expressed as a function of the distance (the “lag”), and eventually the direction, only under the condition of second-order stationarity. Marzban and Sandgathe (2009) advocate for using the variogram instead, as it requires a weaker assumption called intrinsic stationarity. The variogram has been used for precipitation fields by, among other, Achberger et al. (2003), He et al. (2016), and Ekström (2016). Finally, a third mathematical tool for texture analysis has been subject of active research in the recent years (see Weniger et al. 2017): the redundant discrete wavelet transform (RDWT), and more specifically its energy spectrum, which requires an even weaker assumption called local stationarity. Following the theory provided by Eckley et al. (2010), the approach has been used by Weniger et al. (2017), Kapp et al. (2018), Buschow et al. (2019), and Brune et al. (2018). Despite a number of appealing properties, the wavelet approach suffers from a number of limitations. First, the application of the RDWT implies the downscaling domain to be dyadic, as well as to deal with the boundary conditions. Moreover, the power spectrum accounts for anisotropy only to a certain extent, quantifying the energy in the vertical, horizontal, and diagonal directions only

(with up and down diagonal indistinguishable). Finally, a necessary bias correction of the spectrum, due to the redundancy of the wavelet transform, allows the energy in the smallest scales to be negative. Because of the lack of physical interpretation, some of the above-cited references simply discard the smallest scales. In our context, however, the fine-scale details of the downscaled fields are of prime interest, so ignoring the smallest scales is not an option. For all these reasons, we have found the variogram, which we briefly introduce below, to be the best suited mathematical tool for quantifying the texture of fine-scale precipitation fields.

Consider $\mathbf{h} \in \mathbb{R}^2$ any Euclidean vector with $\|\mathbf{h}\| \in [h - \delta, h + \delta]$, where h is a given distance (spatial lag) and δ a fixed tolerance. We define $S_h = \{(i, j) : R_{(i,j)} > 0, R_{(i,j)+\mathbf{h}} > 0, \|\mathbf{h}\| \in [h - \delta, h + \delta]\}$ the set containing all locations (i, j) such that both $R_{(i,j)}$ and $R_{(i,j)+\mathbf{h}}$ exist and are positive, with $\#S_h$ being the number of elements in that set. The function

$$\frac{1}{2\#S_h} \sum_{(i,j) \in S_h} (R_{(i,j)} - R_{(i,j)+\mathbf{h}})^2$$

is known as the empirical, omnidirectional, semivariogram (called simply *variogram* hereinafter) of the field \mathbf{R} . It depends only on the lag h under the assumption of intrinsic stationarity and isotropy of \mathbf{R} . The restriction in S_h to wet pixels aims at evaluating the texture independently of the location of the rainfall features, as recommended by Marzban and Sandgathe (2009).

b. The gridded, stratified variogram

Several suggestions are proposed to generalize the above formulation. To help the understanding, Fig. 2 provides an illustration based on a concrete example. First, we introduce a prior power transformation, to obtain the transformed field \mathbf{R} containing the pixels $\mathbf{R}_{(i,j)} = R_{(i,j)}^\lambda$ for $(i, j) \in \mathcal{D}_D$ (Fig. 2a). Computing the variogram on transformed fields with a power $\lambda < 1$ is a common practice for precipitation (e.g., Schuurmans et al. 2007; Erdin et al. 2012), as a way to reduce the skewness of the variable and thus the strong sensitivity to cases with large precipitation amounts. We also use a variant of the variogram where the squared differences are replaced by absolute differences, thereby reducing the influence of a few large pairwise differences in the fields, and thus increasing robustness to outliers (Cressie and Hawkins 1980). While this variant is sometimes referred to as *madogram*, we will keep using the more general term *variogram* throughout the article. Then, to account for anisotropy we propose a gridded computation of the variogram, replacing the vector \mathbf{h} by $(\Delta i, \Delta j)$ for $\Delta i, \Delta j = (-L, \dots, L)$, where $L \in \mathbb{N}$ is called the window size. Every combination of Δi and Δj thus accounts for a specific lag and direction (Fig. 2b). Finally, while Marzban and Sandgathe (2009) recommend to restrict to pairs of wet pixels, we expand the approach to quantifying the variogram for different ranges (called “strata”) of precipitation, where $K \in \mathbb{N}$ is the number of strata. The rationale is that precipitation at location (i, j) , when plotted against precipitation at another location $(i + \Delta i, j + \Delta j)$,

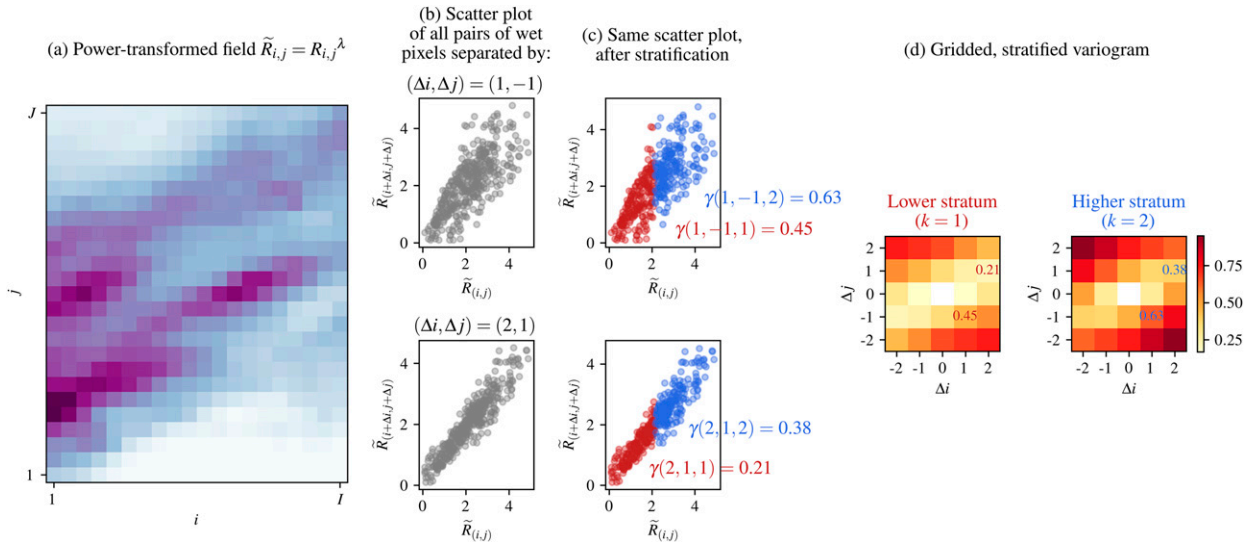


FIG. 2. Illustration of the construction of the gridded, stratified variogram for a given fine-scale precipitation field \mathbf{R} , with parameters $\lambda = 0.5$, $K = 2$, $L = 2$. (a) Power-transformed field $\tilde{\mathbf{R}}$. (b),(c) Unstratified and stratified scatterplots of all pairs of wet pixels separated by $(\Delta i, \Delta j) = (1, -1)$ in the top panels and $(\Delta i, \Delta j) = (2, 1)$ in the bottom panels. These are only two examples among all $(\Delta i, \Delta j)$ for $\Delta i, \Delta j \in (1, \dots, L)$. (d) The gridded, stratified variogram. The values displayed as text correspond to the two vectors $(\Delta i, \Delta j)$ taken as example.

may show heteroscedasticity. The nonstratified variogram, by summing over all pairwise differences, reduces the spatial dependence to a single number, and thus loses the information about heteroscedasticity. With stratification, the degree of heteroscedasticity is reported by the extent to which the variogram values differ between strata (Fig. 2c).

Now that all concepts have been introduced, we define the gridded, stratified variogram of the transformed field $\tilde{\mathbf{R}}$ as the quantity

$$\gamma_{\tilde{\mathbf{R}}}(\Delta i, \Delta j, k) = \frac{1}{2\#S_{\Delta i, \Delta j, k}} \sum_{(i,j) \in S_{\Delta i, \Delta j, k}} \left| \tilde{\mathbf{R}}_{(i,j)} - \tilde{\mathbf{R}}_{(i+\Delta i, j+\Delta j)} \right| \quad (7)$$

for $\Delta i, \Delta j = (-L, \dots, L)$ and $k = (1, \dots, K)$, where

$$S_{\Delta i, \Delta j, k} = \left\{ (i, j) : \hat{q}_{(k-1)/K} < R_{(i,j)} \leq \hat{q}_{k/K}, R_{(i+\Delta i, j+\Delta j)} > 0 \right\}, \quad (8)$$

with $\hat{q}_{(k-1)/K}$ and $\hat{q}_{k/K}$ respectively denoting the $(k-1)$ th and k th empirical K quantile of the population of nonzero values in \mathbf{R} . In the definition (8), the range of value is imposed on $R_{(i,j)}$ only, such that the stratification of the pairwise differences is mutually exclusive but collectively exhaustive. The final gridded stratified variogram is illustrated in Fig. 2d.

c. Application to the loss function \mathcal{L}

Based on the definition (7), we can define a loss function \mathcal{L} as the mean of the absolute errors between the γ values:

$$\mathcal{L}(\mathbf{R}, \mathbf{R}^{\text{obs}}) = \frac{1}{K(2L+1)^2} \sum_{k=1}^K \sum_{\Delta i=-L}^L \sum_{\Delta j=-L}^L \left| \gamma_{\tilde{\mathbf{R}}}(\Delta i, \Delta j, k) - \gamma_{\mathbf{R}^{\text{obs}}}(\Delta i, \Delta j, k) \right|, \quad (9)$$

where \mathbf{R} and \mathbf{R}^{obs} are a downscaled field and its corresponding verifying analysis, respectively. This function thus quantifies the overall difference in texture between the two fields, independently of the spatial position of the fine-scale features. Introducing (9) into (2) we obtain the cost function \mathcal{J} being minimized through the calibration process, which averages \mathcal{L} over all pairs of downscaled and analysis fields in the calibration set. In section 5, the same cost function \mathcal{J} will be used as a verification metric, and in such case it is computed over the validation set.

Note that for any $k = (1, \dots, K)$, we have $\gamma_{\tilde{\mathbf{R}}}(0, 0, k) = 0$, but also due to the symmetry of the anisotropy we can reasonably assume $\gamma_{\tilde{\mathbf{R}}}(\Delta i, \Delta j, k) \simeq \gamma_{\tilde{\mathbf{R}}}(-\Delta i, -\Delta j, k)$ (the strict equality holding only for infinitely large domains because of the stratification). Consequently, the computation of \mathcal{L} can be sped up by only averaging over $K(2L^2 + 2L)$ γ values instead of $K(2L+1)^2$, which is important for calibration purpose.

The loss function \mathcal{L} has three tuning parameters, λ , K , and L , which must be determined through trial and error. This involves a part of subjective judgment, with the objective of reproducing in the function the ability of human eyes for detecting differences of texture. We have found $\lambda = 0.5$ (square root transformation), $K = 3$, and $L = 1$ leading to the most realistic downscaled fields. Considering such a short window size L may seem counterintuitive because only short lags are evaluated, but on the other hand all the effort in calibration is then put into matching the variogram for these short lags, which are known in Markovian fields to have the strongest link with the texture. Furthermore, longer lags will eventually be fully constrained by the coarse-scale conditioning, hence their evaluation within \mathcal{L} becomes unnecessary. As a practical consideration, setting $L = 1$ also reduces to its minimum the

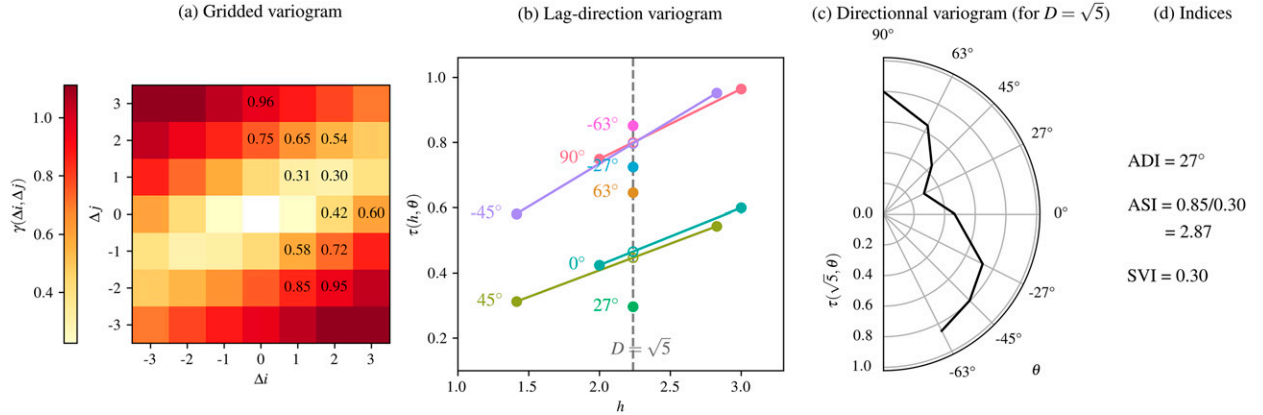


FIG. 3. Illustration of the construction of the indices ASI, ADI, and SVI, for the same field as in Fig. 2, with parameter $D = \sqrt{5}$. (a) Gridded (unstratified) variogram, where only the values displayed as text are actually necessary. (b) Equivalent “lag-direction” variogram, where filled circles represent the values directly available from the gridded variogram, while empty circles represent interpolated values at lag $h = D$. (c) Corresponding directional variogram, as a plot of the variogram values at lag $h = D$ only. (d) Values of the indices.

number of γ values that need to be computed for evaluating \mathcal{L} , which directly affects the computational cost of calibration. Finally, note that the function \mathcal{L} is not fully differentiable because of the use of absolute differences, but this issue is circumvented by the use of a derivative-free optimization algorithm (see appendix).

To illustrate the discrimination capabilities of \mathcal{L} with the above parameters, Fig. 1 provides, for the example date discussed earlier, the value of $\mathcal{L}(\mathbf{R}, \mathbf{R}^{\text{obs}})$ for every downscaled field \mathbf{R} obtained with a specific GSDM variant. We observe that, as the regression models complexify, the difference in texture with the analysis field \mathbf{R}^{obs} decreases, the lowest being here obtained with the variant E32–S31.

d. Application to indices of specific aspects of texture

Quantifying the texture as a whole is useful for calibration. For validation though, it can also be of interest to identify specific attributes of the texture, and evaluate them separately. In this section, we propose novel indices that extract specific information from the gridded variogram. The stratification is here removed ($K = 1$) to reduce the amount of information, while the window size is increased ($L > 1$) to better quantify the anisotropy characteristics. The first step is to transform the gridded variogram, i.e., the set of values $\gamma_{\mathbf{R}}(\Delta i, \Delta j)$ for $\Delta i, \Delta j = (-L, \dots, L)$, to a so-called “lag-direction” variogram denoted $\tau_{\mathbf{R}}(h, \theta)$, where h and θ denote the lag and the direction, respectively. Using trigonometry, the mapping from γ to τ is straightforward [e.g., $\gamma_{\mathbf{R}}(1, 1) = \tau_{\mathbf{R}}(\sqrt{2}, 45^\circ)$], and is illustrated in Figs. 3a,b. Let us also denote by Θ the finite set of θ values for which $\tau_{\mathbf{R}}(h, \theta)$ can be computed. To further synthesize the information, we then reduce the variogram curve for every direction $\theta \in \Theta$ to a single number, assuming a fixed lag $h = D$. The variogram values at combinations (D, θ) that are not directly available from the gridded variogram are obtained via linear interpolation (see Fig. 3b). Finally, the function $\tau_{\mathbf{R}}(D, \theta)$, available for the directions $\theta \in \Theta$, can be plotted on a polar plot to provide a clear picture of the anisotropy (Fig. 3c). Concerning the choice of D , we need to

compromise between short lags that convey more information about the texture, while being large enough to allow the vector Θ to contain a larger variety of directions, and thus better characterize the anisotropy. We here suggest $D = \sqrt{5}$, which is the shortest lag for which one can derive $\tau_{\mathbf{R}}(D, \theta)$ for eight different directions: $\Theta = \{90^\circ, 63^\circ, 45^\circ, 27^\circ, 0^\circ, -27^\circ, -45^\circ, -63^\circ\}$. Due to symmetry we have $\tau_{\mathbf{R}}(h, \theta) = \tau_{\mathbf{R}}(h, \theta \pm 180^\circ)$, hence only a half circle is necessary for representing the function $\tau_{\mathbf{R}}(D, \theta)$. One can note that $D = \sqrt{5}$ implies computing the gridded variogram with $L = 3$, although not all values are actually necessary (see Fig. 3a). For the parameters λ , the same value ($\lambda = 0.5$) as for the loss function is used.

Now that the function $\tau_{\mathbf{R}}(D, \theta)$ has been defined, we propose three indices that summarize specific information about the texture:

$$\text{ADI}_{\mathbf{R}} = \underset{\theta \in \Theta}{\text{argmin}} \tau_{\mathbf{R}}(D, \theta), \quad (10)$$

$$\text{ASI}_{\mathbf{R}} = \frac{\max_{\theta \in \Theta} \tau_{\mathbf{R}}(D, \theta)}{\min_{\theta \in \Theta} \tau_{\mathbf{R}}(D, \theta)}, \quad (11)$$

$$\text{SVI}_{\mathbf{R}} = \min_{\theta \in \Theta} \tau_{\mathbf{R}}(D, \theta). \quad (12)$$

The anisotropy direction index (ADI) is simply the preferential direction (in degrees) of the anisotropy, that is, the direction where precipitation is most correlated. It can only take values in Θ . The second index, the anisotropy strength index (ASI), quantifies the degree of anisotropy. It takes values in $[1, +\infty[$, with 1 corresponding to isotropy. Finally, the small-scale variability index (SVI) aims at quantifying the spatial variability independently from the anisotropy, by extracting the variogram value in the preferential direction. It takes values in \mathbb{R}^+ , with large values associated with high spatial variability. The values of ADI, ASI, and SVI are given in Fig. 1 for the example fields. One can then revisit the qualitative comments made in section 3d, about

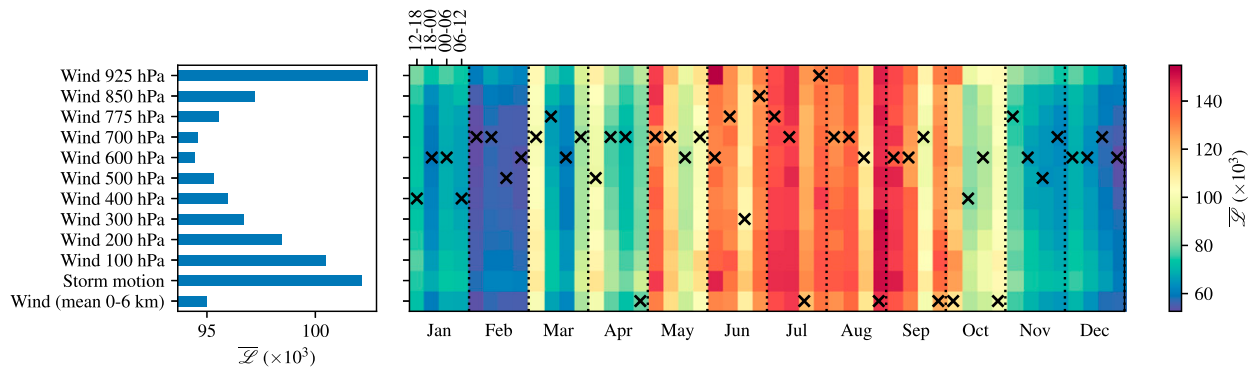


FIG. 4. Calibration results of the cost function $\bar{\mathcal{L}}$ for the GSDM variant E32–S31, for different predictors controlling the anisotropy. The lower $\bar{\mathcal{L}}$ the better. (left) Results aggregated over all months and 6-h periods. (right) Results for every months and 6-h periods, with black crosses indicating the best predictor. For this experiment, the predictor “wind shear sfc–850 hPa” is used for controlling the spatial variability.

the specific effects of each regression model, in the light of these quantitative measures.

For constructing verification metrics, these three indices are computed on each analysis and downscaled fields over the validation period, and the root mean squared error (RMSE) is considered. In the next section, we will refer to these metrics as RMSE_{ADI} , RMSE_{ASI} , and RMSE_{SVI} . Note that because ADI is an angle, the mathematical formulation of differences in the equation of RMSE_{ADI} must be adapted to circular data.

As a concluding remark, the computation of both the loss function \mathcal{L} and the indices ADI, ASI, and SVI requires that the rainfall fields have enough nonzero pixels for the estimation of the γ values to be robust. In this study, an arbitrary threshold of 10% was applied to the analysis fields, that is, all pairs $\{\mathbf{R}, \mathbf{R}^{\text{obs}}\}$ with less than 10% of nonzero pixels in \mathbf{R}^{obs} were discarded from the computation of the metrics.

5. Results and discussion

a. Choice of the best meteorological predictor

Prior to evaluating the performances in validation of the different GSDM variants, we discuss the choice of the predictors used for controlling the anisotropy (in models E21 and E32) and the spatial variability (in model S31). For a fair comparison against the other models, the choice of the predictors must be made from the calibration results only, that is, from the reported minima of the cost function $\bar{\mathcal{L}}$ after optimization. Figure 4 shows the calibration results of the variant E32–S31 for each different anisotropy predictors tested. Recall that these predictors are vectors, where the magnitude is used for controlling the anisotropy strength (in model E32 only) and the direction for controlling the anisotropy direction. The left plot, which aggregates $\bar{\mathcal{L}}$ over all months and 6-h period, shows that the wind at the 600-hPa pressure level is the best predictor overall, followed closely by the wind at 700 hPa, which was used by Gagnon and Rousseau (2014), and the average wind in the 0–6-km layer AGL. The results for every month and 6-h period (right plot) indicate, though, that it is not systematically the same predictor that performs best, even if it is difficult to derive a physical interpretation as the variability seems to be caused by randomness (due to the finite size

of the calibration sample) more than related to the season or time of the day. Figure 5 now shows the results of the same experiment but for the predictors controlling the spatial variability. The vertical wind shear between the surface and the 850-hPa pressure level is found to be the best predictor overall (left plot), but a series of other candidates perform only slightly worse. When looking at the detailed results (right plot), unlike for the anisotropy predictor, a distinguishable pattern is visible. Some indices of the atmosphere instability (CAPE, SI, LI) tend to be selected as the best predictor during the winter months, as well as for the 0000–0600 CST period in summer, that is, when convection has most often not started yet. During the convective 6-h periods, kinematic predictors (wind speed, wind shear) are more informative, as they are less affected by convection. The pressure levels at which the kinematic predictors are selected also vary, although it is again difficult to derive a robust physical interpretation.

To conclude, these results show that, beyond a certain random variability, a pattern for the type of predictor that performs best for each season and time of the day exists, which can potentially be leveraged by the GSDM. On the other hand, setting a framework that picks up different predictors is costly, both in terms of meteorological data access and calibration time. For operational reasons, it can thus be desirable for the downscaling method to restrict to a single set of predictors. Therefore, in what follows we present the results for (i) the models E21, E32, and S31, which uses the optimal predictor overall (wind 600 hPa for E21 and E32; wind shear sfc–850 hPa for S31), and (ii) the models E21^B , E32^B , and S31^B , which pick the best predictor for every month and 6-h period.

b. Validation results

For comparing the performance of the different GSDM variants, the metric $\bar{\mathcal{L}}$ is computed over the validation period. As the downscaling process is stochastic, 10 downscaled field “members” are generated for every date, and the ensemble mean of the metric is reported. As a benchmark, the performance of the bilinear interpolation method is also reported. Monthly results are aggregated by seasons: winter (DJF), spring (MAM), summer (JJA), and fall (SON). To assess the significance of the results, we conduct the following statistical

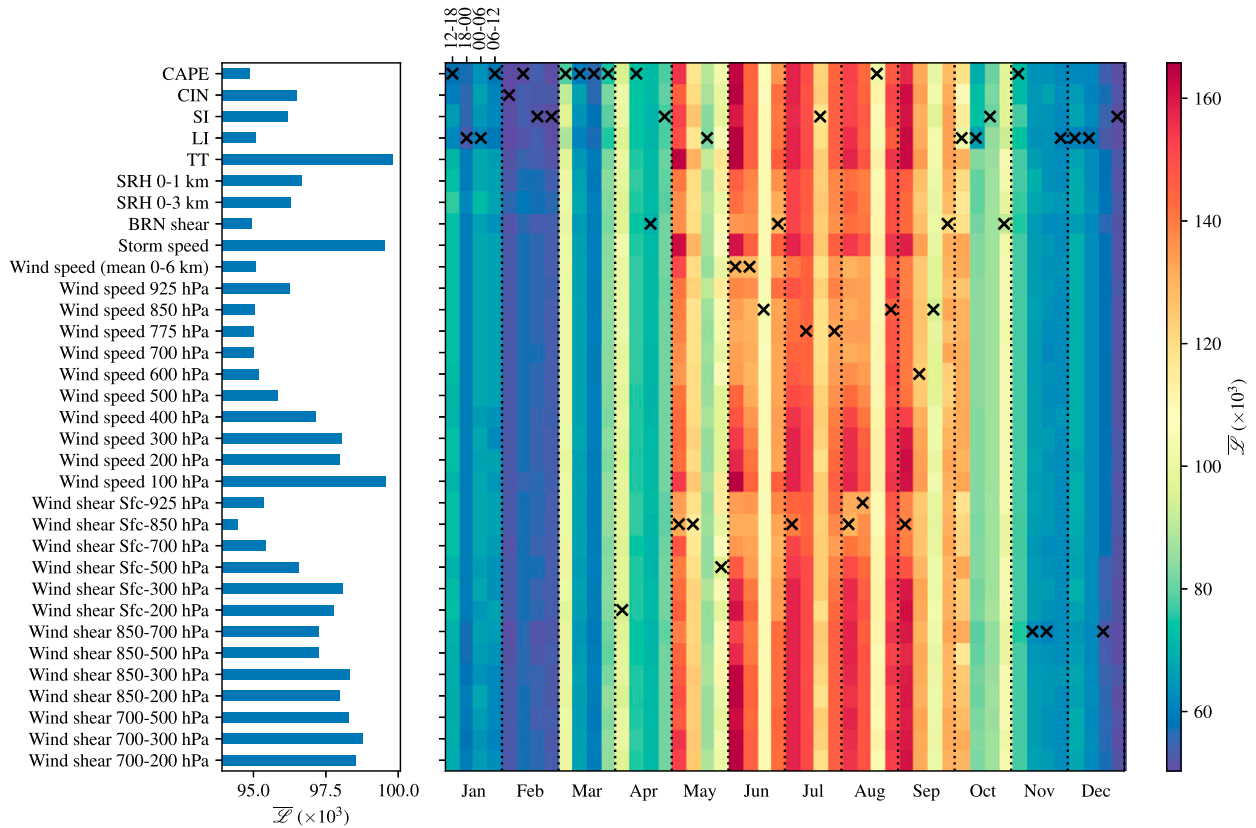


FIG. 5. As in Fig. 4, but for different predictors controlling the spatial variability. For this experiment, the predictor “wind 600 hPa” is used for controlling the anisotropy.

testing. Paired Student’s *t* tests are applied for rejecting—or not—the null hypothesis of two variants having the same \bar{Z} . To account for the multiplicity of statistical testing (all variants are compared to each other), we apply the procedure recommended by Wilks (2016) that consists in controlling the false discovery rate (FDR; Benjamini and Hochberg 1995) at a chosen level $\alpha_{FDR} = 0.05$, that is, we put a ceiling of 0.05 on the statistically expected fraction of null hypotheses that were rejected although they were actually true. This testing is more stringent, as it rejects all null hypotheses with *p* values inferior to a threshold p_{FDR}^* (computed from the distribution of the sorted *p* values) that is lower than α_{FDR} . For more details, we refer to Wilks (2016). From now, all the results that are reported and discussed in the body of the text have been found statistically significant according to this procedure. Figures S2–S5 provide the outcomes of the statistical testing.

The performances of all GSDM variants are shown in Fig. 6. As earlier, two distinct regimes with different results can be observed: the winter season, which includes, to a certain extent, the night and morning of the spring and fall seasons, and the summer season. To help identify the deficiencies of the different variants, Fig. 7 also provides, for 1200–1800 CST, the RMSE of the indices ASI, ADI, and SVI, which specifically quantify the discrepancies in the anisotropy strength, the anisotropy direction, and the spatial variability, respectively.

During winter, a large gap in performance appears between E00/E10 and E30/E21/E32, the first two models being strongly penalized for not reproducing the anisotropy, which appears to be the spatial characteristic of precipitation the most important to reproduce during that season. This finding is confirmed by Fig. 7, where $RMSE_{ASI}$ is about twice as large for E00/E10 than for E30/E21/E32, whereas they perform fairly similar regarding the other indices. Moreover, we observe that the predictor-based models E21/E32 are not significantly better than E30 in winter. At first glance, this result could indicate that the shape of the anisotropy for that season does not vary enough from its climatology for these models to perform better. However, further investigations (not shown) reveal that there exist a number of cases where anisotropy is indeed “unusual” (e.g., the example shown in Fig. 1) and over which the models E21/E32 do perform significantly better. The majority of cases are “usual” though, and in these cases E30 performs slightly better. We conclude that in a regime where the anisotropy shape rarely departs from its climatology, or/and the anisotropy predictor is not informative, E30 should be chosen over E21/E32, due to its simpler regression model. Another finding is that E32 does not systematically outperform E21, meaning that the added value of wind speed for controlling the anisotropy strength is limited, even when the best predictor is picked (model E32^B). In Fig. 7, we indeed observe that using the wind direction results in a sound

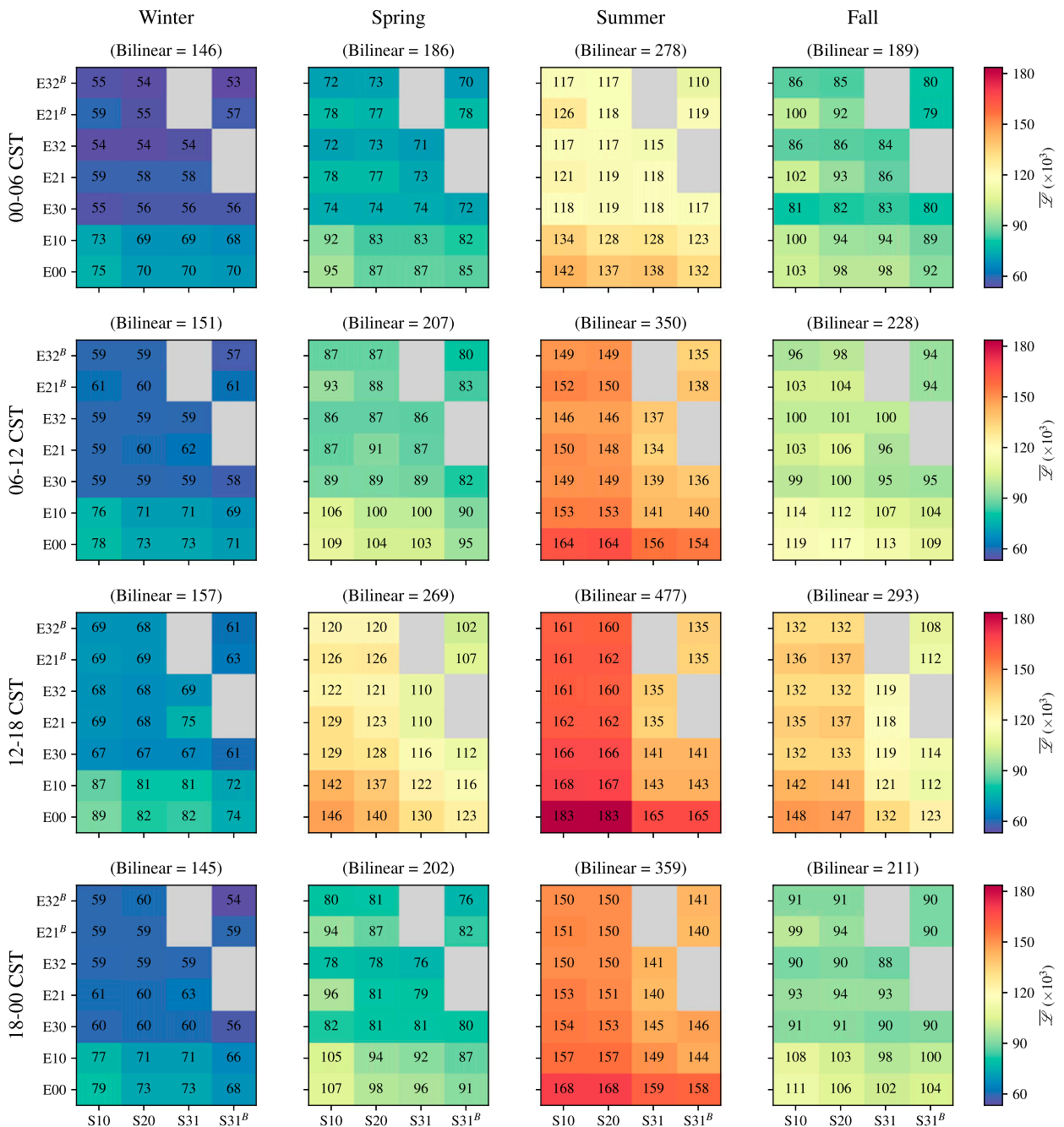


FIG. 6. Validation results of the cost function \bar{Z} for all combinations of regression models for the expectation $E_{(i,j)}$ (y axis) and the standard deviation $SD_{(i,j)}$ (x axis), plus the bilinear interpolation as a benchmark (in parentheses). The values of \bar{Z} are reported in addition to the color scale. Models $E32^B$, $E21^B$, and $S31^B$ use the best predictors for every month and 6-h period, while models $E32$, $E21$, and $S31$ use the best predictors overall. The combinations of both (gray areas) have not been tested.

improvement regarding the ADI, but introducing the wind speed does not improve regarding the ASI. Note that this result is also valid for the summer period. Concerning the regression models for the standard deviation $SD_{(i,j)}$, Fig. 6 indicates that in winter the models S10/S20 perform as well as the predictor-based model S31. However, $S31^B$ significantly improves the performance, at least for the 6-h periods where convection is most

likely to occur (1200–1800 and 1800–0000 CST). As discussed earlier, S31 uses a kinematic predictor that has been found best in overall, although in winter the most informative predictors are indices of instability, which are used by $S31^B$ (Fig. 5).

Over the summer season, the conclusions differ to a certain extent. In Fig. 6 the large gap in performance is now observed between S10/S20 and S31, which highlights the greater

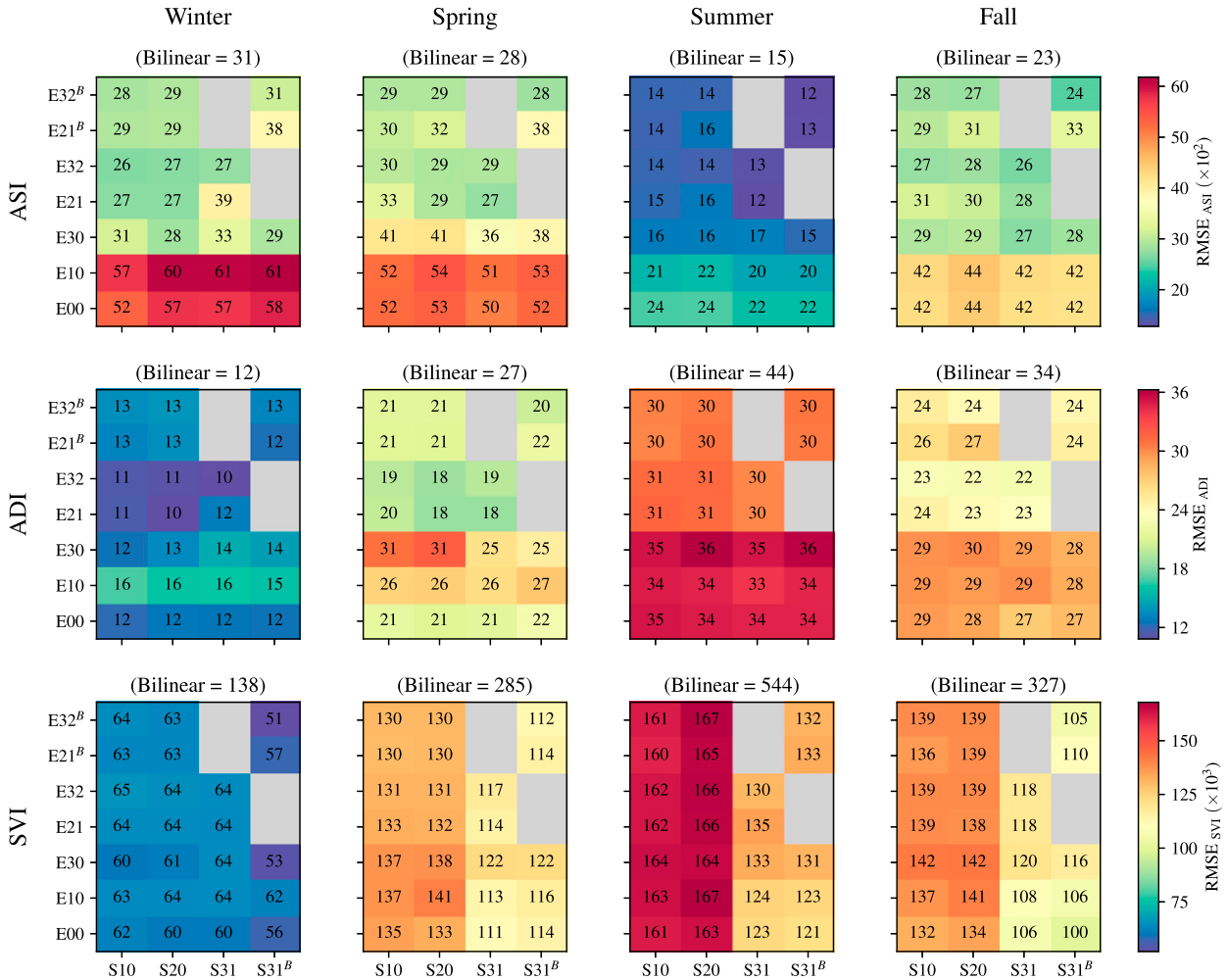


FIG. 7. Validation results of the metrics (top) $RMSE_{ADI}$, (middle) $RMSE_{ASI}$, and (bottom) $RMSE_{SVI}$, over 1200–1800 CST. The values of the metrics are reported in addition to the color scale.

importance of the spatial variability over the anisotropy for that season. This finding is confirmed by Fig. 7, where $RMSE_{SVI}$ for models S10/S20 is substantially higher than for S31, although for the other indices the gap is invisible. We further observe that S31 performs as well as S_{31}^B , which indicates that the predictor used by S31, namely, the wind shear sfc–850 hPa, is actually a robust choice for summer whatever the time of the day. Finally, it is found that E10 significantly outperforms E00 in summer, which seems to demonstrate that the distance term in the regression model

for $E_{(ij)}$ plays an important role in reproducing the correct spatial variability.

c. Discussion of the ability to reproduce field maxima

After having evaluated the ability of the GSDM to reproduce the correct texture, we finally assess its ability to reproduce the maximum over the entire domain, an aspect that can be of interest for users interested in extremes. The downscaling process being stochastic, it is not expected that every downscaled field maximum matches exactly the observed

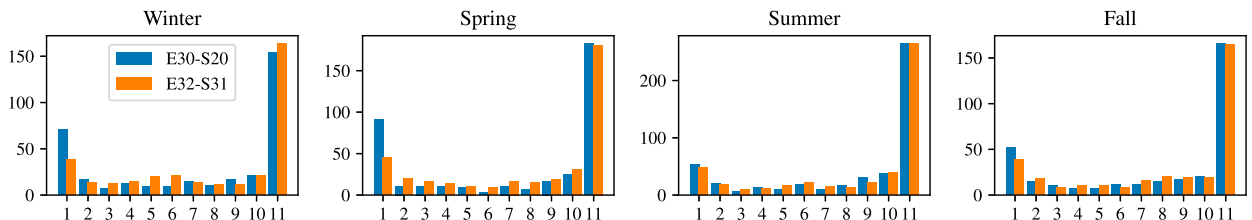


FIG. 8. Verification rank histograms of the maximal precipitation value over the field, for the GSDM variants E30–S20 and E32–S31.

maximum. However, a sample obtained from several different Gibbs sampling runs should encompass the observed maximum in such a way that the simulated maxima are statistically indistinguishable from the observed maximum. This property, referred in the field of probabilistic forecasting to as reliability (Jolliffe and Stephenson 2012), can be verified using rank histograms, whose deviation from flatness indicates specific miscalibrations: underdispersion (U shape), overdispersion (∩ shape), underestimation (∩ shape), and overestimation (∪ shape) [for caveats about this interpretation, see Hamill (2001) and Bellier et al. (2017)]. Figure 8 shows the rank histograms for the two most relevant GSDM variants, E30–S20 and E32–S31. We observe an underdispersion for reproducing the field maxima, with a strong tendency for underestimation, especially in summer. The explanation for this finding is twofold. First, as already mentioned in Gagnon et al. (2012) it seems that the lognormal distribution chosen for constructing the conditional distribution $p(R_{(i,j)}|\mathbf{N}_{(i,j)})$ has a right tail that is too light for correctly fitting the high values of precipitation. If the reproduction of maxima is the most important point of downscaling, we then suggest opting for a heavier-tail distribution. Second, the GSDM has been here calibrated for reproducing the correct texture, but not particularly the field maxima. Indeed, the gridded, stratified variogram used for constructing the loss function \mathcal{L} , especially with the square root prior transformation ($\lambda = 0.5$, see section 4c), is relatively insensitive to precipitation maxima, such that the calibration does not penalize much the models for not reproducing them correctly. However, the revised GSDM proposed in this article is flexible in the sense that every user can define a loss function that corresponds to their specific need. For reproducing the precipitation maxima, although it is possible that using another value for λ helps, we believe that constructing a different loss function is necessary. It could, for instance, encourage sporadic but steep gradients in the downscaled fields, which in convective storms are generally associated with precipitation extremes (Foufoula-Georgiou et al. 2014).

6. Conclusions

We have reviewed the Gibbs sampling disaggregation model (GSDM), a precipitation spatial downscaling method originally proposed by Gagnon et al. (2012). The GSDM is capable of introducing realistic, weather-dependent, and possibly anisotropic fine-scale details, while preserving the mean rain rate over the coarse-scale pixels, without creating blockiness or any other visual discontinuities. It can be applied to any downscaling setup, without constraints on the domain size or the scaling ratio. The downscaling process consists in iteratively updating the fine-scale pixel values by sampling from distributions conditioned on the neighboring pixels, and parameterized via two regressions models (for the expectation and the standard deviation). Three main novelties have been introduced to this method.

First, a modification of the Gibbs sampling algorithm has been proposed that ensures the downscaled fields to have a similar texture to that of the analyses. Much fewer iterations

than in the original version are required, which drastically reduces the computational cost. To quantify discrepancies in rainfall texture, a loss function based on the gridded, stratified variogram has been proposed. Under the perspective mentioned in the introduction, namely, using the GSDM within the U.S. National Blend of Models to recreate a coherent spatial structure within postprocessed probabilistic forecasts, the correct texture in the downscaled fields prevails. However, our algorithm is flexible such that users with different interests (e.g., correctly reproduce the precipitation maxima) can use the GSDM with another loss function that is tailored to their specific need.

Second, equations of various complexity were tested for the two regression models, with the objective of assessing the added value of each different term. Among all combinations, two can be retained as the most relevant in perspective of an operational use. First, E30–S20 should be chosen in case the use of meteorological predictors cannot be afforded, or they are suddenly not available (e.g., failure in data repatriation). Otherwise, the most sophisticated variant E32–S31 is the best choice, as it reduces to less complex models in case the predictors are not informative. The only exception is when the predictor for the wind direction is not informative; in that case E30 should be chosen over E32.

Third, we have conducted an extensive test of various meteorological predictors for controlling the rainfall field anisotropy (strength and direction) and spatial variability. For the anisotropy, which for 6-h accumulations is mainly caused by the advection through time of the rainfall cells, the best predictor all year round is the wind at 600 hPa. For the spatial variability, instability indices such as CAPE, LI, or SI were found most informative in winter (and during the early hours of the day in summer), while kinematic predictors such as the vertical wind shear were found better elsewhere, as they are less contaminated by convection. This indicates that the downscaling performances can be improved for the GSDM variant E32–S31 by selecting the adequate predictor for the two different periods. More work is needed, however, to extend these findings to other regions.

Further work will be conducted in a near future to allow the GSDM to run over a very large area such as the contiguous United States. First, the regression models will be adapted such that it can reproduce the orographic effects, drawing from the work of Gagnon et al. (2013). Then, since calibration can only be performed over a limited domain, a procedure will be proposed that ensures spatial coherence of the model parameters for each date. Finally, the verification aspect will be extended by propagating the downscaled fields into a high-resolution distributed hydrological model, to assess the extent to which improvements in reproducing fine-scale details in precipitation improve the simulation of the streamflow.

Acknowledgments. This research was supported by funding from U.S. NWS Office of Science and Technology Integration through the Meteorological Development Laboratory, project

T8MWQML. We acknowledge Eric Engle from NWS for providing us the CCPA precipitation dataset at the resolution of $1/16^\circ$. We also thank the three anonymous reviewers for their constructive comments.

APPENDIX

Technical Considerations for Minimizing $\overline{\mathcal{L}}$

Because of the stochasticity of the Gibbs sampling, the cost function $\overline{\mathcal{L}}$ to be minimized during the calibration of the GSDM is itself a random quantity, meaning that replications of its evaluation (with identical parameters) will lead to different values. Several approaches exist for tackling optimization problems under a noisy cost function. First, fixing at the beginning of the Gibbs sampling the random “seed,” i.e., the pseudorandom number generator of the machine, will stabilize $\overline{\mathcal{L}}$ across its different evaluations, ensuring convergence of the optimization. Second, considering for every evaluation of $\overline{\mathcal{L}}$ not only one estimate but the average of multiple replications will make the response surface of the model (i.e., the representation of the cost function in the parameter space) smoother. A certain smoothness reduces the risk of minor local minima, but also increases the robustness of the solution to the random seed. Third, using derivative-free optimization algorithms is safer, as they can handle nonsmooth and nondifferentiable cost functions. Given these considerations, one must choose a strategy that ensures a sufficient robustness at an acceptable computational cost. In this study, the optimization has been found satisfactory when using the derivative-free [Nelder and Mead \(1965\)](#) algorithm (also known as the downhill simplex), fixing the random seed, but restricting to a single replication when evaluating $\overline{\mathcal{L}}$. Indeed, in our study the number of dates in the verification period (over which $\overline{\mathcal{L}}$ is averaged) was found high enough for ensuring a sufficient smoothness.

Another consideration for the calibration of the GSDM is the risk of local minima in the response surface. This may typically occur if different parameters induce a similar effect in the model. For instance here, the parameters of the regression model for $SD_{(i,j)}$ influence the spatial variability of the rainfall field, but so does the parameter β_d in the model of $E_{(i,j)}$. Global optimization algorithms are designed to find the global minimum among complex response surfaces, but at a slow convergence rate. As a much faster approach for this specific optimization problem, we suggest a sequential optimization based on the progressive complexification of the model to be calibrated. In [Tables 2 and 3](#), one can note that every model, except the top one, has a “parent” model to which it reduces by neutralizing one (or sometimes two) free parameters (i.e., setting their value to 0 or 1 such that its effect vanishes). For example, E32 has E21 as parent, while E21 and E30 share the same parent, E10. By calibrating the first parent model and using the so-optimized parameters as initial values for the optimization of the next “child” model, and so on, we ensure that any child model performs at least as good as its less complex parent model.

REFERENCES

- Achberger, C., M.-L. Linderson, and D. Chen, 2003: Performance of the Rossby Centre regional atmospheric model in Southern Sweden: Comparison of simulated and observed precipitation. *Theor. Appl. Climatol.*, **76**, 219–234, <https://doi.org/10.1007/s00704-003-0015-6>.
- Ahrens, B., 2003: Rainfall downscaling in an alpine watershed applying a multiresolution approach. *J. Geophys. Res.*, **108**, 8388, <https://doi.org/10.1029/2001JD001485>.
- Allcroft, D. J., and C. A. Glasbey, 2003: A latent Gaussian Markov random-field model for spatiotemporal rainfall disaggregation. *J. Roy. Stat. Soc.*, **52C**, 487–498, <https://doi.org/10.1111/1467-9876.00419>.
- Badas, M. G., R. Deidda, and E. Piga, 2006: Modulation of homogeneous space-time rainfall cascades to account for orographic influences. *Nat. Hazards Earth Syst. Sci.*, **6**, 427–437, <https://doi.org/10.5194/nhess-6-427-2006>.
- Bellier, J., I. Zin, and G. Bontron, 2017: Sample stratification in verification of ensemble forecasts of continuous scalar variables: Potential benefits and pitfalls. *Mon. Wea. Rev.*, **145**, 3529–3544, <https://doi.org/10.1175/MWR-D-16-0487.1>.
- Benjamini, Y., and Y. Hochberg, 1995: Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. Roy. Stat. Soc.*, **57B**, 289–300, <https://doi.org/10.1111/J.2517-6161.1995.TB02031.X>.
- Brune, S., F. Kapp, and P. Friederichs, 2018: A wavelet-based analysis of convective organization in ICON large-eddy simulations. *Quart. J. Roy. Meteor. Soc.*, **144**, 2812–2829, <https://doi.org/10.1002/qj.3409>.
- Bunkers, M. J., B. A. Klimowski, J. W. Zeitler, R. L. Thompson, and M. L. Weisman, 2000: Predicting supercell motion using a new hodograph technique. *Wea. Forecasting*, **15**, 61–79, [https://doi.org/10.1175/1520-0434\(2000\)015<0061:PSMUAN>2.0.CO;2](https://doi.org/10.1175/1520-0434(2000)015<0061:PSMUAN>2.0.CO;2).
- Buschow, S., J. Pidstrigach, and P. Friederichs, 2019: Assessment of wavelet-based spatial verification by means of a stochastic precipitation model (wv_verif v0.1.0). *Geosci. Model Dev.*, **12**, 3401–3418, <https://doi.org/10.5194/gmd-12-3401-2019>.
- Casella, G., and E. I. George, 1992: Explaining the Gibbs sampler. *Amer. Stat.*, **46**, 167–174, <https://doi.org/10.2307/2685208>.
- Cressie, N., and D. M. Hawkins, 1980: Robust estimation of the variogram. I. *J. Int. Assoc. Math. Geol.*, **12**, 115–125, <https://doi.org/10.1007/BF01035243>.
- Deidda, R., 2000: Rainfall downscaling in a space-time multifractal framework. *Water Resour. Res.*, **36**, 1779–1794, <https://doi.org/10.1029/2000WR900038>.
- Ebtehaj, A. M., E. Foufoula-Georgiou, and G. Lerman, 2012: Sparse regularization for precipitation downscaling. *J. Geophys. Res.*, **117**, D08107, <https://doi.org/10.1029/2011JD017057>.
- Eckley, I. A., G. P. Nason, and R. L. Treloar, 2010: Locally stationary wavelet fields with application to the modelling and analysis of image texture. *J. Roy. Stat. Soc.*, **59C**, 595–616, <https://doi.org/10.1111/j.1467-9876.2009.00721.x>.
- Ekström, M., 2016: Metrics to identify meaningful downscaling skill in WRF simulations of intense rainfall events. *Environ. Modell. Software*, **79**, 267–284, <https://doi.org/10.1016/j.envsoft.2016.01.012>.
- Erdin, R., C. Frei, and H. R. Künsch, 2012: Data transformation and uncertainty in geostatistical combination of radar and rain gauges. *J. Hydrometeorol.*, **13**, 1332–1346, <https://doi.org/10.1175/JHM-D-11-096.1>.
- Ferraris, L., S. Gabellani, N. Reborra, and A. Provenzale, 2003: A comparison of stochastic models for spatial rainfall downscaling. *Water Resour. Res.*, **39**, 1368, <https://doi.org/10.1029/2003WR002504>.

- Foufoula-Georgiou, E., A. M. Ebtehaj, S. Zhang, and A. Hou, 2014: Downscaling satellite precipitation with emphasis on extremes: A variational ℓ_1 -norm regularization in the derivative domain. *Surv. Geophys.*, **35**, 765–783, <https://doi.org/10.1007/s10712-013-9264-9>.
- Fowler, H. J., S. Blenkinsop, and C. Tebaldi, 2007: Linking climate change modelling to impacts studies: Recent advances in downscaling techniques for hydrological modelling. *Int. J. Climatol.*, **27**, 1547–1578, <https://doi.org/10.1002/joc.1556>.
- Gagnon, P., and A. N. Rousseau, 2014: Stochastic spatial disaggregation of extreme precipitation to validate a regional climate model and to evaluate climate change impacts over a small watershed. *Hydrol. Earth Syst. Sci.*, **18**, 1695–1704, <https://doi.org/10.5194/hess-18-1695-2014>.
- , —, A. Mailhot, and D. Caya, 2012: Spatial disaggregation of mean areal rainfall using Gibbs sampling. *J. Hydrometeorol.*, **13**, 324–337, <https://doi.org/10.1175/JHM-D-11-034.1>.
- , —, —, and —, 2013: A Gibbs sampling disaggregation model for orographic precipitation. *Int. J. Appl. Earth Obs. Geoinf.*, **22**, 16–26, <https://doi.org/10.1016/j.jag.2011.11.002>.
- Galway, J. G., 1956: The lifted index as a predictor of latent instability. *Bull. Amer. Meteor. Soc.*, **37**, 528–529, <https://doi.org/10.1175/1520-0477-37.10.528>.
- Gilleland, E., D. A. Ahijevych, B. G. Brown, and E. E. Ebert, 2010: Verifying forecasts spatially. *Bull. Amer. Meteor. Soc.*, **91**, 1365–1376, <https://doi.org/10.1175/2010BAMS2819.1>.
- Guan, H., J. L. Wilson, and H. Xie, 2009: A cluster-optimizing regression-based approach for precipitation spatial downscaling in mountainous terrain. *J. Hydrol.*, **375**, 578–588, <https://doi.org/10.1016/j.jhydrol.2009.07.007>.
- Hamill, T. M., 2001: Interpretation of rank histograms for verifying ensemble forecasts. *Mon. Wea. Rev.*, **129**, 550–560, [https://doi.org/10.1175/1520-0493\(2001\)129<0550:IORHFV>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<0550:IORHFV>2.0.CO;2).
- , and M. Scheuerer, 2018: Probabilistic precipitation forecast postprocessing using quantile mapping and rank-weighted best-member dressing. *Mon. Wea. Rev.*, **146**, 4079–4098, <https://doi.org/10.1175/MWR-D-18-0147.1>.
- Harris, D., and E. Foufoula-Georgiou, 2001: Subgrid variability and stochastic downscaling of modeled clouds: Effects on radiative transfer computations for rainfall retrieval. *J. Geophys. Res.*, **106**, 10 349–10 362, <https://doi.org/10.1029/2000JD900797>.
- He, X., N. W. Chaney, M. Schleiss, and J. Sheffield, 2016: Spatial downscaling of precipitation using adaptable random forests. *Water Resour. Res.*, **52**, 8217–8237, <https://doi.org/10.1002/2016WR019034>.
- Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Quart. J. Roy. Meteor. Soc.*, **146**, 1999–2049, <https://doi.org/10.1002/qj.3803>.
- Hou, D., and Coauthors, 2014: Climatology-calibrated precipitation analysis at fine scales: Statistical adjustment of stage IV toward CPC gauge-based analysis. *J. Hydrometeorol.*, **15**, 2542–2557, <https://doi.org/10.1175/JHM-D-11-0140.1>.
- Jolliffe, I. T., and D. B. Stephenson, 2012: *Forecast Verification: A Practitioner's Guide in Atmospheric Science*. 2nd ed. John Wiley & Sons, 274 pp.
- Kapp, F., P. Friederichs, S. Brune, and M. Weniger, 2018: Spatial verification of high-resolution ensemble precipitation forecasts using local wavelet spectra. *Meteor. Z.*, **27**, 467–480, <https://doi.org/10.1127/metz/2018/0903>.
- Lam, S. K., A. Pitrou, and S. Seibert, 2015: Numba: A LLVM-based python JIT compiler. *Proc. Second Workshop on the LLVM Compiler Infrastructure in HPC*, Austin, TX, ACM, 7, <https://doi.org/10.1145/2833157.2833162>.
- Lin, Y., and K. E. Mitchell, 2005: The NCEP Stage II/IV hourly precipitation analyses: Development and applications. *19th Conf. Hydrology*, San Diego, CA, Amer. Meteor. Soc., 1, 2, <https://ams.confex.com/ams/pdfpapers/83847.pdf>.
- Maraun, D., and Coauthors, 2010: Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user. *Rev. Geophys.*, **48**, RG3003, <https://doi.org/10.1029/2009RG000314>.
- Markowski, P., and Y. Richardson, 2011: *Mesoscale Meteorology in Midlatitudes*. John Wiley & Sons, 430 pp.
- Marzban, C., and S. Sandgathe, 2009: Verification with variograms. *Wea. Forecasting*, **24**, 1102–1120, <https://doi.org/10.1175/2009WAF2222122.1>.
- May, R. M., S. C. Arms, P. Marsh, E. Bruning, J. R. Leeman, K. Goebbert, J. E. Thielen, and Z. Bruck, 2019: MetPy: A Python package for meteorological data. GitHub, <https://github.com/Unidata/MetPy>.
- Miller, R. C., 1975: Notes on analysis and severe-storm forecasting procedures of the Air Force Global Weather Central. Tech. Rep. 200, Air Weather Service, 177 pp.
- Nelder, J. A., and R. Mead, 1965: A simplex method for function minimization. *Comput. J.*, **7**, 308–313, <https://doi.org/10.1093/comjnl/7.4.308>.
- Onibon, H., T. Lebel, A. Afouda, and G. Guillot, 2004: Gibbs sampling for conditional spatial disaggregation of rain fields. *Water Resour. Res.*, **40**, W08401, <https://doi.org/10.1029/2003WR002009>.
- Park, N.-W., 2013: Spatial downscaling of TRMM precipitation using geostatistics and fine scale environmental variables. *Adv. Meteor.*, **2013**, 1–9, <https://doi.org/10.1155/2013/237126>.
- Perica, S., and E. Foufoula-Georgiou, 1996a: Linkage of scaling and thermodynamic parameters of rainfall: Results from midlatitude mesoscale convective systems. *J. Geophys. Res.*, **101**, 7431–7448, <https://doi.org/10.1029/95JD02372>.
- , and —, 1996b: Model for multiscale disaggregation of spatial rainfall based on coupling meteorological and scaling descriptions. *J. Geophys. Res.*, **101**, 26 347–26 361, <https://doi.org/10.1029/96JD01870>.
- Raftery, A. E., and S. Lewis, 1992: How many iterations in the Gibbs sampler? *Bayesian Statistics 4*, J. M. Bernardo et al., Ed., Oxford University Press, 763–773.
- Raut, B. A., A. W. Seed, M. J. Reeder, and C. Jakob, 2018: A multiplicative cascade model for high-resolution space-time downscaling of rainfall. *J. Geophys. Res. Atmos.*, **123**, 2050–2067, <https://doi.org/10.1002/2017JD027148>.
- , M. J. Reeder, C. Jakob, and A. W. Seed, 2019: Stochastic space-time downscaling of rainfall using event-based multiplicative cascade simulations. *J. Geophys. Res. Atmos.*, **124**, 3889–3902, <https://doi.org/10.1029/2018JD029343>.
- Rebora, N., L. Ferraris, J. Von Hardenberg, and A. Provenzale, 2006a: Rainfall downscaling and flood forecasting: A case study in the Mediterranean area. *Nat. Hazards Earth Syst. Sci.*, **6**, 611–619, <https://doi.org/10.5194/nhess-6-611-2006>.
- , —, —, and —, 2006b: RainFARM: Rainfall downscaling by a filtered autoregressive model. *J. Hydrometeorol.*, **7**, 724–738, <https://doi.org/10.1175/JHM517.1>.
- Salas, F. R., and Coauthors, 2018: Towards real-time continental scale streamflow simulation in continuous and discrete space. *J. Amer. Water Resour. Assoc.*, **54**, 7–27, <https://doi.org/10.1111/1752-1688.12586>.
- Schefzik, R., T. L. Thorarinsdottir, and T. Gneiting, 2013: Uncertainty quantification in complex simulation models using ensemble copula coupling. *Stat. Sci.*, **28**, 616–640, <https://doi.org/10.1214/13-STS443>.

- Schuurmans, J., and M. Bierkens, 2007: Effect of spatial distribution of daily rainfall on interior catchment response of a distributed hydrological model. *Hydrol. Earth Syst. Sci.*, **11**, 677–693, <https://doi.org/10.5194/hess-11-677-2007>.
- , —, E. Pebesma, and R. Uijlenhoet, 2007: Automatic prediction of high-resolution daily rainfall fields for multiple extents: The potential of operational radar. *J. Hydrometeorol.*, **8**, 1204–1224, <https://doi.org/10.1175/2007JHM792.1>.
- Showalter, A., 1947: A stability index for thunderstorm forecasting. *Bull. Amer. Meteor. Soc.*, **34**, 250–252, <https://doi.org/10.1175/1520-0477-34.6.250>.
- Terzago, S., E. Palazzi, and J. Hardenberg, 2018: Stochastic downscaling of precipitation in complex orography: A simple method to reproduce a realistic fine-scale climatology. *Nat. Hazards Earth Syst. Sci.*, **18**, 2825–2840, <https://doi.org/10.5194/nhess-18-2825-2018>.
- Tsonevsky, I., C. A. Doswell III, and H. E. Brooks, 2018: Early warnings of severe convection using the ECMWF Extreme Forecast Index. *Wea. Forecasting*, **33**, 857–871, <https://doi.org/10.1175/WAF-D-18-0030.1>.
- UCAR, 2019: MetPy. UCAR Unidata, <https://doi.org/10.5065/D6WW7G29>.
- Viterbo, F., and Coauthors, 2020: A multi-scale, hydro-meteorological forecast evaluation of National Water Model forecasts of the May 2018 Ellicott City, Maryland, flood. *J. Hydrometeorol.*, **21**, 475–499, <https://doi.org/10.1175/JHM-D-19-0125.1>.
- Weniger, M., F. Kapp, and P. Friederichs, 2017: Spatial verification using wavelet transforms: A review. *Quart. J. Roy. Meteor. Soc.*, **143**, 120–136, <https://doi.org/10.1002/qj.2881>.
- Wilks, D., 2016: “The stippling shows statistically significant grid points”: How research results are routinely overstated and overinterpreted, and what to do about it. *Bull. Amer. Meteor. Soc.*, **97**, 2263–2273, <https://doi.org/10.1175/BAMS-D-15-00267.1>.
- Xu, G., X. Xu, M. Liu, A. Y. Sun, and K. Wang, 2015: Spatial downscaling of TRMM precipitation product using a combined multifractal and regression approach: Demonstration for South China. *Water*, **7**, 3083–3102, <https://doi.org/10.3390/w7063083>.
- Younger, P. M., J. E. Freer, and K. J. Beven, 2009: Detecting the effects of spatial variability of rainfall on hydrological modelling within an uncertainty analysis framework. *Hydrol. Processes*, **23**, 1988–2003, <https://doi.org/10.1002/hyp.7341>.