Dear Editor and Reviewers,

Thank you for your new comments about our manuscript entitled “Non-stationary stochastic rain type generation: accounting for climate drivers”.

The paper has been revised accordingly. In particular, the supplementary material has been formatted as a technical note, and all figures are now provided with an enhanced resolution.

Hoping that our responses address your concerns, and that our propositions of improvements will fulfil your expectations,

Best regards,

Lionel Benoit, Mathieu Vrac and Gregoire Mariethoz.

Non-stationary stochastic rain type generation: accounting for climate drivers

Lionel Benoit[1], Mathieu Vrac[2], and Gregoire Mariethoz[1]

[1]Institute of Earth Surface Dynamics (IDYST), University of Lausanne, Lausanne, Switzerland
[2]Laboratory for Sciences of Climate and Environment (LSCE-IPSL), CNRS/CEA/UVSQ, Orme des Merisiers, France

Correspondence: Lionel Benoit (lionel.benoit@unil.ch)

Abstract. At sub-daily resolution, rain intensity exhibits a strong variability in space and time, which is favorably modelled using stochastic approaches. This strong variability is further enhanced because of the diversity of processes that produce rain (e.g. frontal storms, mesoscale convective systems, local convection), which results in a multiplicity of space-time patterns embedded into rain fields, and in turn leads to non-stationarity of rain statistics. To account for this non-stationarity in the context of stochastic weather generators, and therefore preserve the relationships between rainfall properties and climatic drivers, we propose to resort to rain types simulation.

In this paper, we develop a new approach based on multiple-point statistics to simulate rain type time series conditional to meteorological covariates. The rain type simulation method is tested by cross-validation using a 17-year long rain type time series defined over central Germany. Evaluation results indicate that the proposed approach successfully captures the relationships between rain types and meteorological covariates. This leads to a proper simulation of rain type occurrence, persistence and transitions. After validation, the proposed approach is applied to generate rain type time series conditional to meteorological covariates simulated by a Regional Climate Model under an RCP8.5 emission scenario. Results indicate that, by the end of the century, the distribution of rain types could be modified over the area of interest, with an increased frequency of convective- and frontal-like rains at the expense of more stratiform events.

1 Introduction

Stochastic weather generators are statistical models designed to simulate realistic random sequences of atmospheric variables (e.g. temperature, rain and wind). Their main target is to reproduce both the internal variability of each variable of interest, and the relationships between these variables (Wilks and Wilby, 1999; Furrer and Katz, 2007; Ailliot et al., 2015). These features make stochastic weather generators particularly well suited for producing synthetic climate histories in view of impact studies (Mavromatis and Hansen, 2001; Verdin et al., 2015; Paschalis et al., 2014), as well as for stochastic downscaling of climate projections (Burton et al., 2010; Wilks, 2010; Volosciuk et al., 2017). Within stochastic weather generators, rainfall has long been recognized as a critical variable, in particular because of the strong intermittency (Pardo-Igúzquiza et al., 2006; Schleiss et al., 2011) and variability (Smith et al., 2009; Gires et al., 2014) of the rain process. The apparent intermittency and variability of rainfall increase with the time resolution of interest (Krajewski et al, 2003; Mascaro et al., 2013), and if resolutions of the

order of 1 h (or finer) are considered, it appears that storms caused by different generation processes (e.g. frontal storms, mesoscale convective systems, local convection) result in different rain field organizations and temporal patterns (Emmanuel et al., 2012; Marra and Morin, 2018).

Such changes in rainfall characteristics make rain statistics time-varying. In terms of stochastic modelling, this implies that the stochastic process used to model rainfall is non-stationary through time, i.e. the parameters of the stochastic model change over time. The most common way to deal with the non-stationarity of rain statistics is to define a priori (i.e. prior to model calibration) the time periods during which stationarity is assumed. Afterwards, a piecewise-stationary modelling is applied, i.e. model parameters are kept constant within a single stationary period, but are allowed to vary between stationary periods. The temporal scale at which non-stationarity occurs is defined by the modeler according to prior knowledge and assumptions about the rain process at hand, and ranges from seasons (Paschalis et al., 2013; Bárdossy and Pegram, 2016; Peleg et al., 2017) to single rain storms (Caseri et al., 2016; Benoit et al., 2018a). However, several empirical studies have shown that at the sub-daily scale, rain statistics can change at a higher rate than alleged in most piecewise-stationary stochastic rainfall models. More precisely, rain statistics have been shown to abruptly change within a single day (Emmanuel et al., 2012), and even within a single rain storm (Kumar et al, 2011; Ghada et al., 2019). To model rain non-stationarity on a more data-driven basis, and thereby account for the sub-daily non-stationarities reported above, it has recently been proposed to classify rain fields into rain types (e.g. based on weather radar images) prior to stochastic modelling of rain intensity (Lagrange et al., 2018; Benoit et al., 2018b). Rain fields belonging to the same rain type are then deemed statistically similar, and periods with a constant rain type can be regarded as stationary periods for the simulation of rain intensity.

In this context, the main goal of this paper is to propose a new approach to leverage the use of rain types for encoding non-stationarity in the framework of stochastic weather generators. However, the finality differs from that of classical weather generators (Richardson, 1981; Wilks and Wilby, 1999; Peleg et al., 2017) since we aim at simulating rainfall conditional to already known meteorological covariates, instead of simulating jointly the whole weather (i.e. all variables). More precisely, we develop a method for stochastic simulation of rain type time series conditional to the current state of the atmosphere, i.e. conditional to meteorological variables such as pressure, temperature, humidity or wind (Fig 1a). These meteorological covariates are assumed to be known beforehand, either from observations, numerical weather model outputs, or from other stochastic simulations. The advantage of the proposed approach is twofold: firstly, using a stochastic simulation to generate rain types allows to properly reproduce the natural variability of rain type occurrence, and thereby to indirectly model the non-stationarity of rain statistics observed in historical datasets. Secondly, the conditioning of the stochastic rain type model to the state of the atmosphere preserves the relationships between rain type occurrence and climatological drivers. Once realistic rain type time series have been simulated (i.e. the core of this study, Fig 1a), high-resolution rain fields can be generated conditional to rain types using any high-resolution stochastic rainfall generator (Vischel et al., 2011; Leblois and Creutin, 2013; Paschalis et al., 2013; Nerini et al., 2017; Benoit et al., 2018a) as illustrated in Fig 1b. Using rain types to guide the stochastic generation of synthetic rains has been shown to improve the realism of the resulting high-resolution space-time simulations (Benoit et al., 2018b).

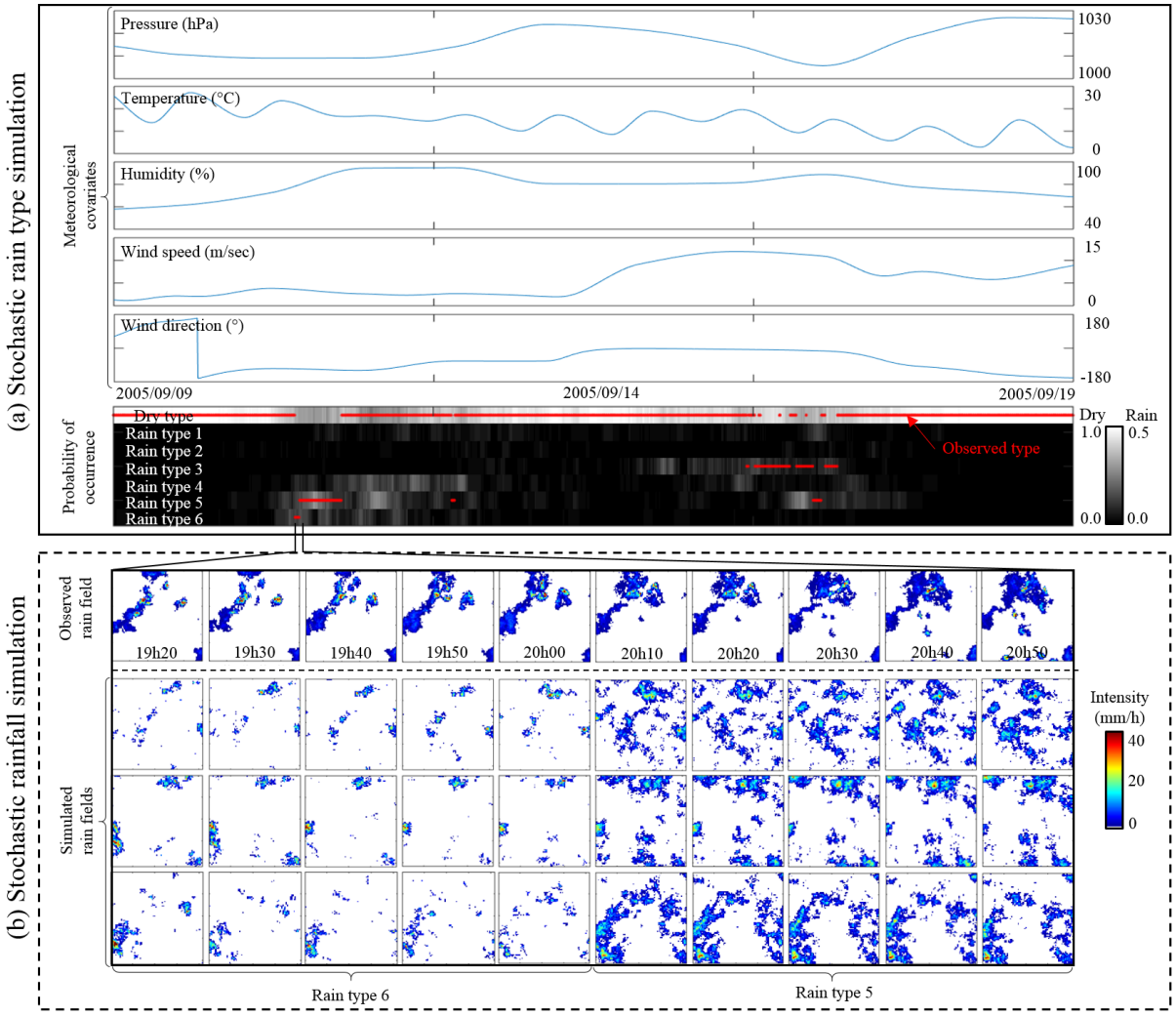

Figure 1. Overview of stochastic rain type generation (core of this study), and its application to simulate high-resolution synthetic rain fields whose statistical properties depend on meteorological conditions. (a) Rain type simulation framework developed in this study. (b) Illustration of stochastic rainfall simulation conditioned to changing rain types.

In the bottom row of (a) the observed rain types are in red, and the gray shaded background denotes the probability of rain type occurrence derived from stochastic rain type simulations conditioned to the meteorological covariates displayed in the 4 top lines.

In (b), the upper row displays actual rain fields observed by radar imagery, and the two bottom rows display two stochastic simulations of synthetic rain fields for the same period generated using the approach of Benoit et al. (2018a).

The remainder of the paper is structured as follows. First, Sect. 2 presents an example of sub-daily rain type time series, and Sect. 3 proposes a stochastic model able to capture the main statistical features of this dataset. Next, Sect. 4 assesses the

performance of the model through a cross-validation procedure and Sect. 5 illustrates the application of the proposed approach to the downscaling of EURO-CORDEX RCM precipitation future projections. Finally, Sect. 6 provides some conclusions about stochastic rain type modelling.

2 Example dataset of rain type time series and related meteorological covariates

Before proposing a stochastic model able to mimic the rain type occurrence process (Sect. 3), the present section explains how rain type time series are derived from weather radar observations, and investigates the main features of rain type occurrence in a mid-latitude climate.

2.1 Rain type time series

We focus hereafter on a 100 km x 100 km squared area centered on the city of Jena in the Land of Thüringen, Germany (Fig. 2a). This area has been chosen because its flat topography and its location far from coastlines or major topographic barriers ensure spatially homogeneous rain fields, allowing to focus on the temporal component of rainfall non-stationarity. Over this area, data used for rain typing consists of radar images extracted from the RADOLAN (RAdar-OnLine-ANeichung) dataset (Winterrath et al., 2012; Kaspar et al., 2013), which is provided in open access by the German meteorological agency (Deutscher Wetterdienst - DWD). It consists of raw (i.e. not adjusted on rain gauges) radar image composites over entire Germany from 01/01/2001 to present. RADOLAN radar image resolution is 1 km x 1km in space, and 5 min in time. In practice, however, we resampled radar images at a 10 min resolution and restricted our study to the period 01/01/2001 – 31/12/2017. The RADOLAN dataset is used as baseline information for rain typing, following a space-time classification approach (Benoit et al., 2018b). Using raw radar images can lead to biases in estimated rain intensities, but the impact of such biases on the classification are deemed negligible since the adopted approach focuses on rainfall space-time behavior rather than rainfall intensity. A problematic source of errors would be the change of radar biases along time, which could alter the inter-annual frequency of rain types. To alleviate this problem, uniformly reprocessed radar images are used as basis for the classification, which ensures a consistent data-cube throughout the period of interest. In practice, no adverse trend is noted in the observed rain type distribution (Fig. 3).

In a nutshell, the classification method proposed by Benoit et al. (2018b) and used hereafter for rain typing consists of the following. First, rainy time steps are defined as periods with more than 10% radar pixels measuring rain. The other time steps are classified as dry, and are not considered for rain typing. For classification, 10 statistical metrics are computed for each rainy radar image in order to assess the space-time-intensity behavior of the rain field observed in the image. Among these metrics, 3 relate to the statistical distribution of rain intensity observed in the radar image, 3 characterize the spatial arrangement of rain patterns within the image, and 4 evaluate the temporal evolution of the rain field due to rain advection and diffusion between consecutive periods. Then, the 10 metrics are used as a basis for classification using a Gaussian Mixture Model (GMM). All details about the 10 metrics and the clustering approach with GMM can be found in Benoit et al. (2018b). The resulting clusters correspond to rain fields with similar space-time behaviors. The number of rain types is selected as a compromise

between goodness of fit to rain field statistics and model parsimony. In the present case, the parsimony is favored in order to allow for a physical interpretation of the resulting rain types. As a result, 6 rain types are identified in the example dataset (Fig. 2b). Among them, rain types 1 and 4 correspond to rather stratiform and spread rain events, rain type 3 corresponds to frontal rain storms, and rain types 5 and 6 can be associated with rather convective rains. Rain type 2 cannot be associated to a specific rain behavior, but rather gather rain fields that are not classified otherwise, and often correspond to partial rain coverage.

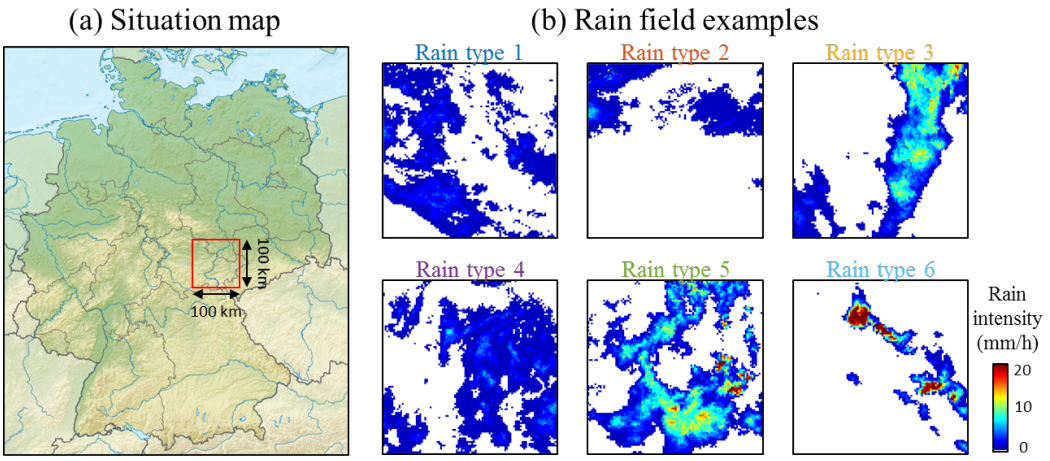

Figure 2. Radar dataset used for rain typing. (a) Study area. The red square denotes the area of interest, centered on Jena, Thüringen, Germany. (Background map from Wikipedia.org, licensed under CC BY 3.0) (b) Example of RADOLAN radar images (cropped over the area of interest) for each rain type.

Using only radar images with more than 10% wet pixels to define rain types ensures a reliable classification, but at the cost of a dry bias (in the present dataset 32% of the images have a rain fraction between 0% and 10%, and encompass 19% of the rain total, cf Fig. 3). To deal with images with less than 10% wet pixels, Benoit et al. (2018b) proposed to classify images with a small rain fraction (i.e. 0% < rain fraction < 10%) in a second step by assigning them the type of the closest classified image (i.e. nearest neighbor interpolation in time). This post-processing scheme is not directly transferable to the context of simulation because no information about low rain coverage images is available in simulation outputs. Two options can be considered to alleviate this problem. First, the rain type model defined in Sect 3 can be calibrated on the final classification (i.e. including images with low rain coverage), which results in simulations that preserve the actual rain proportion. However, using a classification that includes the onset and the end of rain storms leads to less clear relationships between climate covariates and rain type occurrence, which may degrade simulation results. Hence the second option, which we follow in this paper, which consists in (1) calibrating and running the rain type model for rain types defined only from radar images with more than 10% rain coverage, and next (2) re-adjusting the wet/dry balance by post-processing. The dry bias is corrected assuming the ratio $R = \frac{N_s}{N_l}$ between the number N_s of images with low rain coverage and the number N_l of images with large rain coverage

as constant in observations and simulations. Subsequently, the number of epochs for which rain is simulated is increased by propagating the closest rain type to the $R \times N_l$ dry epochs located at the beginning and at the end of rain storms. Supplementary material 1 shows that such post-processing performs well to mitigate the dry bias originating from the use of a 10% rain coverage threshold to define a wet image. However, since the present study focuses on climate - rain type relationships, which are better defined when considering only the first step of the classification, the aforementioned post-processing is not applied in the remainder of this paper. Hence, one should keep in mind that in the following the dry type also includes epochs with a low rain coverage (under 10%), and that a post-processing is required if the end-use application involves stochastic simulation of actual rain fields.

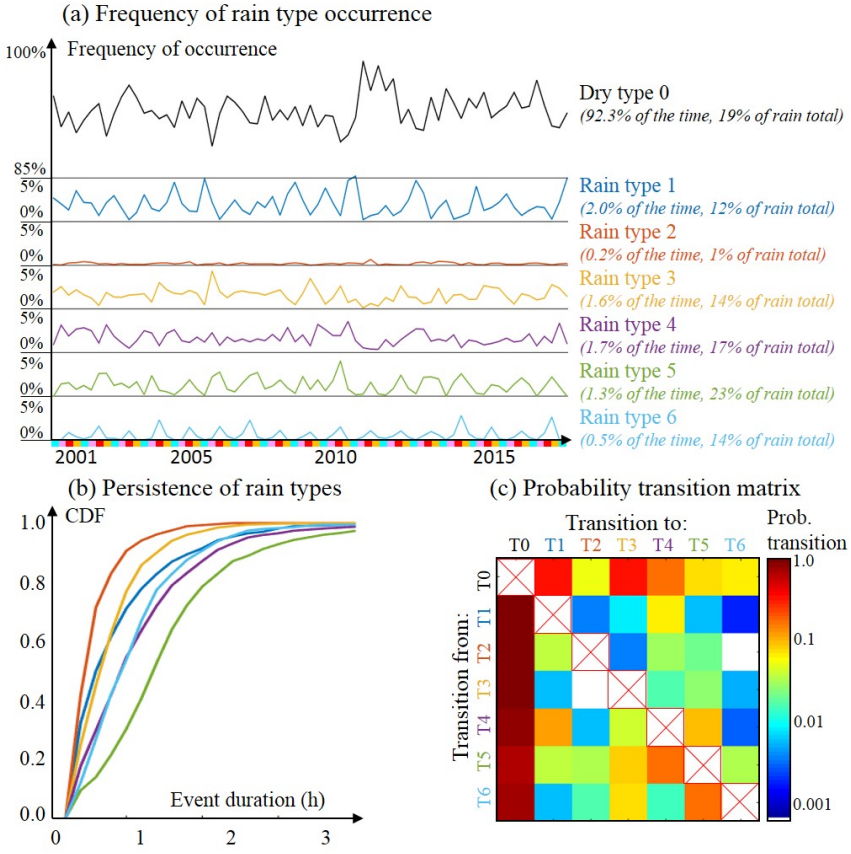

Figure 3. Main features of a rain type time series (2000-2017) observed over central Germany. (a) Frequency of rain type occurrence computed at a seasonal basis (Seasons are DJF (light blue), MAM (pink), JJA (red) and SON (yellow)). (b) CDF of event duration stratified by rain type. (c) Empirical matrix of transition probability between rain types.

Fig. 3 investigates the occurrence of rain types through time, and highlights some features. Fig. 3a displays the frequency of each rain type at the seasonal scale. It appears from Fig. 3a that the frequency of occurrence of individual rain types is strongly variable across the year. For example, stratiform rain types 1 and 4 occur mostly in winter, while convective rain types 5 and

6 are most common in summer. In addition, one can notice a strong inter-annual variability in rain occurrence, summers 2003 and 2011 having a low occurrence of rain, while rain occurrence is particularly high during winters 2006 and 2011. Figure 3b displays the Cumulative Density Function (CDF) of the duration of each rain type. Here rain type duration is defined as the duration (i.e. length along the time axis) of a segment of constant rain type. Each curve in Fig. 3b therefore corresponds to the probability that a rain event of a given type does not exceed the duration given in abscissa. This figure shows that all rain types are persistent in time with durations ranging from few minutes to more than 3 hours, and that some types (e.g rain types 4 and 5) are more persistent than others (e.g. rain types 2 and 3). Finally, Fig. 3c displays the empirical transition matrix between rain types, and focuses on inter-type transitions (i.e. transitions to the same type are ignored and denoted by red crosses). This figure shows that the patterns of transition between rain types are complex, and that the transitions involving type 0 (i.e. no rain) are largely dominant.

2.2 Meteorological covariates

The strong seasonality and inter-annual variability of rain type occurrence emerging from Fig. 3a can be explained to a large extent by regional meteorological conditions. Hereafter, we investigate the links between rain type occurrence and a set of 5 meteorological covariates that are deemed to influence rainfall behavior (Vrac et al., 2007; Willems, 2001; Rust et al., 2013), namely: sea level pressure, temperature at 2 m, relative humidity, and direction and intensity of synoptic wind at 850 hPa. The actual values of the meteorological covariates used in this study are extracted from the ERA-5 reanalyzes (Hersbach et al., 2018) and averaged over the whole area of interest before further use. Only parameters provided at daily resolution are used hereafter in order to be compatible with the temporal resolution of RCM projections used for the illustration of our framework (see Sect. 5). To match the resolution of rain type data, the above daily-resolution meteorological covariates are disaggregated to a 10 min resolution. To this end, the mean daily pressure, humidity, wind direction and wind intensity are assumed to occur at 12 PM local time, and are then interpolated at a 10 min resolution using a polynomial interpolation. For temperature, the daily minimum is assumed to occur at 5 AM local time, the daily maximum at 3 PM, and the diurnal cycle is captured by spline interpolation. This disaggregation framework leads to 10-min resolution meteorological covariates in good agreement with actual 10-min resolution observations carried out by a weather station located close to the center of the area of interest, and rain type simulations conditioned to these two sets of covariates (in-situ observations and disaggregated reanalysis data) give very similar results (Supplementary material 2, Fig. 3).

Figure 4 displays the relationships between meteorological covariates and rain type occurrence, and confirms the strong influence of temperature and wind speed on rain types. Indeed, stratiform rain types 1 and 4 occur at lower temperatures than convective types 5 and 6, and co-occur with stronger winds. The frontal rain type 3 is characterized by strong westerlies, but occurs for a broad range of temperatures. In contrast to temperature and wind speed, the standalone knowledge of pressure, relative humidity or wind direction does not allow to discriminate between all rain types. However, when considered jointly (Supplementary material 2, Fig. 4), all meteorological covariates bring information about rain type occurrence. In particular, pressure and humidity are key drivers for rain occurrence (no matter the type) and are therefore useful to predict dry and wet spells. Furthermore, wind direction informs the occurrence of rain type 3 and helps discriminating between rain types 5 and 6.

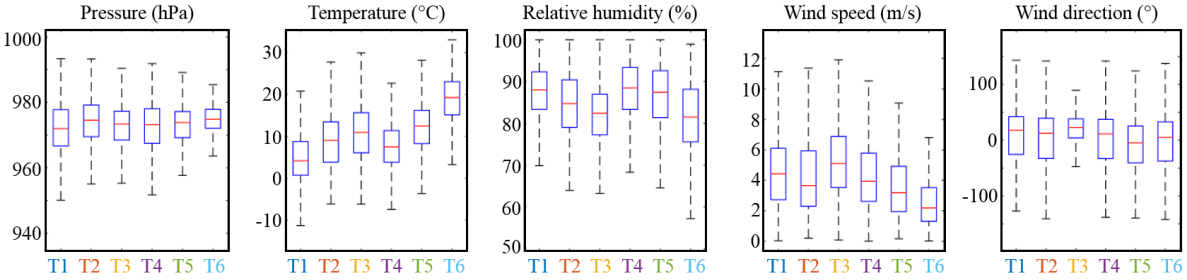

Figure 4. Statistics of meteorological covariates for each rain type. The meteorological data have been extracted from the ERA-5 reanalysis dataset.

3 Stochastic rain type model

Daily resolution stochastic weather generators do not distinguish between rain types and typically resort to Markov-chain models to simulate rain occurrence (Richardson, 1981; Wilby, 1994; Wilks and Wilby, 1999). Within Markov-chain models, semi-Markov models are often favored when the persistence of wet and dry spells is of prime importance (Foufoula-Georgiou and Lettenmaier, 1987; Bárdossy and Plate, 1991), and non-homogeneous Markov models are preferred when the dry/wet sequence has to be conditioned to meteorological covariates (Hughes and Guttorp, 1999; Vrac et al., 2007). An easy option to deal with rain type simulation conditional to meteorological covariates would therefore be to extend one of these Markov-chain-based frameworks by simply increasing the number of states of the Markov chain in order to account for the diversity of rain types. However, both rain type persistence and conditioning to covariates are equally important for the targeted application, which led us to combine both frameworks to build a non-homogeneous semi-Markov model. Unfortunately, semi-Markov models do not allow for an easy conditioning to continuous-time covariates, and our attempt to build a non-homogeneous semi-Markov model led to a significant dry bias in rain type simulations (Supplementary material 3). Hence, it appeared that even relatively sophisticated parametric models are challenged by the complexity of rainfall at sub-daily resolution (Oriani et al., 2018).

One alternative to account for such complexity is to consider high-order properties through a non-parametric approach based on the resampling of historical datasets (Oriani et al., 2014). For the simulation of rain types conditional to meteorological covariates, we therefore adopt the framework of multiple-point simulations (MPS). MPS consist of using a training dataset (here a past rain type record) to estimate empirically the probability distribution of the variable of interest (here rain type occurrence at a given time step) conditionally to the values already simulated in its temporal neighborhood (Fig. 5a). In the specific MPS algorithm we use (Gravey and Mariethoz, 2018, Under Review), the conditional pdf is indirectly assessed by making a random sampling of the training dataset that aims at finding a pattern that is similar to the local conditioning neighborhood (Fig. 5b). In practice, we use a 100 h neighborhood for the present application. Once a match is found (i.e. a pattern in the training image that minimizes the Hamming distance with the target pattern), the corresponding value is imported in the simulation grid (Fig. 5c), and the procedure is iterated until the full simulation grid is filled. In the MPS framework, the

dependence of rain type occurrence to meteorological covariates can be handled by multivariate-MPS simulation (Mariethoz et al., 2010). It consists of stacking time series of meteorological covariates with the time series of rain type occurrence, and to evaluate the conditioning neighborhood on the time series of the resulting vector-variable. Here we use a simplified version of multivariate-MPS where only the co-located covariates (i.e. the values of the covariates observed at the exact time step to simulate) are accounted for during the matching procedure.

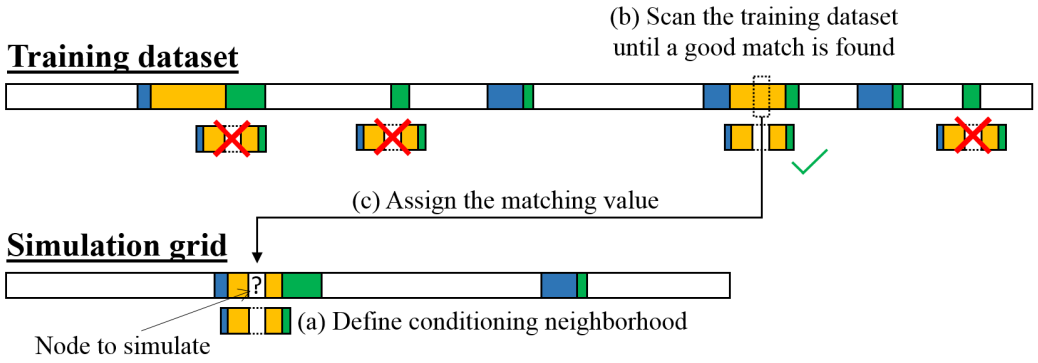

Figure 5. Schematic view of the MPS algorithm used for non-parametric resampling of historical rain type time series.

Since MPS is a non-parametric approach, it does not require model calibration strictly speaking. Instead, it requires a training dataset to resample, which should include both the variable of interest (here a rain type time series) and optional covariates (here meteorological covariates). To produce reliable results, and in particular meaningful uncertainty estimates, MPS requires large training datasets (Emery and Lantuéjoul, 2014). In the present case, the training dataset is the historical record of joint rain types and meteorological covariates observations available over the target area (17 years). After selection of a training dataset, simulations are obtained by resampling the training dataset using the MPS algorithm described above.

4 Model assessment

4.1 Cross-validation

Model performance is assessed by cross-validation, using the dataset introduced in Sect. 2. In practice, we adopt a leave-one-year-out procedure. For a given simulation year, the rain type model is first trained using data from the 2001-2017 period, excluding the year to simulate. Next, 50 realizations of rain type time series are generated for the year of interest by MPS simulation, conditioned to observations of the meteorological covariates derived from the ERA5 reanalysis as described in Sect. 2. Finally, the same procedure is iterated for each year of the test period (i.e. 2001-2017), and 50 realizations of 17-years-long rain type time series are obtained by concatenating in time the 17 yearly simulations.

The 50 simulations are compared to the reference rain type time series in Fig. 6. Focusing first on the ability of the model to simulate dry and wet conditions (i.e. without distinction between rain types), the first panel of Fig. 6a compares the observed

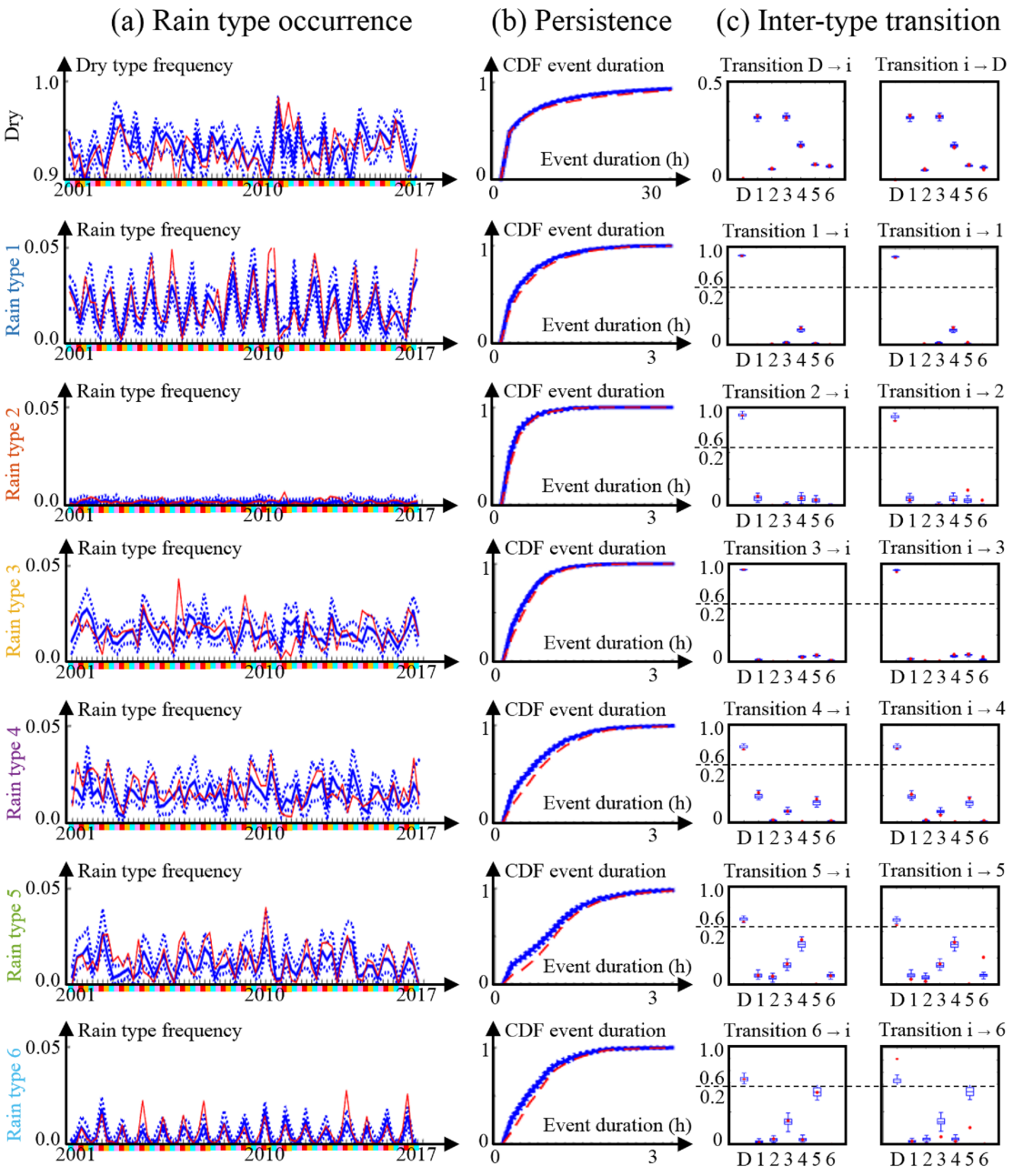

Figure 6. Results of the cross-validation experiment. (a) Seasonality of rain (and dry) type occurrence (Seasons are DJF (light blue), MAM (pink), JJA (red) and SON (yellow)), (b) rain type persistence, and (c) probability of transition between rain types. Observations are in red and simulations in blue. In simulations, continuous lines represent the median of the simulated ensembles (50 realizations), and dashed lines represent the Q10 and Q90 quantiles.

and simulated frequencies of dry occurrence for each season of the validation period 2001-2017. The results show that our model properly simulates the overall proportion of rain (ratio simulated/observed rain frequency = 0.93). In addition, the chronology of dry/wet occurrence at the seasonal scale is reasonably simulated (correlation between observed and simulated dry type occurrence=0.6).

Focusing next on the simulation of rain types, Fig. 6 (lines 2-7) assesses the ability of the model to reproduce the typical features of rain type occurrence highlighted in Sect. 2, namely seasonality, persistence and transition. Figure 6a assesses rain type seasonality and shows that for strongly seasonal rain types (types 1, 5 and 6) the annual cycle of rain type occurrence is properly simulated. It is also worth noting that the inter-annual variability of the annual cycle is reasonably well simulated, as well as the inter-annual variability of weakly seasonal rain types (types 3 and 4). This shows that the proposed approach not only reproduces the annual cycle of rain type occurrence driven by monthly-scale variations of the covariates, but also captures the impact of short term fluctuations of meteorological conditions that trigger rain storms of type 3 and 4. However, when scrutinizing the minima and maxima in Fig 6a, one can notice that peaks in observations are smoothed out in simulations, which traduces the difficulty of the model to simulate extreme cases. This is a known drawback of MPS simulations (Mariethoz and Caers, 2015), and it constitutes the main limitation of the present approach for the simulation of future climates in which unusual climatic conditions are deemed to become more frequent. Regarding rain type persistence, Fig 6b shows that this feature is in general very well reproduced, except for long lasting types 4 and 5 for which persistence is slightly underestimated. Finally Fig. 6c assesses inter-type transitions. One can notice that the transition patterns are almost perfectly reproduced, except for the transition from dry (and to a lesser extend from type 5) to rain type 6 that is underestimated (respectively overestimated). Overall, the proposed stochastic rain type model properly reproduces the main features of observed rain type time series. This good performance is linked to the ability of MPS to accurately reproduce high-order statistics of rain type time series.

4.2 Sensitivity to climate variability

To ensure that the proposed rain type model is able to capture the impact of climatic signals on rain type occurrence, the results of the cross-validation procedure are stratified according to annual climatic signatures. To this end, Fig. 7 compares simulated and observed rain type occurrences at the monthly scale for four sub-datasets: the 5 coldest years of the 2001-2017 period (Fig. 7a), the 5 warmest years (Fig. 7b), the 5 driest years (Fig. 7c), and finally the 5 wettest years (Fig. 7d). Observations (red curves in Fig. 7) show that years with different climatic signatures indeed develop distinct dry/wet ratios and rain type distributions. Simulation results (blue curves in Fig. 7) show that the proposed model properly reproduces these climatically-driven differences in rain type distribution.

Figure 7 therefore allows for a detailed investigation of the impact of climatic signals on local rain type distribution over central Germany. Comparing first cold and warm years (Fig. 7a-b), one can notice that warm years tend to be drier, in particular in late winter and spring. This is mostly caused by a deficit of type 1 precipitations during warm years, which correspond to less snow (type 1 occurs mostly at low temperatures, cf Fig. 4). Drier springs during warm years are also caused by a deficit of type 5 (slightly convective), which probably correspond for this time of the year to rain and sleet showers. Finally, one can notice an increased occurrence of type 6 (strongly convective) during warm years, which is captured by the model despite a

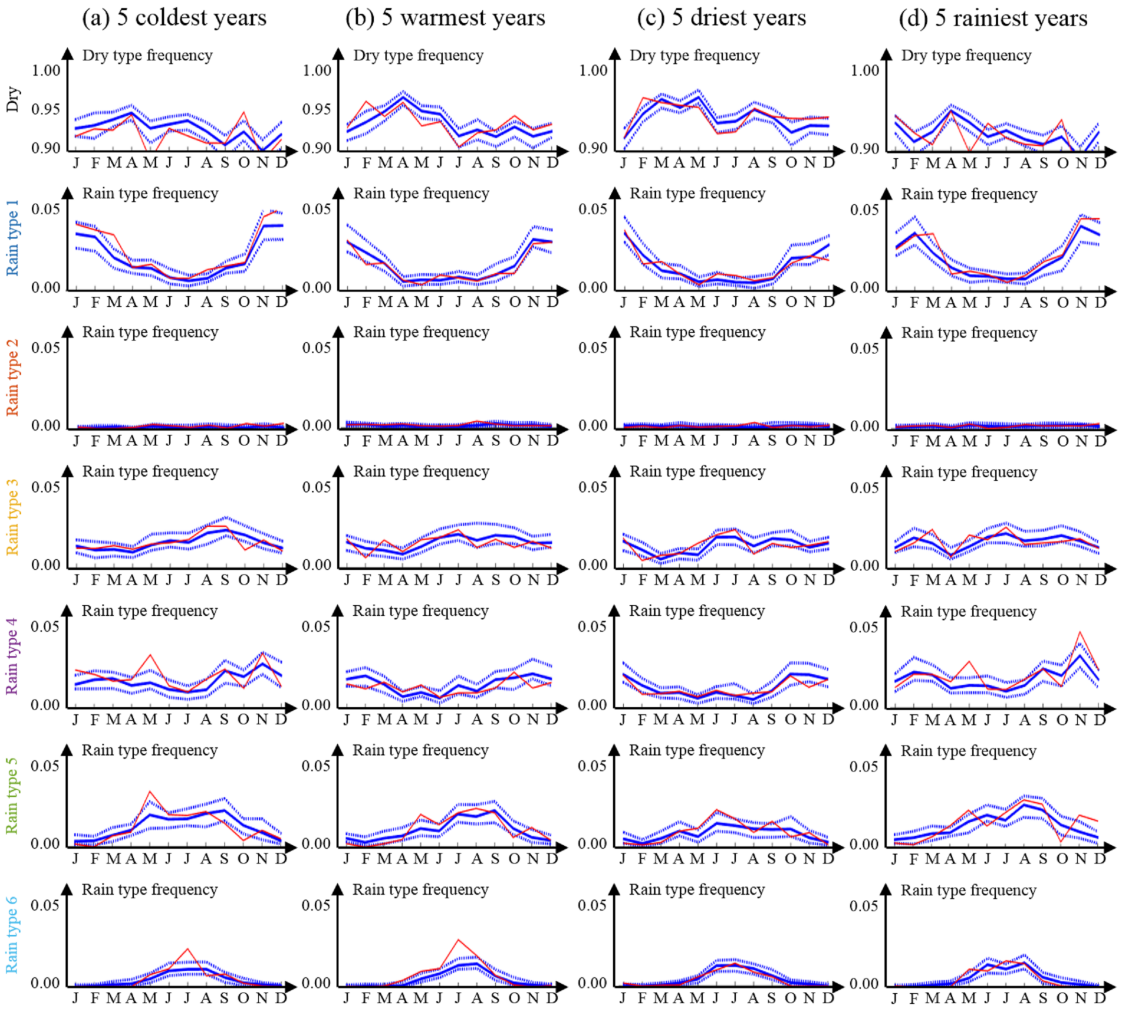

Figure 7. Monthly rain type occurrence stratified according to climate forcing: (a) 5 coldest years of the 2001-2017 period (2001, 2004, 2005, 2010, 2013), (b) 5 warmest years (2007, 2011, 2014, 2015, 2017), (c) 5 driest years (2003, 2011, 2012, 2015, 2016), and (d) 5 wettest years (2001, 2002, 2007, 2009, 2010). Observations are in red and simulations in blue. In simulations, continuous lines represent the median of the simulated ensembles (50 realizations), and dashed lines represent the Q10 and Q90 quantiles.

slight underestimation of this type for both cold and warm years. Comparing next dry and wet years (Fig. 7c-d), one can notice that all months contribute to the rain imbalance, but that the rain deficit is more pronounced in spring and autumn. In terms of rain type distribution, this is mostly caused by a deficit of rain types 1 and 4 (stratiform) as well as 5 (slightly convective) during dry years. It is also worth noting that rain type 3 (frontal) tends to be slightly more common during rainy years, but in contrast to other types its increased occurrence is spread along the whole year.

Overall, the proposed approach properly captures the impact of climatic signals on rain type occurrence. This property is

essential to preserve the relationships between rain types and climatological drivers, and paves the way to RCM precipitation downscaling.

5 Application to RCM precipitation downscaling

For illustration purposes, the stochastic rain type model developed in Sect. 3 is used to simulate the evolution of rain type occurrence in a changing climate simulated by one RCM run extracted from the EURO-CORDEX climate downscaling experiment (Jacob et al., 2014). To drive the simulation of rain types in a changing climate, the same set of meteorological covariates as the one used for cross-validation is derived from one RCM run, more precisely from the Regional Atmospheric Climate MOdel of the Dutch national weather service (RACMO-KNMI (Van Meijgaard et al., 2008)) driven by the CNRM-CM5 Earth system model (Voldoire et al., 2013) forced according to the RCP8.5 emission scenario. Three intervals of 20 years each are selected to investigate the evolution of rain type occurrence over the 21^{st} century: 1997-2017 (reference period that encompasses the 2001-2017 calibration period for which rain type observations are available), 2037-2057, and 2077-2097. For each period, the meteorological covariates are extracted from the RCM simulation, averaged over the area of interest, and disaggregated at a 10 min resolution as described in Sect. 2. In addition, RCM outputs are bias-corrected using the CDF-t method for each variable separately (Vrac et al., 2012). After bias-correction of RCM data, the performance of the model to simulate rain types in the present climate is almost identical for meteorological covariates derived from the RACMO-KNMI RCM and the ones derived from the ERA-5 reanalysis (Supplementary material 4).

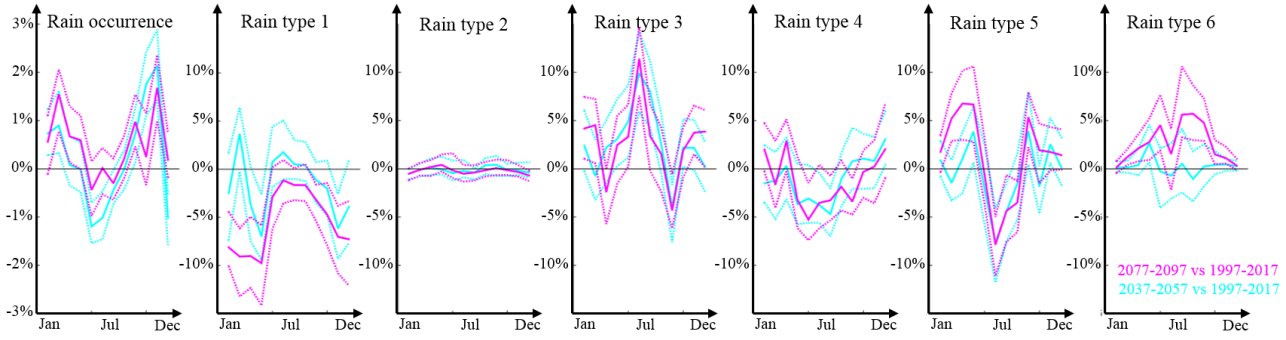

Figure 8. Changes in rain occurrence frequency (left panel) and rain type distribution conditional to the presence of rain (other panels) simulated using the stochastic rain type model developed in Sect. 3: 2037-2057 vs 1997-2017 (light blue) and 2077-2097 vs 1997-2017 (purple). Continuous lines represent the median of the simulated ensembles (50 realizations), and dashed lines represent the Q10 and Q90 quantiles.

For each 20-year period, 50 realizations are simulated conditional to bias-corrected RCM-derived meteorological covariates. To evaluate the projected changes in rain type distribution, Fig. 8 displays the evolution of the monthly frequency of rain type occurrence between the reference period 1997-2017 and the two future periods 2037-2057 and 2077-2097. Observed changes

in rain type occurrence frequency are considered as significant if they exceed the uncertainty of the projection that is defined as the Q10-Q90 interval of the 50 realizations. It should be noted that in contrast to rain type occurrence, the simulated persistence and transition behavior of rain types remain constant over the whole test period (not shown). Results in Fig. 8 show that the frequency of rain occurrence slightly decreases in summer and increases in winter, spring and autumn. Among these changes, only the increase of rain occurrence during autumn and winter is significant. The distribution of rain types conditional to the presence of rain is more significantly modified than rain occurrence. More precisely, during winter, rain type 1 (stratiform) significantly declines while the frequency of rain type 3 (frontal) and 5 (moderately convective) significantly increases. During spring and fall, rain types 1 (stratiform) significantly declines while rain types 5, 6 (convective) tend to increase but not significantly. Finally, during summer, rain types 1, 4 (stratiform) and 5 (moderately convective) decrease while rain types 3 (frontal) and 6 (strongly convective) increase, most of these changes being significant. Overall, it appears from this exploratory study that under the assumption of the specific RCM run used to simulate the meteorological covariates, convective and frontal rains could become more frequent at the expense of stratiform rains by the end of the 21^{st} century. The most significant changes are obtained during winter and summer. It is worth mentioning that the evolution of rain behavior along the 21^{st} century simulated in the present study is qualitatively in line with results obtained over Western Europe by studies using physical models, which anticipate more frequent heavy rains driven by convection or active fronts (Molnar et al., 2013; Faranda et al., 2019) at the expense of low intensity stratiform precipitations.

6 Concluding remarks

6.1 Discussion

By introducing a step of rain type simulation in the framework of stochastic rainfall generators, we suggest that for high temporal resolution applications, the simulation of rain can be split in two steps (Fig. 1). In a first instance, rain types are simulated conditional to meteorological covariates to account for the diversity of rain storms at the regional scale. This first step is the main focus of the present paper. For subsequent applications, we assume that the rain intensity can be simulated conditional to rain types. This is the classical aim of space-time distributed stochastic rainfall generators, which are becoming more and more common to address the needs of high-resolution hydrometeorological impact studies (Vischel et al., 2011; Leblois and Creutin, 2013; Paschalis et al., 2013; Nerini et al., 2017; Benoit et al., 2018a).

Hence, two main applications can be considered for stochastic rain type simulation. The first one, briefly illustrated in Sect. 5, consists of assessing the evolution of the statistical signature of rainfall in a changing climate simulated by RCMs. It is worth noting that if one want to carefully evaluate the change in rain type occurrence that may emerge in the future, one should rely on a large ensemble of RCM-GCM-Emission scenarios combinations to properly capture the uncertainty on meteorological covariates. In addition, one should keep in mind that the present approach only accounts for changes in the distribution of existing rain types, and therefore ignores the possible emergence of new rain types in response to climate conditions that have never been observed over the area of interest. Such new rain types could potentially by modelled by reparametrizing the stochastic rainfall model used to simulate local rain fields (Peleg et al., 2019), by using rain analogs from areas that experience

today the climate that is simulated in the future over the area of interest (Hallegatte et al., 2007; Fitzpatrick and Dunn, 2019), or by running a convection-permitting climate model (Prein et al., 2015) for the newly emerging climate conditions. The development of a framework to model emerging rain types is however left for future research.

The second application is the simulation of rain intensity at high space-time resolution while preserving consistency with climatological drivers such as temperature, pressure, humidity and wind. As mentioned in the introduction, simulating rain intensity would require setting up and calibrating a high-resolution stochastic rainfall model for each rain type over the area of interest, and was therefore not considered in the present study except in Fig 1b for illustration purposes. However, two advantages are expected to emerge from adding a rain type simulation step into stochastic rainfall modelling: first, a relatively low number of rain types can be specified, which implies that the model of rain intensity has to be calibrated a limited number of times. This ensures that enough observations are available to calibrate the rain intensity model for each rain type, and therefore prevents model overfitting. The second advantage is the added flexibility to simulate rain storm dynamics, which allows to generate intra-storm variations of the space-time rainfall statistics.

6.2 Outlook

In this paper, a non-parametric approach based on the resampling of historical records using multiple-point statistics has been proposed and thoroughly tested for the simulation of rain types time series conditional to meteorological covariates. Evaluation results based on a 17-year long rain type dataset in a mid-latitude climate (central Germany) show that MPS simulations are able to reproduce both the internal variability of rain type time series, as well as relationships with meteorological covariates. After validation, stochastic rain type simulation is applied to the downscaling of RCM projections over the 21^{st} century. Rain type simulations conditioned to meteorological covariates simulated by a Regional Climate Model under an RCP8.5 emission scenario indicate a possible change in rain type distribution by the end of the century, with an increased frequency of heavy rains driven by convection or active fronts, and a decline of low intensity stratiform precipitations.

The ability of stochastic simulations to generate realistic rain type time series when conditioned to meteorological covariates advocates for including stochastic rain type simulation into rainfall generators in order to: (1) reproduce the internal variability or rain type occurrence, in particular inter-annual variability, seasonality, persistence and inter-type transitions, and (2) preserve the relationships between rain statistics and meteorological covariates, in the present case temperature, pressure, humidity and wind. The above features make stochastic rain type simulation a convenient tool to account for the non-stationnarity of rain statistics driven by meteorological conditions. This opens the door to sub-daily stochastic downscaling of climate projections, and to improved stochastic rainfall simulations.

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

Acknowledgements. All data and codes used in this study are open source and freely available in the following repositories:

- Radar data: https://opendata.dwd.de/climate_environment/CDC/grids_germany/5_minutes/radolan.
- Rain type data: https://github.com/LionelBenoit/Stochastic_Raintype_Generator/Raintype_data.
- Rain typing software: https://github.com/LionelBenoit/Rain_typing.
- Stochastic rain type models: https://github.com/LionelBenoit/Stochastic_Raintype_Generator/codes.
- MPS simulation software: https://github.com/GAIA-UNIL/G2S.

References

Ailliot, P., Allard, D., Monbet, V. and Naveau, P.: Stochastic weather generators: an overview of weather type models. Journal de la société franąise de statistiques, 156(1): 101-113, ISSN: 2102-6238, 2015.

Bárdossy, A. and Plate, E.J.: Modelling daily rainfall using a semi-Markov representation of circulation pattern occurence, Journal of Hydrology, 122, 33-47, doi: 10.1016/0022-1694(91)90170-M, 1991.

Bárdossy, A. and Pegram, G. G. S.: Space-time conditional disaggregation of precipitation at high resolution via simulation. Water Resources Research, 52, 920-937, doi: 10.1002/2015WR018037, 2016.

Benoit, L., Allard, D. and Mariethoz, G.: Stochastic Rainfall Modelling at Sub-Kilometer Scale, Water Resources Research, 54, 4108-4130, doi:10.1029/2018WR022817, 2018a.

Benoit, L., Vrac, M. and Mariethoz, G.: Dealing with non-stationarity in sub-daily stochastic rainfall models, Hydrology and Earth System Sciences, 22, 5919-5933, doi:10.5194/hess-22-5919-2018, 2018b.

Burton, A., Fowler, H. J., Blenkinsop, S. and Kilsby, C. G.: Downscaling transient climate change using a Neyman-Scott Rectangular Pulses stochastic rainfall model. Journal of Hydrology, 381, 18-32, doi: 10.1016/j.jhydrol.2009.10.031, 2010.

Caseri, A., Javelle, P., Ramos, M. H., and Leblois, E.: Generating precipitation ensembles for flood alert and risk management. Journal of Flood Risk Management, 9, 402-415, doi: 10.1111/jfr3.12203, 2016.

Emery, X. and Lantuéjoul, C.: TBSIM: A computer program for conditional simulation of three-dimensional Gaussian random fields via the turning bands method. Computers and Geosciences, 32, 1615-1628, doi: 10.1016/j.cageo.2006.03.001, 2014.

Emmanuel, I., Andrieu, H., Leblois, E., and Flahaut, B.: Temporal and spatial variability of rainfall at the urban hydrological scale. Journal of Hydrology, 430-431, 162-172, doi: 10.1016/j.jhydrol.2012.02.013, 2012.

Faranda, D., Alvarez-Castro, C., Messori, G., Rodrigues, D., and Yiou, P.: The hammam effect or how a warm ocean enhances large scale atmospheric predictability, Nature communications, 10, Article number: 1316, doi: 10.1038/s41467-019-09305-8, 2019.

Foufoula-Georgiou, E., and Lettenmaier, D.: A Markov Renewal Model for Rainfall Occurence, Water Resources Research, 23, 875-884, doi: 10.1029/WR023i005p00875, 1987.

Fitzpatrick, M.C., and Dunn, R.R.: Contemporary climatic analogs for 540 North American urban areas in the late 21st century, Nature Communications, 10, 614, doi: 10.1038/s41467-019-08540-3, 2019.

Furrer, E.M. and Katz, R.W.: Generalized linear modeling approach to stochastic weather generators, Climate Research, 34, 129-144, doi: 10.3354/cr034129, 2007.

Ghada, W., Estrella, N. and Menzel, A.: Machine Learning Approach to Classify Rain Type Based on Thies Disdrometers and Cloud Observations, Atmosphere, 10, article number: 251, doi: 10.3390/atmos10050251, 2019.

Gires, A., Tchiguirinskaia, I., Schertzer, D., Schellart, A., Berne, A. and Lovejoy, S.: Influence of small scale rainfall variability on standard comparison tools between radar and rain gauge data, Atmospheric Research, 138, 125-138, doi:10.1016/j.atmosres.2013.11.008, 2014.

Gravey, M. and Mariethoz, G.: Quantile Sampling: a new approach for multiple-point statistics simulation, IAMG 2018 Conference - Olomouc - Czech Republic, 2018.

Gravey, M. and Mariethoz, G.: Quantile Sampling: a robust and simplified pixel-based multiple-point simulation approach, Geoscientific Model Development, Under Review.

Hallegatte, S., Hourcade, J.C., Ambrosi, P.: Using climate analogues for assessing climate change economic impacts in urban areas, Climatic Change, 82, 47-60, doi: 10.1007/s10584-006-9161-z, 2007.

Hersbach, H., de Rosnay, P., Bell, B., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Alonso-Balmaseda, M., Balsamo, G., Bechtold, P., Berrisford, P., Bidlot, J-R., de Boisséson, E., Bonavita, M., Browne, P., Buizza, R., Dahlgren, P., Dee, D., Dragani, R., Diamantakis, M., Flemming, J., Forbes, R., Geer, AJ., Haiden, T., Hólm, E., Haimberger, .L., Hogan, R., Horányi, A., Janiskova, M., Laloyaux, P., Lopez, P., Munoz-Sabater, J., Peubey, C., Radu, R., Richardson, D., Thépaut, J-N., Vitart, F., Yang, X., Zsótér, E., Zuo, H.: Operational global reanalysis: progress, future directions and synergies with NWPRep, ECMWF report, ERA Report Series, 2018.

Hughes, J.P. and Guttorp, P.: A non-homogeneous hidden Markov model for precipitation occurence, Applied Statistics, 48, 15-30, doi: 10.1111/1467-9876.00136, 1999.

Jacob, D. et al.: EURO-CORDEX: new high-resolution climate change projections for European impact research, Regional Environmental Change, 14, 563-578, 2014.

Kaspar, F., Muller-Westermeier, G., Penda, E., Machel, H., Zimmermann, K., Keiser-Weiss, A. and Deutschlander, T.: Monitoring of climate change in Germany - data, products and services of Germany's National Climate Data Centre, Advances in Science and Research, 10, 99-106, doi: 10.5194/asr-10-99-2013, 2013.

Krajewski, W.F., Ciach, G. and Habib, E.: An analysis of small-scale rainfall variability in different climatic regimes, Hydrological Sciences Journal, 48, 151-162, doi: 10.1623/hysj.48.2.151.44694, 2003.

Kumar, L.S., Lee, Y.H., Yeo, J.X. and Ong, J.T.: Tropical rain classification and estimation of rain from Z-R relationships, Progress In Electromagnetics Research, 32, 107-127, doi: 10.2528/pierb11040402, 2011.

Lagrange, M., Andrieu, H., Emmanuel, I., Busquets, G., Loubrié, S.: Classification of rainfall radar images using the scattering transform, Journal of Hydrology, 556, 972-979, doi:10.1016/j.jhydrol.2016.06.063, 2018.

Leblois, E., Creutin, J. D.: Space-time simulation of intermittent rainfall with prescribed advection field: Adaptation of the turning band method, Water Resources Research, 49, 3375-3387, doi: 10.1002/wrcr.20190, 2013.

Mariethoz, G., Renard, P. and Straubhaar, J.: The Direct Sampling method to perform multiple-point geostatistical simulations, Water Resources Research, 46, W11536, doi:10.1029/2008WR007621, 2010.

Mariethoz, G., and Caers, J.: Multiple-Point Geostatistics: Stochastic Modeling with Training Images, WILEY Blackwell, 364 pp, 2015.

Marra, F. and Morin, E.: Autocorrelation structure of convective rainfall in semiarid-arid climate derived from high-resolution X-Band radar estimates, Atmospheric Research, 200, 126-138, doi:10.1016/j.atmosres.2017.09.020, 2018.

Mascaro, G., Deidda, R., Hellies, M.: On the nature of rainfall intermittency as revealed by different metrics and sampling approaches, Hydrology and Earth System Sciences, 17, 355-369, doi:10.5194/hess-17-355-2013, 2013.

Mavromatis, T. and Hansen, J. W.: Interannual variability characteristics and simulated crop response of four stochastic weather generators. Agricultural and Forest Meteorology, 109, 283-296, doi: 10.1016/S0168-1923(01)00272-6, 2001.

Molnar, P., Fatichi, S., Gaàl, L., Szolgay, J. and Burlando, P.: Storm type effects on super Clausius–Clapeyron scaling of intense rainstorm properties with air temperature, Hydrology and Earth System Sciences, 19, 1753–1766, doi: 10.5194/hess-19-1753-2015, 2013.

Nerini, D., Besic, N., Sideris, I.V., Germann, U. and Foresti, L.: A non-stationary stochastic ensemble generator for radar rainfall fields based on the short-space Fourier transform, Hydrology and Earth System Sciences, 21, 2777-2797, doi: 10.5194/hess-21-2777-2017, 2017.

Oriani, F., Mehrotra, R., Mariethoz, G., Straubhaar, J., Sharma, A. and Renard, P.: Simulation of rainfall time series from different climatic regions using the direct sampling technique, Hydrology and Earth System Sciences, 18, 3015-3031, doi: 10.5194/hess-18-3015-2014, 2014.

Oriani, F., Straubhaar, Renard, P. and Mariethoz, G.: Simulating rainfall time-series: how to account for statistical variability at multiple scales?, Stochastic Environmental Research and Risk Assessment, 32, 321-340, doi: 10.1007/s00477-017-1414-z, 2018.

Pardo-Igúzquiza, E., Grimes, D. I. F. and Teo, C. K.: Assessing the uncertainty associated with intermittent rainfall fields. Water Resources Research, 42, doi:10.1029/2004WR003740, 2006.

Paschalis, A., Molnar, P., Fatichi, S., and Burlando, P.: A stochastic model for high-resolution space-time precipitation simulation. Water Resources Research, 49, doi:10.1002/2013WR014437, 2013.

Paschalis, A., Fatichi, S., Molnar, P., and Burlando, P.: On the effects of small scale space-time variability of rainfall on basin flood response. Journal of Hydrology, 514, 313-327, doi: 10.1016/j.jhydrol.2014.04.014, 2014.

Peleg, N., Fatichi, S., Paschalis, A., Molnar, P. and Burlando, P.: An advanced stochastic weather generator for simulating 2-D high-resolution climate variables, Journal of Advances in Modeling Earth Systems, 9, 1595-1627, doi:10.1002/2016MS000854, 2017.

Peleg, N., Molnar, P., Burlando, P. and Fatichi, S.: Exploring stochastic climate uncertainty in space and time using a griddedhourly weather generator, Journal of Hydrology, 571, 627-641, doi:10.1016/j.jhydrol.2019.02.010, 2019.

Prein, A.F., Langhans, W., Fosser, G., Ferrone, A., Ban, N., Goergen, K., Keller, M., Tolle, M., Gutjahr, O., Feser, F., Brisson, E., Kollet, S., Schmidli, J., van Lipzig, N., Leung, R.: A review on regional convection-permitting climate modeling: Demonstrations, prospects, and challenges, Reviews of Geophysics, 53, 323-361, doi:2014RG000475, 2015.

Richardson, C. W.: Stochastic simulation of daily precipitation, temperature, and solar radiation. Water Resources Research, 17, 182-190, doi: 10.1029/WR017i001p00182, 1981.

Rust, H.W., Vrac, M., Sultan, B. and Lengaigne, M.: Mapping Weather-Type Influence on Senegal Precipitation Based on a Spatial–Temporal Statistical Model, Journal of Climate, 26, 8189-8209, doi: 10.1175/jcli-d-12-00302.1, 2013.

Schleiss, M., Jaffrain, J. and Berne, A.: Statistical analysis of rainfall intermittency at small spatial and temporal scales, Geophysical Research Letters, 38, doi:10.1029/2011GL049000, 2011.

Smith, J.A., Hui, E., Steiner, M., Krajewski, W.F. and Ntelekos, A.A.: Variability of rainfall rate and raindrop size distributions in heavy rain, Water Resources Research, 45, doi:10.1029/2008WR006840, 2009.

Van Meijgaard, E., Van Ulft, L.H., Van de Berg, W.J., Bosveld, F.C., Van den Hurk, J.M. Lenderink, G. and Siebesma, A.P.: The KNMI regional atmospheric climate model RACMO version 2.1., Report of Koninklijk Nederlands Meteorologisch Instituut, 2008.

Verdin, A., Rajagopalan, B., Kleiber, W. and Katz, R.W.: Coupled stochastic weather generation using spatial and generalized linear models, Stochastic Environmental Research and Risk Assessment, 29, 347-356, doi: 10.1007/s00477-014-0911-6, 2015.

Vischel, T., Quantin, G., Lebel, T., Viarre, J., Gosset, M., Cazenave, F. and Panthou, G.: Generation of High-Resolution Rain Fields in West Africa: Evaluation of Dynamic Interpolation Methods, Journal of Hydrometeorology, 12, 1465-1482, doi: 10.1175/JHM-D-10-05015.1, 2011.

Voldoire, A. et al.: The CNRM-CM5.1 global climate model: description and basic evaluation, Climate Dynamics, 40, 2091-2121, doi: 10.1007/s00382-011-1259-y, 2013.

Volosciuk, C., Maraun, D., Vrac, M. and Widmann, M.: A combined statistical bias correction and stochastic downscaling method for precipitation, Hydrology and Earth System Sciences, 21, 1693-1719, doi: 10.5194/hess-21-1693-2017, 2017.

Vrac, M., Stein, M. and Hayhoe, K.: Statistical downscaling of precipitation through nonhomogeneous stochastic weather typing, Climate Research, 34, 169-184, doi: 10.3354/cr00696, 2007.

Vrac, M., Drobinski, P., Merlo, A., Herrmann, M., Lavaysse, C., Li, L. and Somot, S.: Dynamical and statistical downscaling of the French Mediterranean climate: uncertainty assessment, Natural Hazards and Earth System Sciences, 12, 2769-2784, doi: 10.5194/nhess-12-2769-2012, 2012.

Wilby, R.L.: Stochastic weather type simulation for regional climate change assessment, Water Resources Research, 30, 3395-3403, doi: 10.1029/94wr01840, 1994.

Wilks, D. S.: Use of stochastic weather generators for precipitation downscaling. WIREs Climate Change, 1, 898-907, doi: 10.1002/wcc.85, 2010.

Wilks, D. S. and Wilby, R. L.: The weather generation game: a review of stochastic weather models. Progress in Physical Geography, 23, 329-357, doi: 10.1177/030913339902300302, 1999.

Willems, P.: A spatial rainfall generator for small spatial scales. Journal of Hydrology, 252, 126-144, doi: 10.1016/S0022-1694(01)00446-2, 2001.

Winterrath, T., Rosenow, W. and Weigl, E.: On the DWD quantitative precipitation analysis and nowcasting system for real-time application in German flood risk management, Weather radar and hydrology, 351, 323-329, 2012.

Supplementary material to 'Non-stationary stochastic rain type generation: accounting for climate drivers'

Lionel Benoit[1], Mathieu Vrac[2], and Gregoire Mariethoz[1]

[1]Institute of Earth Surface Dynamics (IDYST), University of Lausanne, Lausanne, Switzerland
[2]Laboratory for Sciences of Climate and Environment (LSCE-IPSL), CNRS/CEA/UVSQ, Orme des Merisiers, France

This technical note complements the paper *Non-stationary stochastic rain type generation: accounting for climate drivers*. It consists of several side experiments that were carried out to support technical details of the main study. More precisely, Supplementary material 1 reiterates the cross-validation of Sect 4.1 in case radar images with a low rain coverage are assigned to rain types, and not classified as dry (cf Sect 2.1). Supplementary material 2 assesses the impact of using high-resolution and in-situ data as meteorological covariates, instead of the reanalysis data introduced in Sect 2.2. Supplementary material 3 explores the use of a parametric model as an alternative to the non-parametric model detailed in Sect 3. Finally, Supplementary material 4 details the bias correction of RCM data prior to their use for precipitation downscaling in Sect 5.

Supplementary material 1: Cross-validation when accounting for low rain coverage

In this supplementary material, we reiterate the cross-validation of Sect 4.1 for rain type time series in which radar images with a low rain coverage (rain fraction between 5% and 10%) are regarded as wet. To build such time series, radar images with more than 10% rainy pixels are first classified into rain types according to their space-time-intensity statistical signature as detailed in Sect 2.1. However, in contrast with the main study, the images with rain fraction between 5% and 10% are classified into rain types in a second step by assigning them the type of the closest classified image (i.e. nearest neighbor interpolation in time, cf. (Benoit et al., 2018b)).

The exact same leave-one-year-out cross-validation procedure than in Sect 4.1 is applied to the above rain type time series: for a given simulation year, the rain type model is first trained using data from the 2001-2017 period, excluding the year to simulate. Next, 50 realizations of rain type time series are generated for the year of interest by MPS simulation, conditioned to meteorological covariates. This procedure is iterated for each year of the test period (i.e. 2001-2017), and 50 realizations of 17-years long rain type time series are obtained by concatenating in time the 17 yearly simulations.

Results in Fig. 1 show that classifying radar images with low coverage in a rain type instead of considering them dry does not influence the performance of rain type simulation. Indeed, the only noticeable difference with the Fig 6 of the main manuscript is the correction of the dry bias in the present case. Hence, the fact that the performance of rain type simulation is insensitive to the way images with low rain coverage are classified opens the doors to re-adjusting the wet/dry balance by post-processing, as proposed in Sect 2.1.

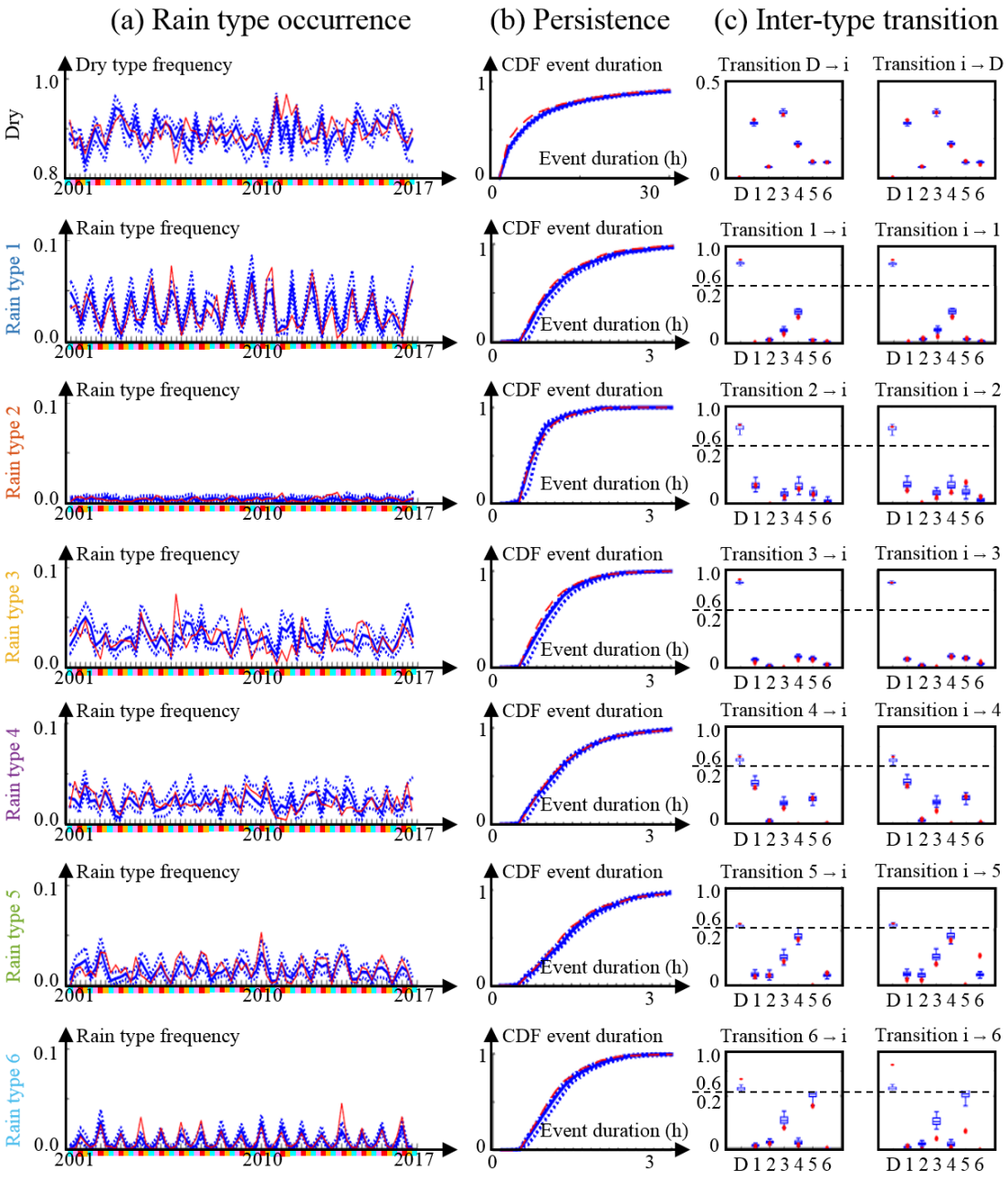

Figure 1. Results of the cross-validation experiment when images with low rain coverage (rain fraction between 5% and 10%) are regarded as wet. (a) Seasonality of rain (and dry) type occurrence (Seasons are DJF (light blue), MAM (pink), JJA (red) and SON (yellow)), (b) rain type persistence, and (c) probability of transition between rain types. Observations are in red and are obtained by assigning the type of the closest classified image to epochs with rain fraction between 5% and 10%. Simulations are in blue and are obtained by propagating the closest rain type to the beginning and to the end of each rain event. In simulations, continuous lines represent the median of the simulated ensembles (50 realizations), and dashed lines represent the Q10 and Q90 quantiles.

Supplementary material 2: Selection of meteorological covariates

The stochastic rain type generator developed in this paper requires meteorological covariates in order to (1) ensure the climatological consistency of the simulations, and (2) reproduce the annual cycle as well as the inter-annual variability of rain type occurrence. As mentioned in Sect. 2.2 we focus on meteorological parameters that are known to influence the triggering and the behavior of rain storms (Vrac et al., 2007; Willems, 2001; Rust et al., 2013), namely: pressure, temperature, relative humidity, and wind direction and intensity. Here we choose to use the actual values of the covariates rather than weather types defined by classification of the spatial patterns of one or several of these covariates (Vrac et al., 2007; Rust et al., 2013; Milrad et al., 2014). There are two main reasons for this: First, in case of weather types derived from several covariates, each weather type combines in an intractable manner the influence of the different meteorological parameters, thus making the identification of the climatological drivers of rain type occurrence difficult. Second, the use of weather types drastically reduces the dimensionality of the covariate space, which allows fewer nuances in the links between meteorological conditions and rain types.

Rain type data and simulations are used at 10-min resolution, and therefore the meteorological covariates must be provided at the same resolution. In addition, rainfall is often related to short time changes in meteorological conditions (in particular temperature and air pressure), and high resolution covariates are therefore expected to better explain rain types. However, one of the two foreseen applications of the proposed stochastic rain type generator is the stochastic downscaling of regional climate model projections, which are most of the time available only at daily resolution. Hence the hybrid solution adopted in this study, which uses daily resolution meteorological covariates and disaggregates them to a 10-min resolution as described in Sect. 2.2.

Figure 2 compares disaggregated data derived from the ERA5 reanalysis with in-situ observations carried out by a weather station (located in Erfurt, within the study area) at high resolution (10-min for pressure, temperature and relative humidity, and 6h for wind). Results show that pressure and temperature have similar behaviors in the two datasets, while relative humidity and wind present significant dissimilarities. Regarding relative humidity, the observed dissimilarities are mostly caused by the influence of the daily temperature cycle in the high resolution dataset. When replacing relative humidity by vapor pressure, which is not correlated with temperature, it appears that the differences between the two datasets are considerably reduced. Hence, the independent information (i.e. not correlated with other covariates) carried by the relative humidity does not drastically differ between the two datasets. Regarding wind-related covariates (here the Eastward and Northward components of the wind vector are assessed), it is interesting to note that the dataset derived from the reanalysis carries more signal than the in-situ observations. This is because in-situ observations refer to low altitude winds, which are much more variable than the wind at 850hPa extracted from the reanalysis. Hence, the signal-to-noise ratio of the weather station dataset is lower than the one derived from ERA5. All in all, the proposed disaggregation method performs well to capture sub-daily variations of meteorological conditions, despite some mismatches during stormy periods.

To ensure that the small fluctuations of the meteorological covariates that are missed by the disaggregated dataset do not negatively affect the simulation of rain types, we performed an additional cross-validation experiment where the rain type simulation is forced by the two datasets of covariates described above, namely: (1) ERA5 data disaggregated at 10 min resolution

(i.e. the dataset used throughout the paper) and (2) in-situ observations from a weather station located within the study area. Fig. 3 summarizes the results of this experiment and shows very little differences between the two simulations. The almost similar performance when using the two sets of covariates can on the one hand be explained by the fact that the additional signal embedded into in-situ observations hardly emerges from the measurement noise, and on the other hand by the fact that the additional information brought by the wind data in the reanalysis dataset compensates for the uncertainties related to the 10-minute reconstruction of pressure, temperature and relative humidity data. Finally, since both sets of covariates lead to similar performances, we favor the disaggregation of daily data because it is compatible with the targeted application of regional climate model downscaling.

After disaggregation, the meteorological covariates can be used to investigate how weather conditions influence rain type occurrence. To this end, Fig. 4 displays the impact of pairs of covariates on rain type occurrence. It shows that although temperature and wind intensity are the main drivers for rain type occurrence, the joint knowledge of all covariates is important to assess rain type distribution.

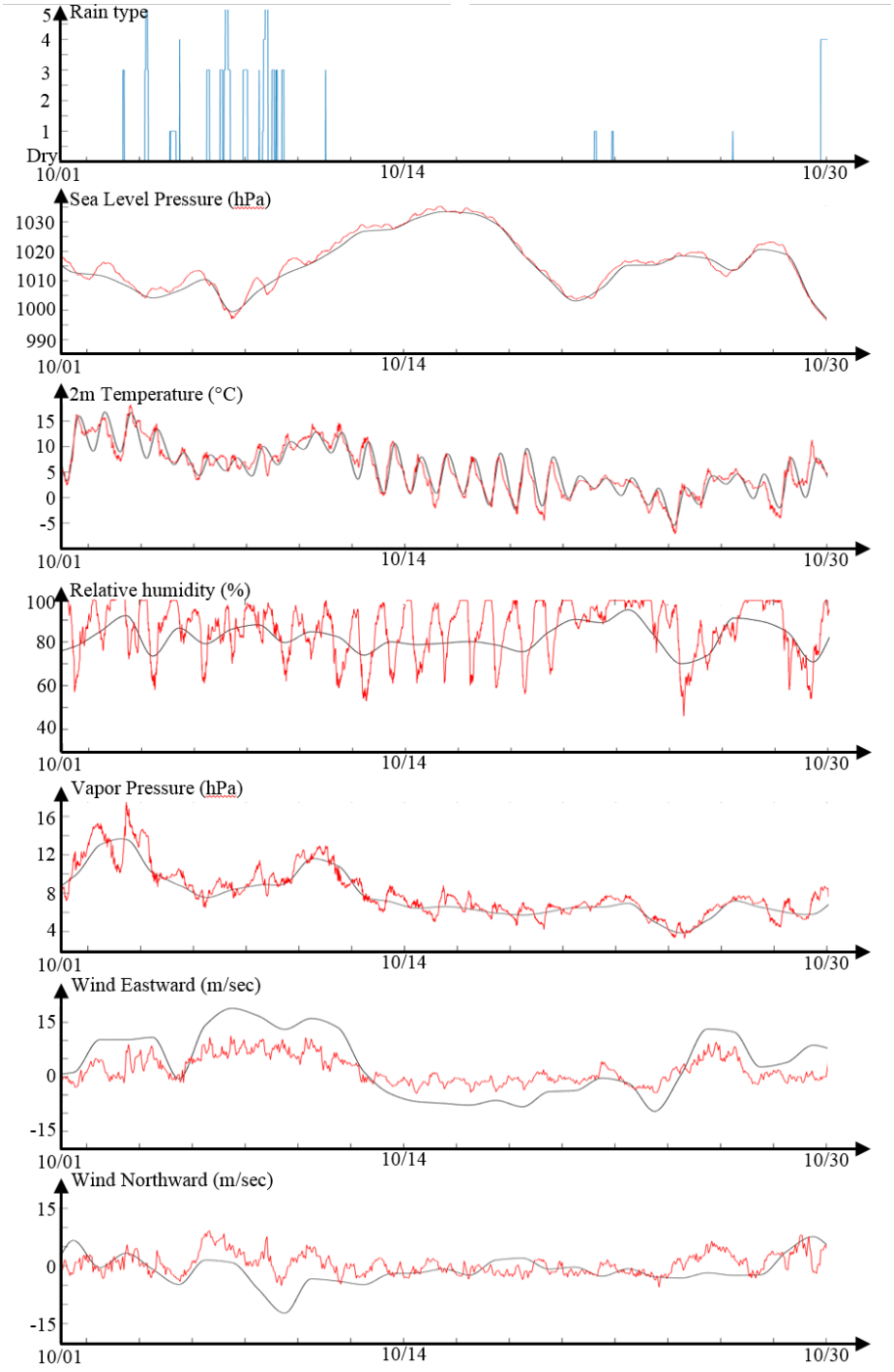

Figure 2. Comparison between in-situ observations (red) and disaggregated ERA5 reanalysis data (black) of the meteorological covariates considered in this study. Period of interest: October 2003. From top to bottom: observed rain types (derived from radar images, cf Sect. 2.1, pressure, temperature, relative humidity, vapor pressure, Eastward wind component, Northward wind component.

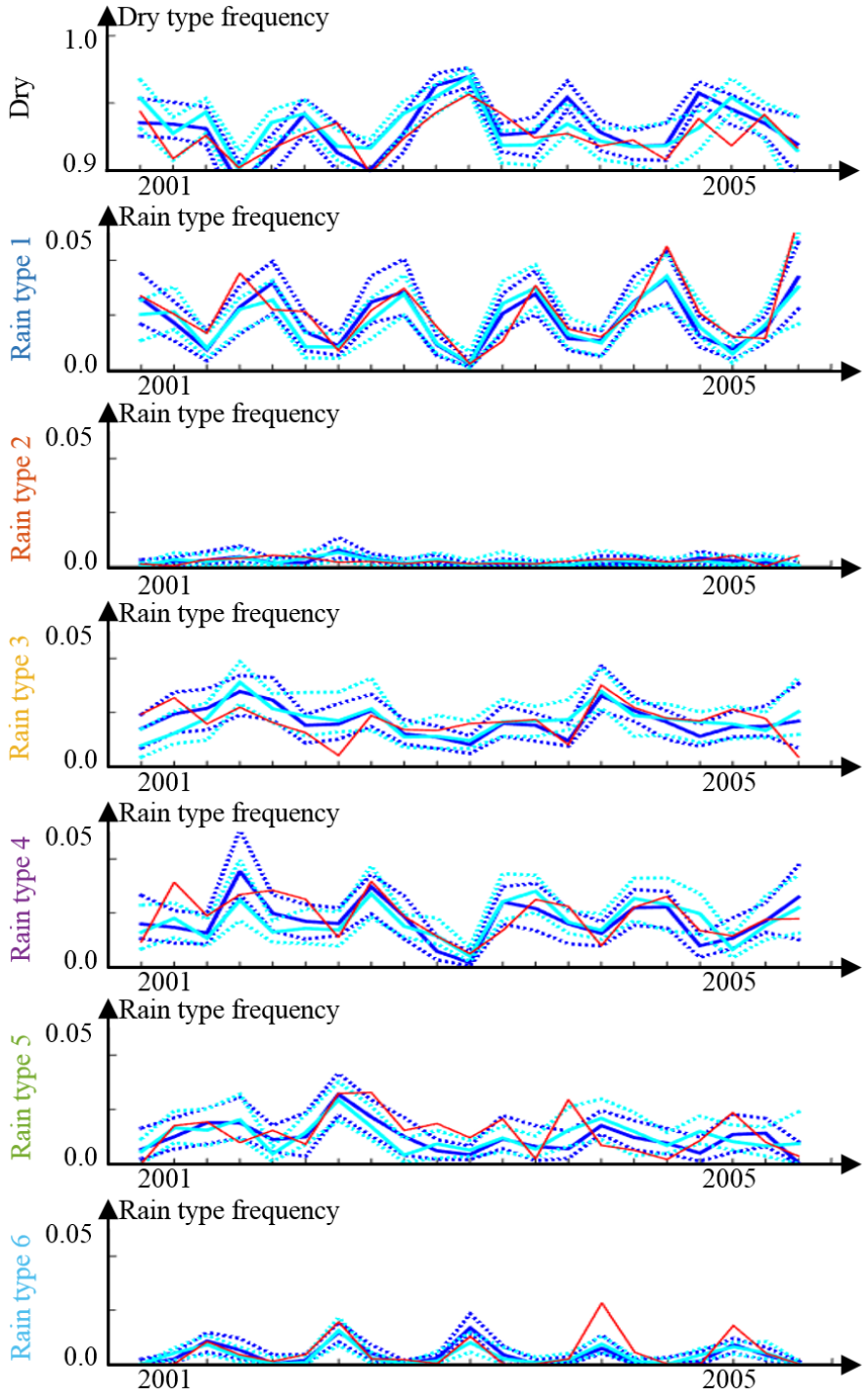

Figure 3. Cross-validation using the leave-one-year out method described in Sect. 4.1 applied to the 2001-2005 period and two different sets of covariates. Observations are in red, simulations using the disaggregated ERA5 covariates are in dark blue, and simulations using in-situ observations of covariates are in light blue. Continuous lines represent the median of the simulated ensembles (30 realizations), and dashed lines represent the Q10 and Q90 quantiles.

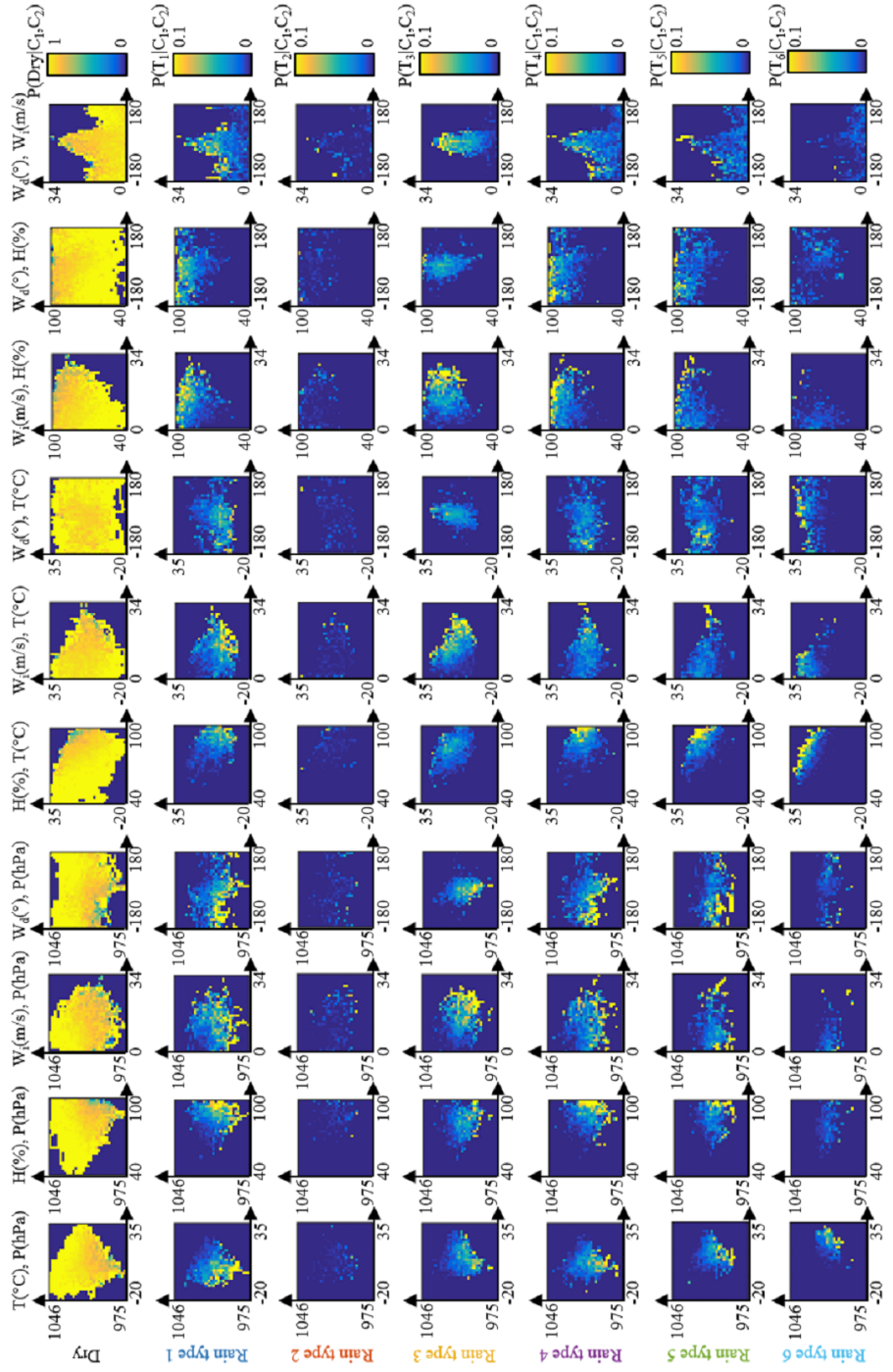

Figure 4. Joint probability of rain type occurrence conditional to pairs of meteorological covariates. Rows correspond to different rain types, and columns correspond to different pairs of meteorological covariates. The title of the rows indicates the pair of covariates of interest; the first covariate defines the abscissa axis, and the second the ordinate axis. T denotes temperature, P is pressure, H is relative humidity, W_i is wind intensity and W_d is wind direction. In each graph, the probability of rain type occurrence is coded by the color-scale.

Supplementary material 3: Markov-chain-based model of rain type occurrence

As mentioned in Sect. 3, our first attempt of building a stochastic rain type model was based on a parametric approach using a Markov-chain model. Despite not fully satisfactory results, this model is presented hereafter as a possible benchmark for the non-parametric framework proposed in the main paper (Sect. 3).

The parametric approach that has been tested to model rain type occurrence builds on Markov-chain models, which were originally developed to model dry/wet sequences in daily resolution applications (Richardson, 1981; Wilby, 1994; Wilks and Wilby, 1999). The existing models are substantially amended to match with the features of sub-daily resolution rain type time series, which significantly differ from dry/wet sequences at daily resolution. To model sub-daily resolution rain type time series, we adopt a non-homogeneous semi-Markov model (i.e. with non-stationary transition probabilities and time-varying step lengths) with N+2 states (Fig. 5). Among these N+2 states, N states model rain types (in the present case N=6), and 2 states model dry periods that are split into 'short dry' (duration <24h) and 'long dry' (duration >24h) states. In practice, if a dry period exceeds 24h, it is split into as many 'long dry' spells as possible, and the remaining time is distributed into two 'short dry' spells of equal length at the beginning and at the end of the overall dry period. The 'long dry' state can only transition to 'short dry', and the rain types (as well as the 'short dry' type) can transition to each other and with the 'short dry' type. A semi-Markov approach (Foufoula-Georgiou and Lettenmaier, 1987; Bárdossy and Plate, 1991) is used to account for the persistence of rain (and 'short dry') types. In our model, the duration of these types is explicitly defined by a Probability Density Function (PDF) of event duration, and the Markov chain is not allowed to be twice in the same state (i.e. transitions from state i to i are censored). On the contrary, the 'long dry' state always lasts exactly 24h, but is allowed to transition to itself to generate long lasting dry spells. Finally, to account for the non-stationarity of rain type occurrence in time, the semi-Markov chain is made non-homogeneous (Hughes and Guttorp, 1999; Vrac et al., 2007). It consists of changing the probability transition matrix of the Markov chain over time conditionally to a set of meteorological covariates X:

$$P\left(S_t = j | S_{t-1} = i, X_t\right) \propto \gamma_{ij}.\exp\left(-\frac{1}{2}(X_t - \mu_{ij})\Sigma^{-1}(X_t - \mu_{ij})^T\right) \quad (1)$$

Where S_t is the state of the Markov chain at time t, Σ is the covariance matrix of the covariates, μ_{ij} is the mean vector of the covariates when the transition from type i to type j occurs, and γ_{ij} is the baseline (i.e. long term averaged) transition probability from state i to state j. It should be emphasized here that since the 'long dry' state is allowed to transition to itself, the probability of transition from 'long dry' to 'long dry' is driven by the meteorological covariates, and indirectly the length of dry spell duration is made dependent on the state of the atmosphere. Conversely, the persistence of all other states (i.e. rain types and 'short dry' type) is stationary in time, and only the probabilities of occurrence of these states depend on meteorological conditions. The non-homogeneous semi-Markov model of rain type occurrence is summarized in Fig. 5.

The parameters of the model can be inferred from historical records using the following procedure. First, the baseline transition matrix is estimated by counting all transitions between each pair of rain types (incl. dry states) occurring in the calibration dataset, and by normalizing the result by the total number of transitions. Then, the parameters required to make the transition matrix non-homogeneous (i.e. μ_{ij} and Σ, cf Eq. (1)) are estimated using a synchronous record of covariate observations. More

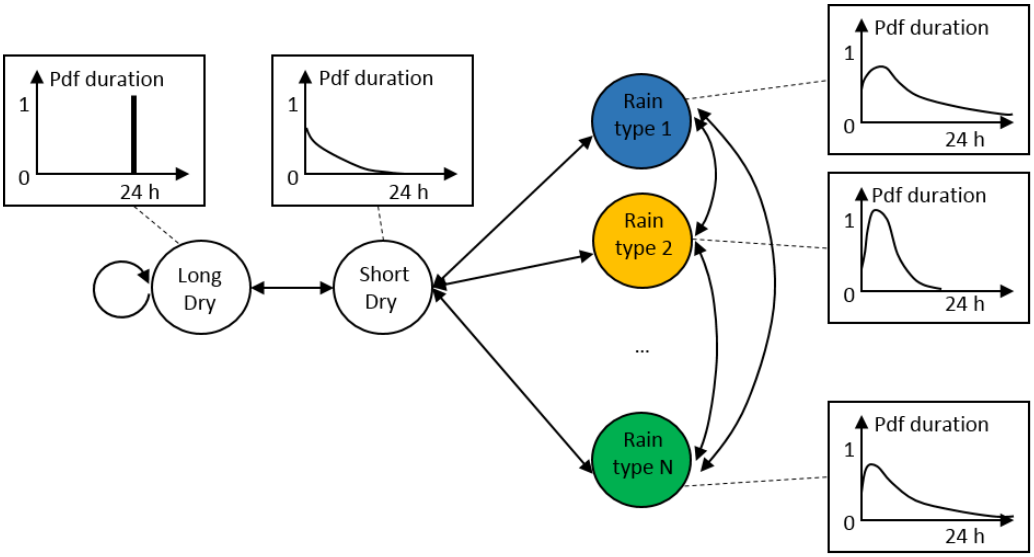

Figure 5. Schematic view of the non-homogeneous semi-Markov model used to model sub-daily rain type occurrence.

precisely, Σ is the empirical covariance matrix and μ_{ij} is the empirical mean of the covariates for the time steps where the transition from i to j occurs. Finally, the PDFs of rain type duration are assumed to be gamma distributions whose parameters are inferred by likelihood maximization using the observed rain type durations. After inference of model parameters, the calibrated model can be used to generate synthetic rain type time series conditioned to meteorological covariates. To this end, time series of covariates have to be available for the target simulation period. Next, synthetic rain types are stochastically generated by iteratively (i) simulating a rain type (or dry / wet) transition, and (ii) generating the duration of the current event by sampling the PDF of duration of the rain (or dry) type of interest.

Figure 6 shows the results of the same cross-validation experiment than in Sect. 4.1 of the main study, but applied to the parametric model described above. Results show that this model generates a strong dry bias (ratio simulated/observed rain frequency = 0.61) and does not properly capture the inter-annual variability of rain occurrence (correlation between observed and simulated time series = 0.4). This can be explained by the fact that the relationships between the meteorological covariates and the presence of rain are probably more complex than the linear relationship assumed in the non-homogeneous Markov chain formulation of the parametric model, and by the fact that semi-Markov models do not allow for an easy conditioning to continuous-time covariates. These hypotheses are reinforced by the fact that simulations driven by daily mean temperature only do not generate a dry bias (not shown here). To conclude, it should be noted here that the lower performance of the parametric model is due to the complexity of the problem at hand, rather than a deficiency in the model itself, which extends to high temporal resolution problems some approaches that are state-of-the-art in daily resolution stochastic weather generators.

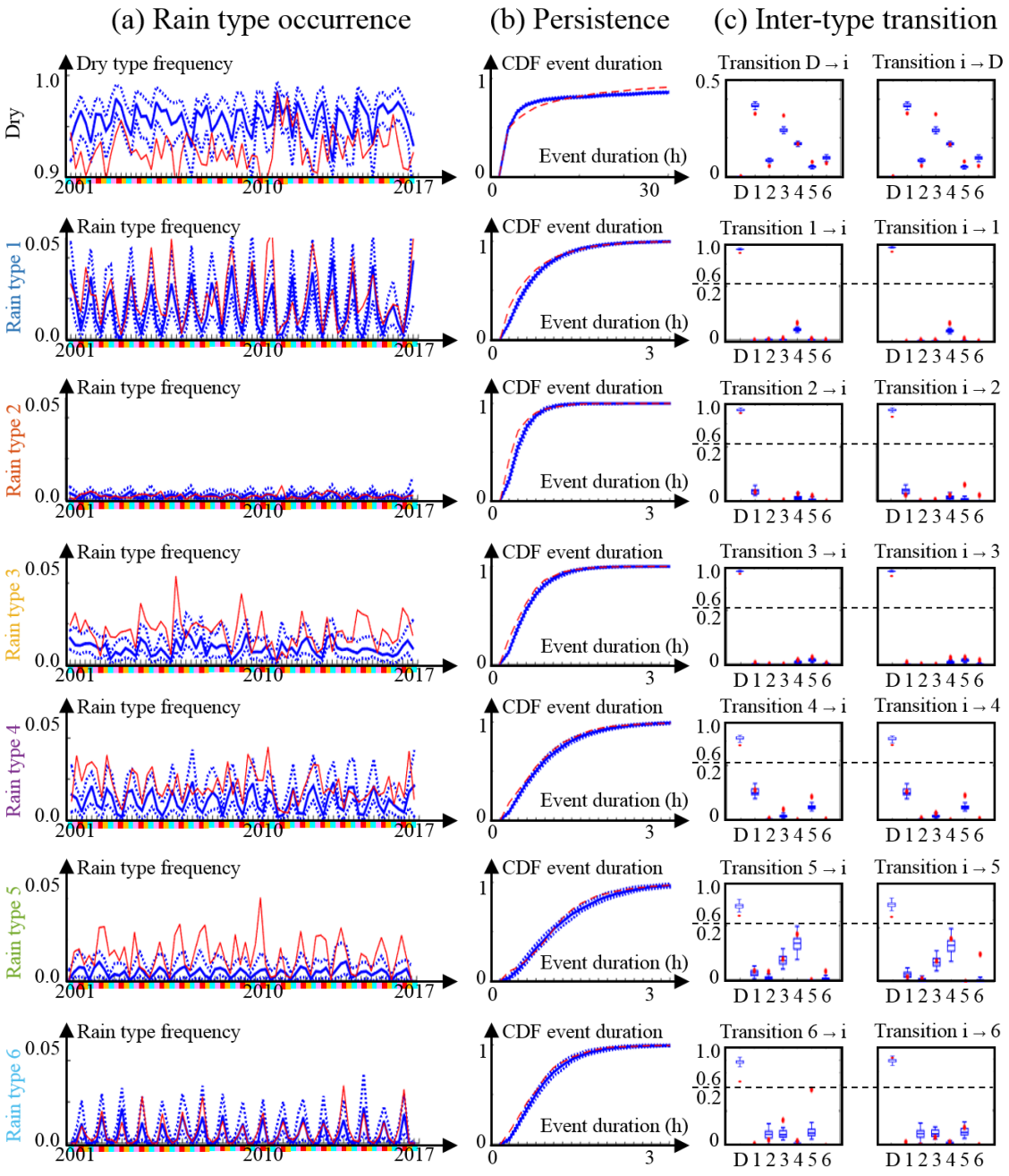

Figure 6. Results of the cross-validation experiment for the parametric model. (a) Seasonality of rain (and dry) type occurrence (Seasons are DJF (light blue), MAM (pink), JJA (red) and SON (yellow)), (b) rain type persistence, and (c) probability of transition between rain types. Observations are in red and simulations in blue. In simulations, continuous lines represent the median of the simulated ensembles (50 realizations), and dashed lines represent the Q10 and Q90 quantiles.

Supplementary material 4: RCM bias correction and performance of associated rain type simulations

Meteorological covariates are used to encode rain type non-stationarity (Sect 3 of the main study). When used in the context of a changing climate, these covariates are derived from regional climate model (RCM) outputs. In the present case, the meteorological covariates are derived from the Regional Atmospheric Climate MOdel of the Dutch national weather service (RACMO-KNMI (Van Meijgaard et al., 2008)) driven by the CNRM-CM5 Earth system model (Voldoire et al., 2013) forced according to the RCP8.5 emission scenario. However, like almost all RCM projections, the specific RCM that is used to derive meteorological covariates can be biased. The meteorological covariates must therefore be bias-corrected before further use for stochastic rain type simulation (Sect 5 of the main study). In the present study, we apply the CDF-t method for bias-correction of each variable separately (Vrac et al., 2012).

Figure 7 shows the impact of bias-correction on the distribution of each meteorological covariate. One can notice the significant impact of bias-correction on daily minimum temperature and on humidity. Figure 8 shows that after bias-correction, the performance of the model to simulate rain types in the present climate is almost identical for meteorological covariates derived from the RACMO-KNMI RCM and the ones derived from the ERA-5 reanalyses. This result paves the way to precipitation downscaling through stochastic rain type simulation, as detailed in Sect 5.

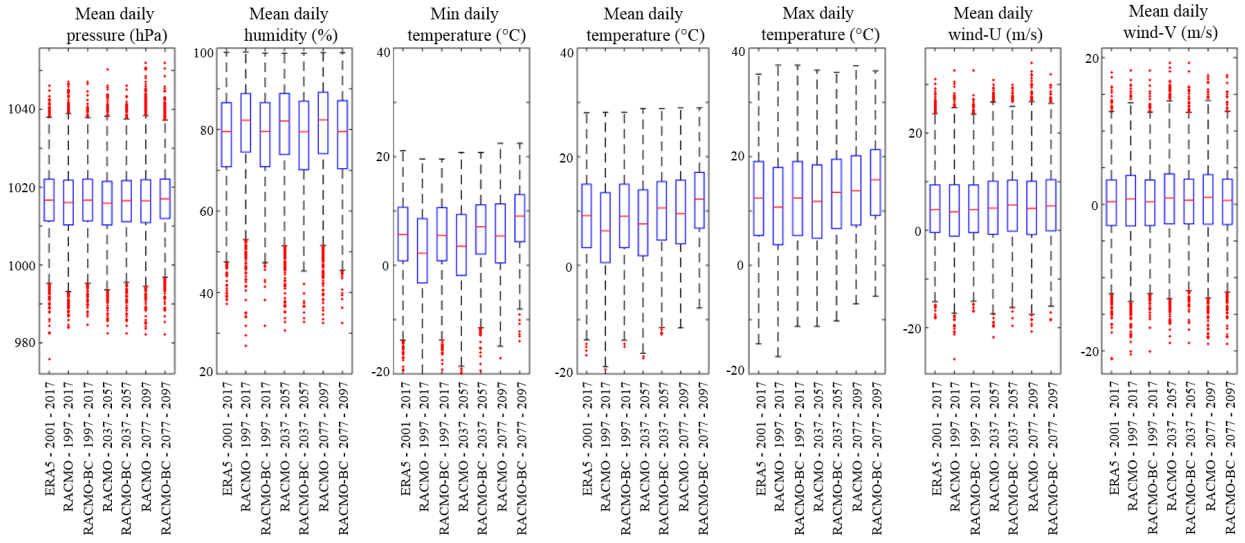

Figure 7. Evolution of the meteorological covariates along the different periods tested in this study: 2001-2017 (present climate, cross-validation), 1997-2017 (present climate, application to RCM downscaling), 2037-2057 and 2077-2097 (future climate, application to RCM downscaling). ERA5 refers to meteorological data derived from the ERA-5 reanalysis that are used for calibration and cross-validation. RACMO refers to meteorological data derived directly from the RACMO – CNRM regional climate model. RACMO-BC is similar to RACMO but has been bias-corrected using the CDF-t method and ERA5 data as reference for the present climate. RACMO-BC is the dataset that is used to condition the rain type simulations in subsection 4.2. In the boxplots the central red line denotes the median, the blue box encompasses the 25%-75% quantiles, the whiskers encompass all non-outliers data points (+/- 2.7σ), and the red crosses denote outliers.

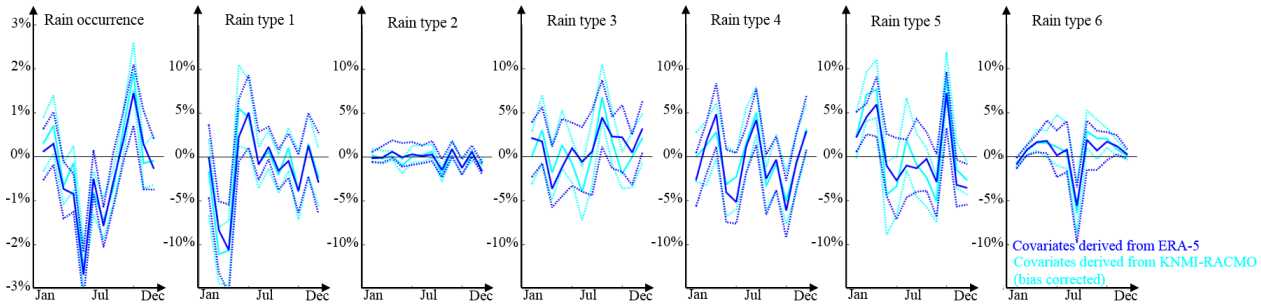

Figure 8. Performance of the proposed stochastic rain type model to simulate rain type occurrence for the 2001-2017 period using two different sets of meteorological covariates: ERA-5 reanalysis (dark blue) and bias-corrected RACMO-KNMI RCM (light blue). This figure shows the difference between observed and simulated monthly rain occurrence (left panel) and rain type frequency conditional to the presence of rain (other panels). Continuous lines represent the median of the simulated ensembles (50 realizations), and dashed lines represent the Q10 and Q90 quantiles.

References

Bárdossy, A. and Plate, E.J.: Modelling daily rainfall using a semi-Markov representation of circulation pattern occurence, Journal of Hydrology, 122, 33-47, doi: 10.1016/0022-1694(91)90170-M, 1991.

Benoit, L., Vrac, M. and Mariethoz, G.: Dealing with non-stationarity in sub-daily stochastic rainfall models, Hydrology and Earth System Sciences, 22, 5919-5933, doi:10.5194/hess-22-5919-2018, 2018b.

Foufoula-Georgiou, E., and Lettenmaier, D.: A Markov Renewal Model for Rainfall Occurence, Water Resources Research, 23, 875-884, doi: 10.1029/WR023i005p00875, 1987.

Hughes, J.P. and Guttorp, P.: A non-homogeneous hidden Markov model for precipitation occurence, Applied Statistics, 48, 15-30, doi: 10.1111/1467-9876.00136, 1999.

Milrad, S.M., Atallah, E.H., Gyakum, J.R. and Dookhie, G.: Synoptic Typing and Precursors of Heavy Warm-Season Precipitation Events at Montreal, Quebec. Weather and forecasting, 29, 419-444, doi: 10.1175/WAF-D-13-00030.1, 2014.

Richardson, C. W.: Stochastic simulation of daily precipitation, temperature, and solar radiation. Water Resources Research, 17, 182-190, doi: 10.1029/WR017i001p00182, 1981.

Rust, H.W., Vrac, M., Sultan, B. and Lengaigne, M.: Mapping Weather-Type Influence on Senegal Precipitation Based on a Spatial–Temporal Statistical Model, Journal of Climate, 26, 8189-8209, doi: 10.1175/jcli-d-12-00302.1, 2013.

Van Meijgaard, E., Van Ulft, L.H., Van de Berg, W.J., Bosveld, F.C., Van den Hurk, J.M. Lenderink, G. and Siebesma, A.P.: The KNMI regional atmospheric climate model RACMO version 2.1., Report of Koninklijk Nederlands Meteorologisch Instituut, 2008.

Vrac, M., Stein, M. and Hayhoe, K.: Statistical downscaling of precipitation through nonhomogeneous stochastic weather typing, Climate Research, 34, 169-184, doi: 10.3354/cr00696, 2007.

Vrac, M., Drobinski, P., Merlo, A., Herrmann, M., Lavaysse, C., Li, L. and Somot, S.: Dynamical and statistical downscaling of the French Mediterranean climate: uncertainty assessment, Natural Hazards and Earth System Sciences, 12, 2769-2784, doi: 10.5194/nhess-12-2769-2012, 2012.

Voldoire, A. et al.: The CNRM-CM5.1 global climate model: description and basic evaluation, Climate Dynamics, 40, 2091-2121, doi: 10.1007/s00382-011-1259-y, 2013.

Wilby, R.L.: Stochastic weather type simulation for regional climate change assessment, Water Resources Research, 30, 3395-3403, doi: 10.1029/94wr01840, 1994.

Wilks, D. S. and Wilby, R. L.: The weather generation game: a review of stochastic weather models. Progress in Physical Geography, 23, 329-357, doi: 10.1177/030913339902300302, 1999.

Willems, P.: A spatial rainfall generator for small spatial scales. Journal of Hydrology, 252, 126-144, doi: 10.1016/S0022-1694(01)00446-2, 2001.