Supplement of Hydrol. Earth Syst. Sci., 24, 2841–2854, 2020
https://doi.org/10.5194/hess-24-2841-2020-supplement

Hydrology and Earth System Sciences
Open Access
EGU

Supplement of

Nonstationary stochastic rain type generation: accounting for climate drivers

Lionel Benoit et al.

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

Supplementary material

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.

Sect. S1: 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.

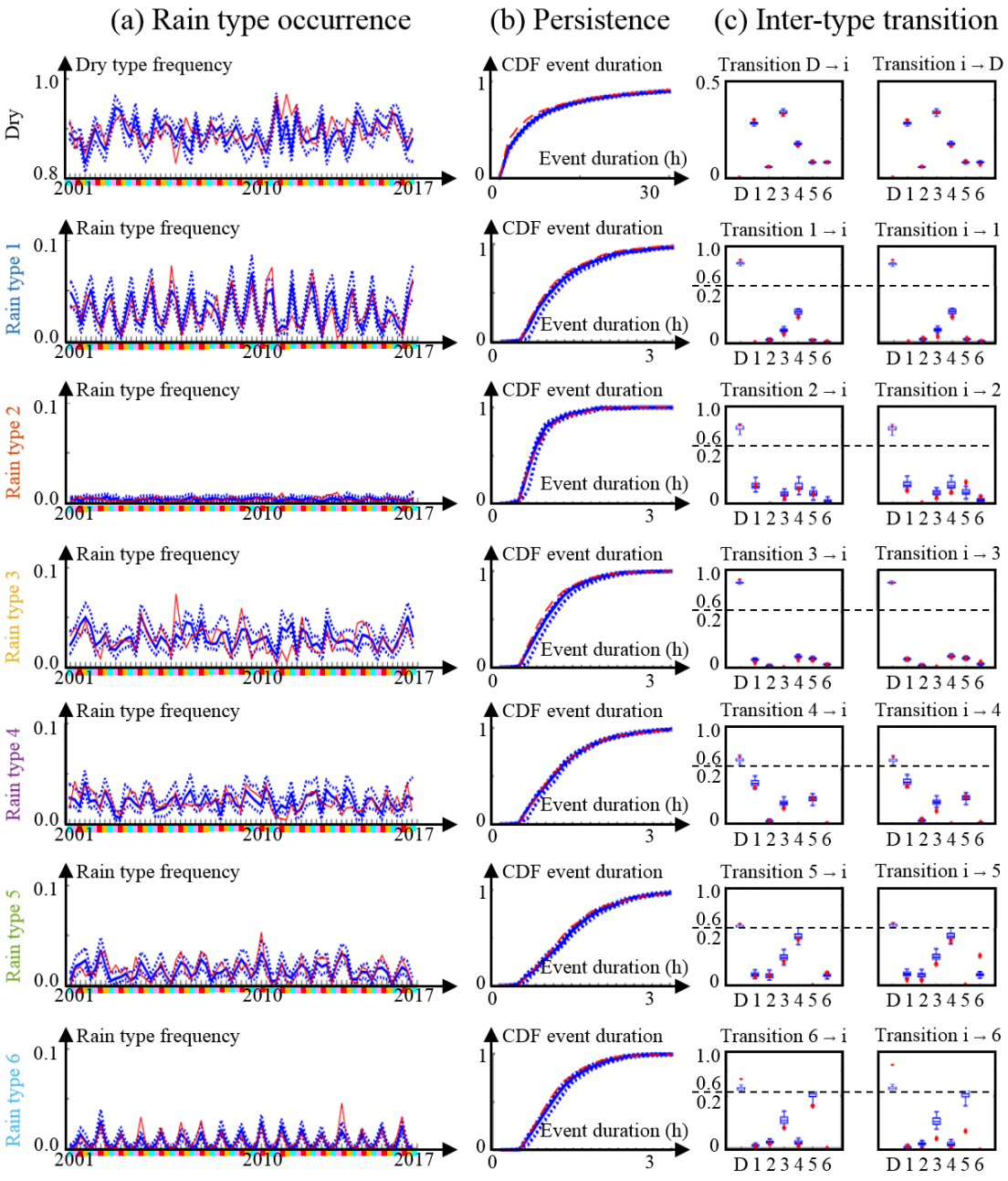

Figure S1. Results of the cross-validation experiment when images with low rain coverage (rain fraction between 5% and 10%) are regarded as wet. (a) Seasonality of rain (and dry) type occurrence (Seasons are DJF (light blue), MAM (pink), JJA (red) and SON (yellow)), (b) rain type persistence, and (c) 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.

Sect. S2: 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.

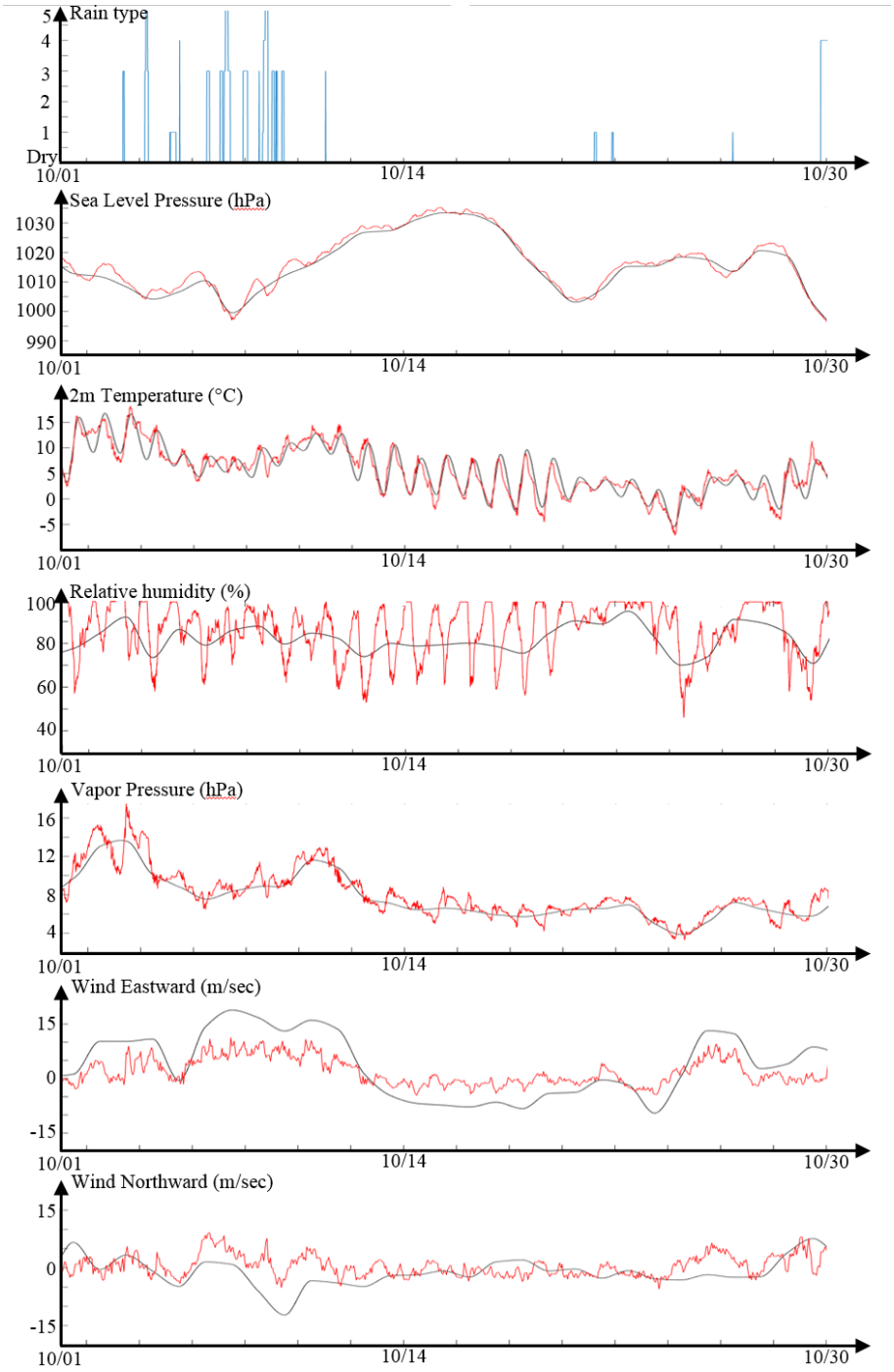

Figure S2. Comparison between in-situ observations (red) and disaggregated ERA5 reanalysis data (black) of the meteorological covariates considered in this study. Period of interest: October 2003. From top to bottom: observed rain types (derived from radar images, cf Sect. 2.1, pressure, temperature, relative humidity, vapor pressure, Eastward wind component, Northward wind component.

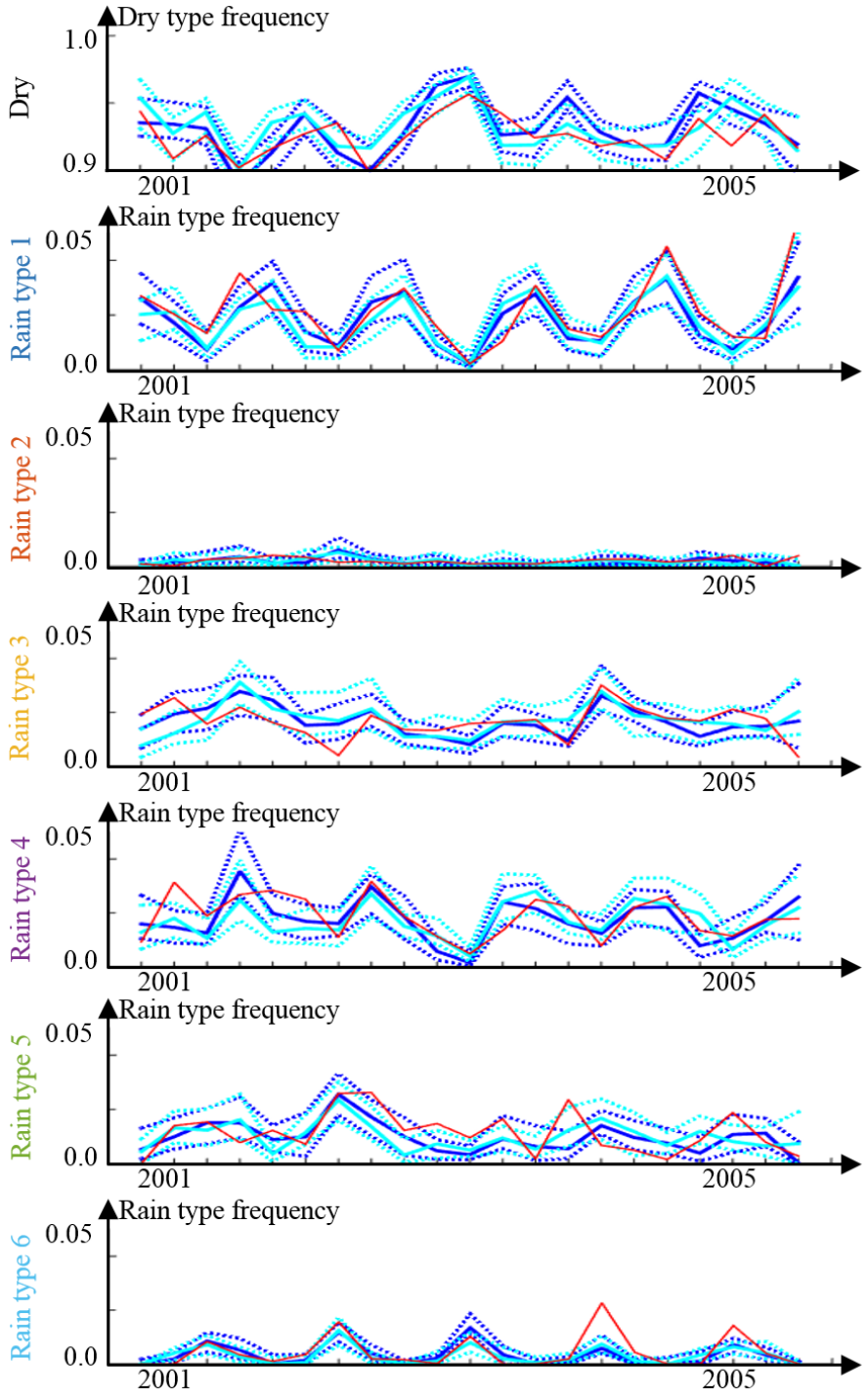

Figure S3. Cross-validation using the leave-one-year out method described in Sect. 4.1 applied to the 2001-2005 period and two different sets of covariates. Observations are in red, simulations using the disaggregated ERA5 covariates are in dark blue, and simulations using in-situ observations of covariates are in light blue. Continuous lines represent the median of the simulated ensembles (30 realizations), and dashed lines represent the Q10 and Q90 quantiles.

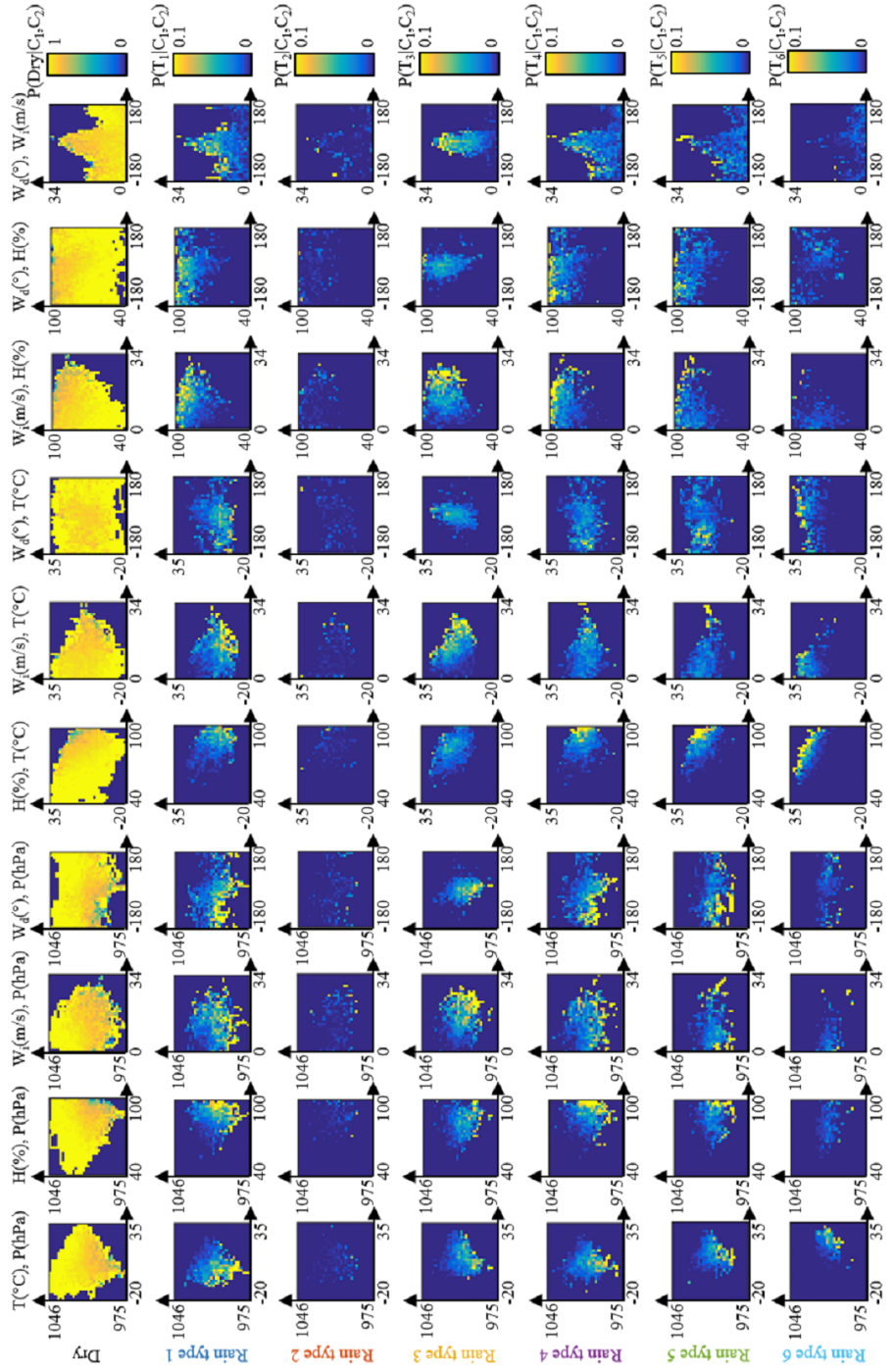

Figure S4. Joint probability of rain type occurrence conditional to pairs of meteorological covariates. Rows correspond to different rain types, and columns correspond to different pairs of meteorological covariates. The title of the rows indicates the pair of covariates of interest; the first covariate 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.

Sect. S3: 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

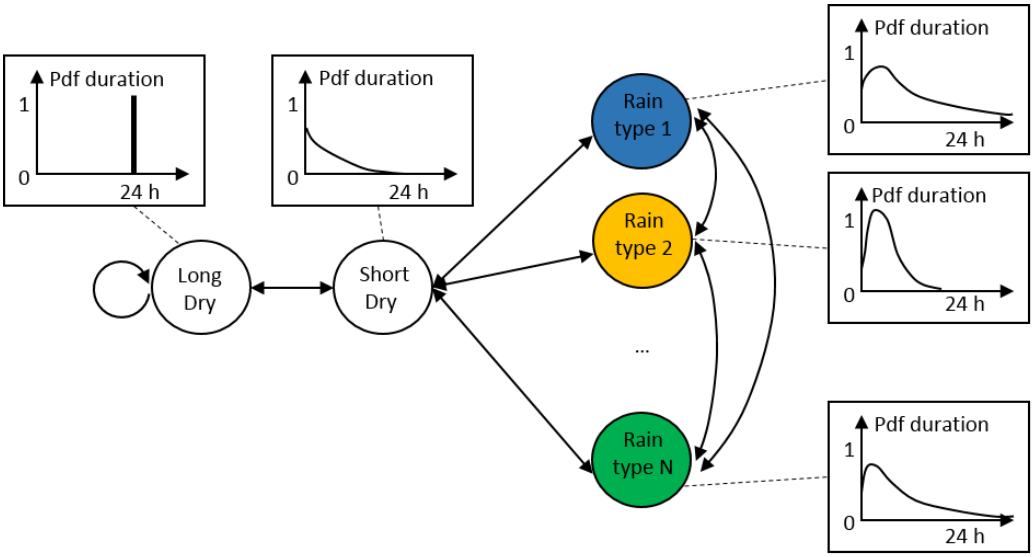

Figure S5. Schematic view of the non-homogeneous semi-Markov model used to model sub-daily rain type occurrence.

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.

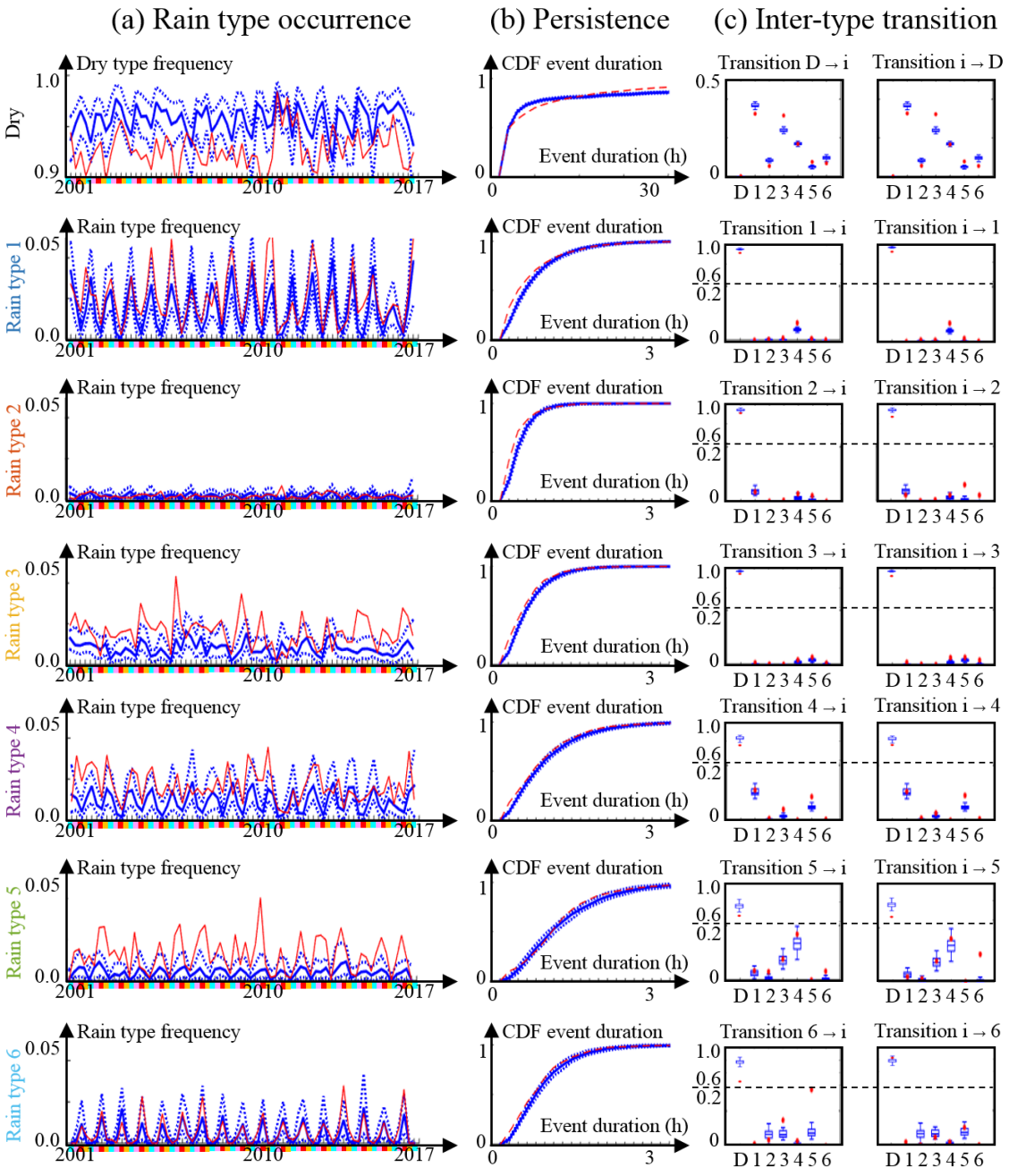

Figure S6. Results of the cross-validation experiment for the parametric model. (a) Seasonality of rain (and dry) type occurrence (Seasons are DJF (light blue), MAM (pink), JJA (red) and SON (yellow)), (b) rain type persistence, and (c) probability of transition between rain types. Observations are in 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.

Sect. S4: 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.

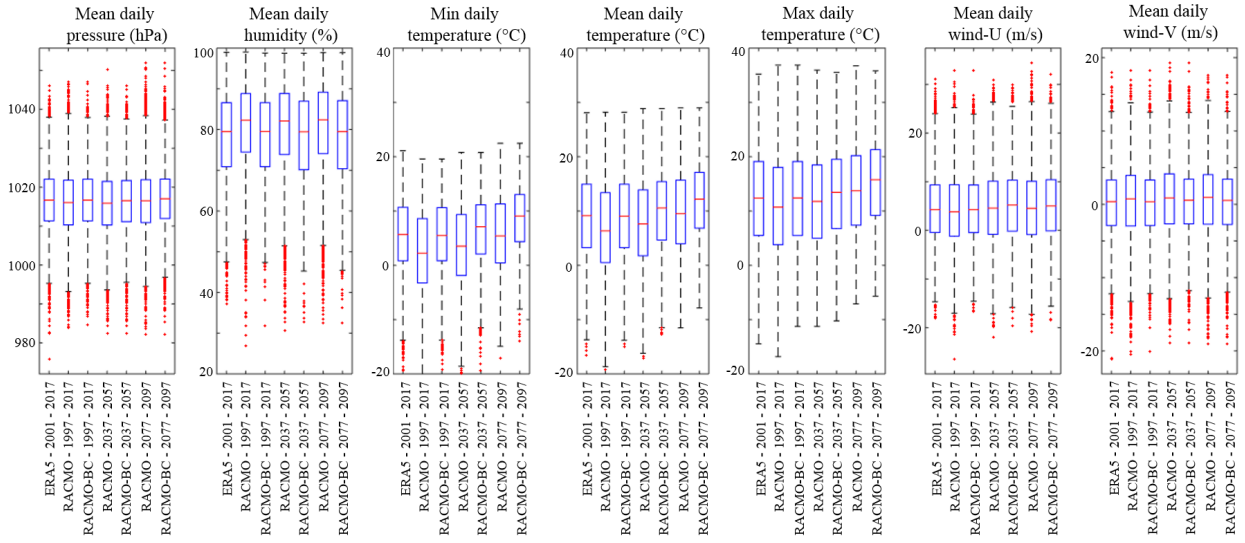

Figure S7. Evolution of the meteorological covariates along the different periods tested in this study: 2001-2017 (present climate, cross- validation), 1997-2017 (present climate, application to RCM downscaling), 2037-2057 and 2077-2097 (future climate, application to RCM downscaling). ERA5 refers to 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.

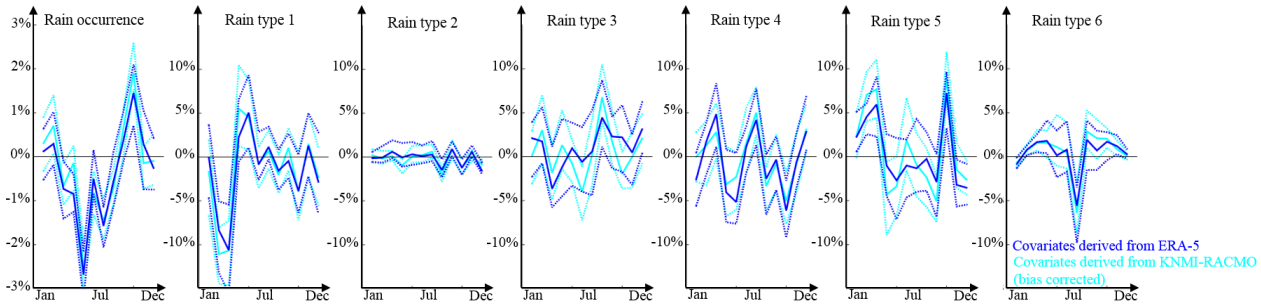

Figure S8. Performance of the proposed stochastic rain type model to simulate rain type occurrence for the 2001-2017 period using two different sets of meteorological covariates: ERA-5 reanalysis (dark blue) and bias-corrected RACMO-KNMI RCM (light blue). This figure shows the difference between observed 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.