

Supplement of Hydrol. Earth Syst. Sci., 19, 2163–2177, 2015
<http://www.hydrol-earth-syst-sci.net/19/2163/2015/>
doi:10.5194/hess-19-2163-2015-supplement
© Author(s) 2015. CC Attribution 3.0 License.



Hydrology and
Earth System
Sciences

Open Access



Supplement of

Implementation and validation of a Wilks-type multi-site daily precipitation generator over a typical Alpine river catchment

D. E. Keller et al.

Correspondence to: D. E. Keller (denise.keller@meteoswiss.ch)

1 Supplementary

2 1 Stochastic simulation of daily precipitation

3 1.1 Single-site

4 The simulation of precipitation at a given day and a given station (say A) is accomplished in
5 four main steps (see yellow circles in Supplementary Figure 1):

- 6 1. A random number $u_{b,A}$ is drawn from a standard normal distribution.
- 7 2. The conditional wet day probability p_A is determined depending on the state of the
8 previous day. It is set to $p_{11,A}$ or $p_{01,A}$, depending on whether the previous simulated day
9 was wet or dry, respectively.
- 10 3. The random number $u_{t,A}$ is compared to the standard normal quantile function Q ,
11 evaluated at p_A : if $u_{t,A}$ is larger than $Q[p_A]$, a dry day ($J_{t,A}=0$) is simulated and else a wet
12 day ($J_{t,A}=1$) is set.

13 4.1 In case of a dry day, the simulated amount $X_{t,A}$ is set to zero.

14 4.2.1 In case of a wet day, a second random number $v_{t,A}$ (independent from $u_{b,A}$) is drawn
15 from a standard normal distribution.

16 4.2.2 The corresponding quantile of the random number $v_{t,A}$ is then inserted into the quantile
17 function of the mixture model yielding the corresponding precipitation amount (x_A) at a given
18 day.

19 Note that this simulation procedure could be simplified by taking random uniform $[0,1]$
20 numbers instead of Gaussian random numbers. We use the latter here in order to be consistent
21 with the multi-site extension introduced later.

22 Steps 1-4 are repeated over all remaining days within a certain simulation period. Based on
23 this procedure time-series of arbitrary length can be generated that resemble observed
24 climatological precipitation statistics, both in terms of frequency and intensity.

1 1.2 Multi-site

2 So far, the procedure to generate precipitation consists of multiple single-site WGs only.
3 Specifically, no spatial dependence in the simultaneous simulation of precipitation at several
4 sites was taken into account. To close this gap several single-site WGs are driven
5 simultaneously with spatially correlated but serially independent random numbers (Wilks,
6 1998). For simplicity, the concept is illustrated in Supplementary Fig. 1 for the example of two
7 fictitious sites (A and B) only. The extension to several sites is straightforward. One of the
8 main hurdles in simultaneously generating precipitation at several sites is the prescription of
9 the spatial correlation matrices such that the dependence is also preserved in the final
10 generated time-series (Wilks & Wilby, 1999; Wilks, 1998). This difficulty mainly arises from
11 the stochastic process that partly destroys the initially imposed correlation structure again
12 (Wilks, 1998). We will come back to this problem later. For the moment, let us assume that
13 the optimal correlation matrices for both, occurrence and amount (i.e. $\phi_{AB, optim}$ and $r_{AB, optim}$),
14 are known. In this case, the main extensions to single-site WGs are two spatially correlated
15 but serially independent random number streams (dashed boxes in Supplementary Figure 1):
16 one for the occurrence (u) and the other for the amount (v) process. They are determined prior
17 to the simulation process (see below) and contain the same number of days as the simulation
18 period. Once these correlated random number streams are generated, the simulation proceeds
19 as in Supplementary Sect. 1.1 for all stations simultaneously. In practice, the multi-site WG
20 implies the handling of three main methodological hurdles that are the following:

21 (a) Calculating spatial correlation coefficients ϕ_{AB} and r_{AB}

22 Spatial dependence in binary series at site A and B is inferred by the phi-coefficient (ϕ_{AB}).
23 Similarly as the Pearson correlation coefficient, the phi-coefficient ϕ_{AB} is bounded by -1 and
24 1 . For the precipitation amounts, the spatial correlation coefficient (r_{AB}) is determined by the
25 conventional Pearson product-moment correlation coefficient. The correlation is calculated
26 over the whole precipitation series that also include time-steps with zero amounts. From a
27 statistical point of view, this is not an optimal procedure, since the correlation coefficients
28 could be strongly affected by the number of zeros in the time-series. However, the purpose
29 here is to use this spatial similarity measure rather as a tool to compare the observed spatial
30 dependencies with those in artificial data. It is assumed that the statistical limitations in the
31 calculation apply similarly to observations and generated data. The spatial correlations
32 between different sites are determined pair-wise. Note that the pair wise estimation of the

1 inter-station correlation can result in matrices that are not positive definite, especially when
2 the number of station number is large or when there are incomplete station records.

3 (b) *Finding optimal spatial correlation coefficients $\phi_{AB, optim}$ and $r_{AB, optim}$*

4 As mentioned above, imposing observed inter-site correlations as input to our WG does not
5 guarantee its reproduction in the generated series. This is due to a randomization process
6 through transition probabilities calibrated at each site separately. In general, the imposed
7 correlation is reduced by the stochastic process, both in terms of occurrence and amount
8 process. This characteristic is illustrated at an artificial example of two fictitious sites A and B
9 in Supplementary Fig. 2. While the random number streams (u_A and u_B) perfectly incorporate
10 the observed spatial correlation in occurrence between A and B, it is essentially the two
11 distinct transition probabilities at the two sites that lead to a final correlation in the binary
12 series that is much reduced ($\phi_{AB, sim} = 0.6$ compared to $\phi_{AB, obs} = 0.8$). In case of precipitation
13 amounts the mismatch in correlation magnitude is also present ($r_{AB, sim} = 0.38$ compared to $r_{AB, obs} = 0.5$)
14 and can be mainly explained by two factors. First, precipitation amount is only
15 simulated at wet days (i.e. where $J_t' = 1$), while the correlated random number streams ($v_A(t)$
16 and $v_B(t)$) are representative for the full time-series. Hence, the number of zeros introduced by
17 distinct transition probabilities impact on the generated correlation coefficient. Second, if the
18 two fitted PDFs at the two sites are markedly different, the correlation of the observed and
19 simulated precipitation time-series will deviate, even in absence of any zeros.

20 To overcome this inherent problem of a multi-site WG after Wilks (1998), an optimization
21 procedure was proposed to find an input spatial correlation that ultimately yield the target
22 correlation of the observations. This has to be done first for the occurrence process ($\phi_{AB, optim}$)
23 and then in a subsequent step for the amount process ($r_{AB, optim}$). The optimization procedure
24 iterates over an interval of input correlations, thereby running at each iteration the full
25 occurrence and amount model of the multi-site WG (see Supplementary Fig. 3). After each
26 iteration, the resulting correlation is compared to the target correlation of observations. To
27 find an optimal correlation, we use a bisection method (Burden & Faires, 2010) as non-linear
28 root finding algorithm. The iteration is repeated until the generated correlation equals the one
29 of observations with a precision of 0.005 (see Supplementary Fig. 3). Note that this estimation
30 procedure is done prior to the simulation and has to be repeated for each station pair and
31 month.

1 (c) *Generation of correlated random number streams*

2 There are several approaches to generate spatially correlated random numbers streams (e.g.
3 Monahan 2011). For the study at hand we applied a Cholesky decomposition (e.g. Higham
4 2009):

- 5 1. Sample for each station a random number stream from a standard Gaussian
6 distribution.
- 7 2. Apply a Cholesky decomposition to the optimized correlation matrix to get a lower
8 triangular matrix and its transposed.
- 9 3. Multiply the resulting lower triangular matrix with the matrix of random number
10 streams.

11 Cholesky decomposition requires matrices that are positive definite, i.e. that contain no
12 negative eigenvalues. However, in case of the applied pairwise optimization process (see
13 section (2) above) this is not always fulfilled. In absence of positive definite matrices, a fall-
14 back solution based on the nearest positive correlation matrix was chosen. The nearest
15 positive definite matrix was found by using the algorithm proposed by Higham (1989), which
16 uses a weighted version of the Frobenius norm. This problem occurred in our study only a
17 few times. Note, that the temporal correlation structure of the precipitation time-series at one
18 specific site is not altered by the imposed spatial correlation, since the spatially correlated
19 random number streams exhibit no serial correlation.

20

1 **2 References**

2 Burden, R., & Faires, J. D. (2010). *Numerical Analysis*. (Michelle Julet, Ed.)*Numerical*
3 *Analysis* (9th Ed., p. 895).

4 Higham, N.J. (2009). Cholesky factorization. *Wiley Interdisciplinary Reviews: Computational*
5 *Statistics*, 1(2), 251–254. doi:10.1002/wics.18

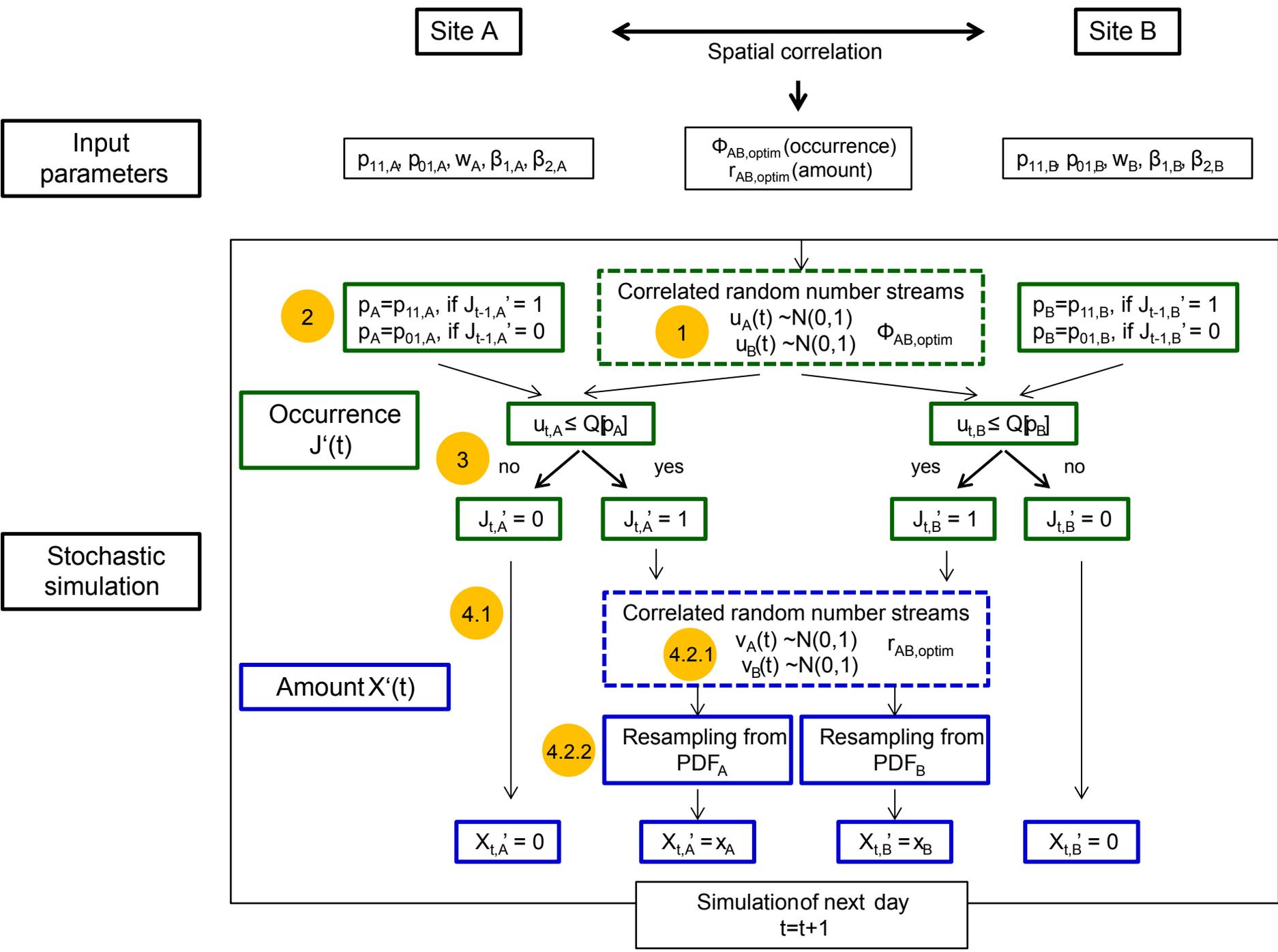
6 Higham, Nicholas J. (1989). Matrix nearness problems and applications. In M. Gover & S.
7 Barnett (Eds.), *Applications of Matrix Theory* (pp. 1–27). Oxford University Press.

8 Monahan, J. F. (2011). *Numerical Methods of Statistics*. *Numerical Methods of Statistics* (p.
9 447). Cambridge: Cambridge University Press.

10 Wilks, D. S. (1998). Multisite generalization of a daily stochastic precipitation generation
11 model. *Journal of Hydrology*, 210(1-4), 178–191. doi:10.1016/S0022-1694(98)00186-3

12 Wilks, D. S., & Wilby, R. L. (1999). The weather generation game: a review of stochastic
13 weather models. *Progress in Physical Geography*, 23(3), 329–357.
14 doi:10.1177/030913339902300302

15



Supplementary Figure 1: Technical workflow of a multi-site precipitation generator at the example of two fictitious sites A and B. In general, it is a combination of multiple single-site precipitation generators that are calibrated at each site individually (see input parameters) and run simultaneously with spatially correlated random number streams (dashed boxes). The correlated random number streams (of similar length as the simulation period) are determined beforehand. The orange-labelled numbers indicate the steps for single-site precipitation simulation.

Occurrence

Correlated random number streams

Observed correlation
 $\Phi_{AB, OBS} = 0.8$

$u_A(t)$
 $u_B(t)$

| | | | | | | | | | | |
|-------|-------|------|-------|------|-------|------|------|------|-------|-----|
| 0.47 | -1.88 | 0.43 | -1.40 | 1.48 | -0.61 | 0.18 | 0.71 | 0.80 | -1.67 | ... |
| -0.03 | -1.45 | 1.15 | -0.86 | 1.95 | -0.37 | 0.51 | 0.85 | 0.87 | -1.74 | ... |

Transition probabilities
 $p_{11,A} = 0.60$ $p_{01,A} = 0.40$
 $p_{11,B} = 0.55$ $p_{01,B} = 0.45$

Simulated precipitation occurrence

Simulated correlation
 $\Phi_{AB, Sim} = 0.6$

$J_A'(t)$
 $J_B'(t)$

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|-----|
| 0 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 1 | ... |
| 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | ... |

time →

Amount

Correlated random number streams

Observed correlation
 $r_{AB, OBS} = 0.5$

$v_A(t)$
 $v_B(t)$

| | | | | | | | | | | |
|-------|-------|-------|------|------|-------|------|------|------|-------|-----|
| -1.04 | -0.05 | -1.34 | 0.69 | 0.20 | -0.01 | 1.09 | 1.53 | 0.19 | -1.20 | ... |
| 0.25 | 0.54 | -1.64 | 0.11 | -0.1 | 0.76 | 2.25 | 1.39 | 1.20 | 0.01 | ... |

Distribution parameters
 $\beta_{1,A} = 1.2$ $\beta_{1,B} = 1.5$
 $\beta_{2,A} = 7.2$ $\beta_{2,B} = 8.0$
 $\alpha_A = 0.7$ $\alpha_B = 0.6$

Simulated precipitation amount [mm/day]

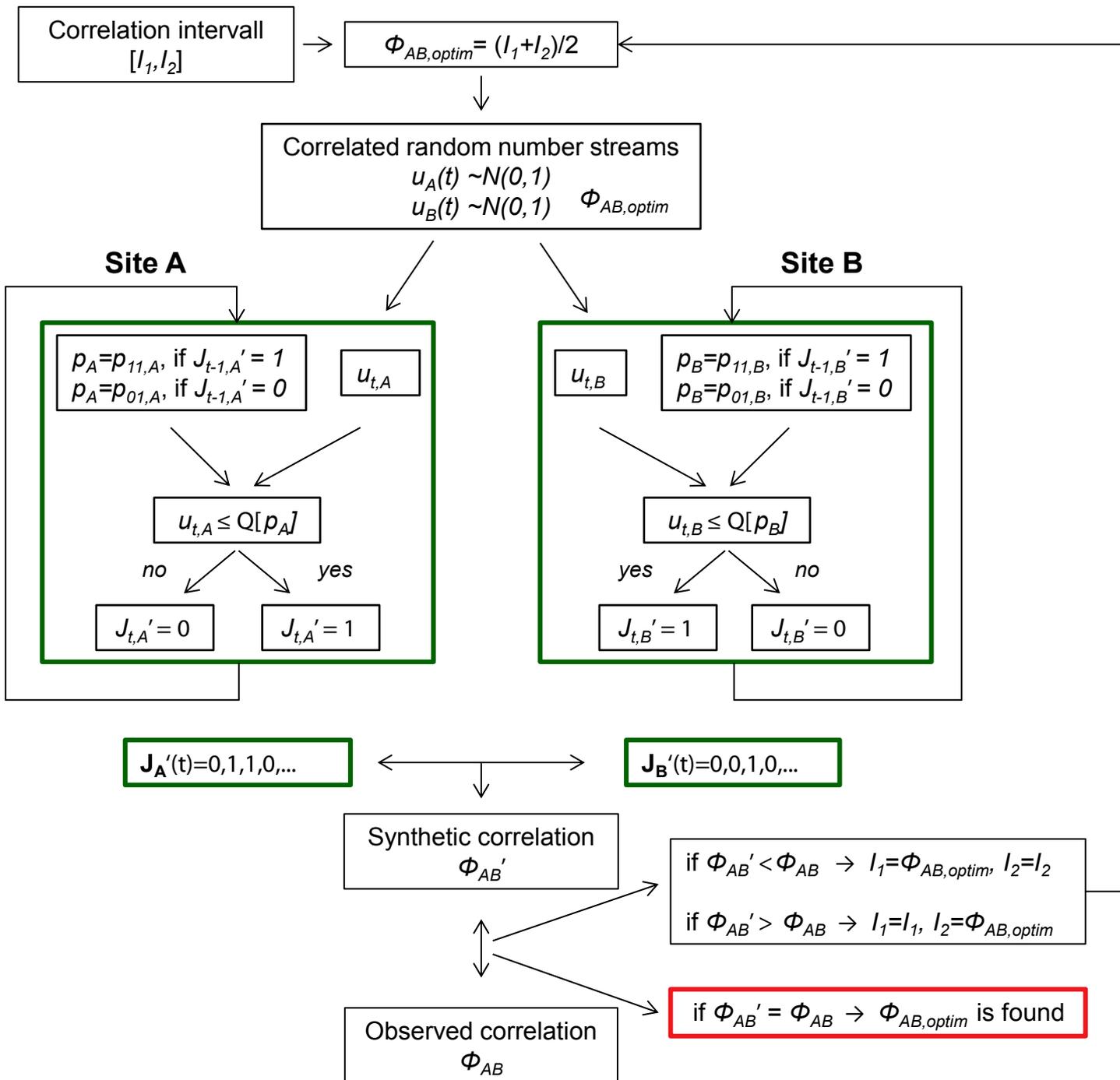
Simulated correlation
 $r_{AB, Sim} = 0.38$

$X_A'(t)$
 $X_B'(t)$

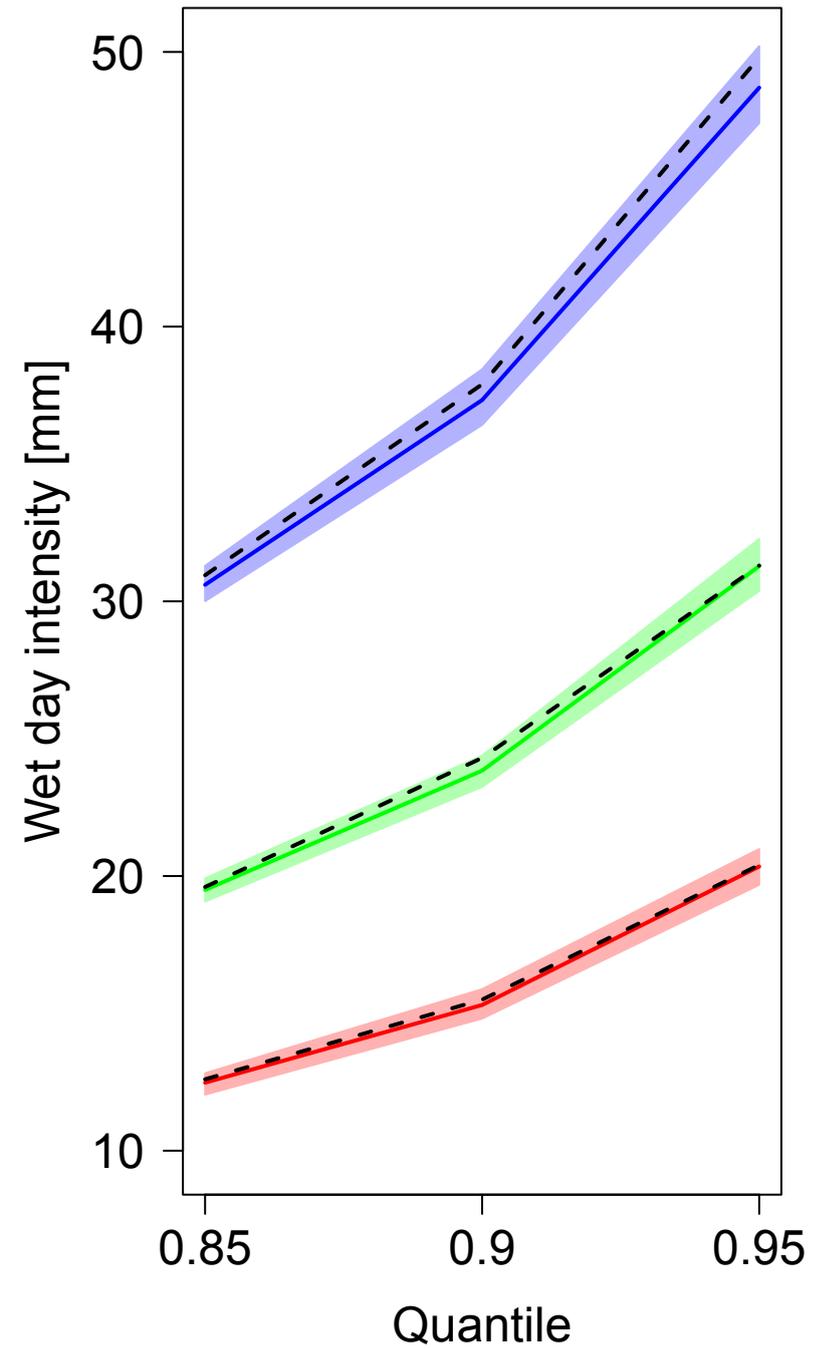
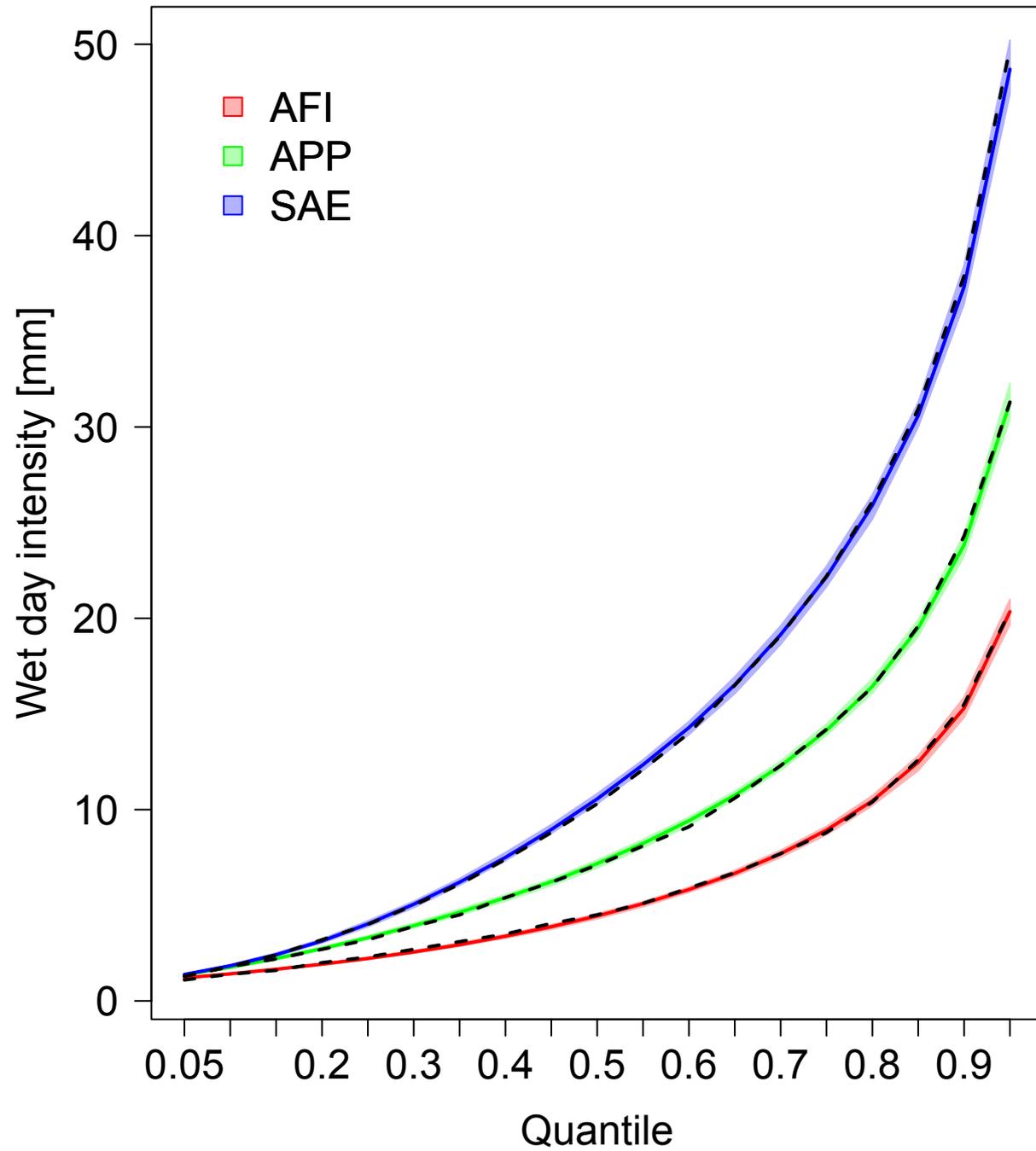
| | | | | | | | | | | |
|------|------|---|------|---|------|------|---|---|------|-----|
| 0 | 1.3 | 0 | 1.88 | 0 | 1.32 | 2.35 | 0 | 0 | 1.04 | ... |
| 1.32 | 1.49 | 0 | 1.26 | 0 | 1.66 | 0 | 0 | 0 | 1.22 | ... |

time →

Supplementary Figure 2: The issue of simulating spatially correlated time-series in a Richardson-type WG for two stations “A” and “B”. The observed spatial correlation (solid box) as input for the occurrence (green) and amount (blue) process is partly destroyed by the stochastic step-by-step simulation of precipitation that is calibrated at single stations (transition probabilities and distribution parameters). This results in a simulated spatial correlation that is lower than observed (dashed box). On the right-hand side, a fictitious example of correlated random number streams and simulated time-series are shown. The pink boxes indicate instances of dry conditions.



Supplementary Figure 3: Iterative procedure to find an optimal spatial correlation coefficient between two stations “A” and “B” ($\Phi_{AB,optim}$) in case of precipitation occurrences. The synthetic correlation is derived from the binary series ($J_A'(t)$ and $J_B'(t)$) and compared to the observed target correlation (Φ_{AB}). A similar procedure is used for precipitation amounts.



Supplementary Figure 4: Quantiles of daily non-zero precipitation amounts, aggregated over 51 years (1961-2011) for the three stations “Andelfingen” (AFI), “Appenzell” (APP) and “Saentis” (SAE). The coloured areas (lines) show for each quantile the estimates of the 95% interval (median) across the 100 model realizations. The black dashed lines represent the observed quantiles.