



# BINGO

a better future under  
CLIMATE CHANGE

BRINGING INNOVATION TO ONGOING  
WATER MANAGEMENT

## D2.6

Ensembles for decadal prediction extremal  
episodes downscaled to 3-1km/ 1h);  
Spatial stochastic precipitation generator  
for catchments

April 2017

[www.projectbingo.eu](http://www.projectbingo.eu)



The BINGO project has received funding from the European Union's Horizon 2020 Research and Innovation programme, under the Grant Agreement number 641739.



Horizon 2020 Societal challenge 5:  
Climate action, environment, resource  
efficiency and raw materials

# BINGO

## Bringing INnovation to onGOing water management – a better future under climate change

Grant Agreement n° 641739, Research and Innovation Action

<b>Deliverable number:</b>	<b>D2.6</b>
<b>Deliverable name:</b>	<b>Ensembles for decadal prediction extremal episodes downscaled to 3-1km/ 1h); Spatial stochastic precipitation generator for catchments</b>
<b>WP / WP number:</b>	WP2
<b>Delivery due date:</b>	30/04/2017
<b>Actual date of submission:</b>	28/04/2017
<b>Dissemination level:</b>	public
<b>Lead beneficiary:</b>	FUB
<b>Responsible scientist/administrator:</b>	Uwe Ulbrich, Henning Rust
<b>Estimated effort (PM):</b>	2
<b>Contributor(s):</b>	Komlan Kpogo-Nuwoklo, Edmund Meredith, Christos Vagenas (all FUB)
<b>Estimated effort contributor(s) (PM):</b>	14
<b>Internal reviewer:</b>	Corrado Camera (CYI), Tim aus der Beek (IWW)

### **Changes with respect to the DoW**

In the DoW, FUB was committed to creating high-resolution extremal episode ensembles for all research sites except Cyprus. With respect to this for the decadal predictions, the same change as was made for D2.4 (high-resolution downscaling of extremal episodes for the present climate) has been made for D2.6, too. Namely, high-resolution downscaling was not performed for the Veluwe catchment. As explained in full detail in D2.4, we were informed by the colleagues responsible for the Veluwe catchment that, different from the original plan, they do not require such high-resolution simulations. The computational resources saved by these changes were re-allocated to create larger ensembles of extremal episodes for the other research sites, providing more data for the hydrological modellers to work with and to produce more robust results.

### **Dissemination and uptake**

Both downscaled extremal episodes and data from the stochastic precipitation generator will be used by the Research Sites' modelling groups.

### **Short Summary of results (<250 words)**

This deliverable consists of two components. First, a spatial stochastic precipitation generator is developed for the catchments. Second, extremal episodes are identified from the decadal predictions and downscaled with the COSMO-CLM to a resolution of 2.2-km and 1 hour for all sites except Veluwe (see changes with respect to the DoW) and Cyprus. The corresponding high-resolution downscaling for Cyprus is performed with WRF and presented in D2.7.

### **Evidence of accomplishment**

This report and its attachments, the regional climate model simulations available for download via DECO (the Data Extraction and Conversion platform of FUBs FREVA Web-System).

## Table of contents

<b>1.Introduction .....</b>	<b>4</b>
<b>2.Spatial stochastic precipitation generator .....</b>	<b>6</b>
<b>3.Dynamical downscaling of extremal episodes from decadal predictions .....</b>	<b>14</b>
3.1 MiKlip decadal predictions.....	14
3.2 High-resolution ensembles of extremal episodes from decadal-predictions (2015-2024). .....	15
3.3 Accessing the extremal episode data .....	16
<b>4.Summary .....</b>	<b>17</b>
<b>5.References .....</b>	<b>18</b>
<b>Appendix A: .....</b>	<b>20</b>

## 1. Introduction

Deliverable 2.6 consists of two major tasks. The first task is the development of a spatial stochastic precipitation generator for the catchments. The second task is the identification of extremal-episodes from the decadal predictions and their high-resolution downscaling to (0.02°/2.2-km), for all research sites except Cyprus (please see also the changes with respect to the DoW above). The high-resolution downscaling for Cyprus is performed separately by CYI, with a different regional climate model, and is described in D2.7.

The first part, development of a daily spatial stochastic precipitation model of rainfall, is discussed in Section 2. The developed stochastic generator is based on a censored and transformed latent Gaussian field. This type of stochastic generator is widely used and many previous studies have demonstrated its ability to simulate realistic spatial rainfall fields (see [Glasbey and Nevison, 1997], [Guillot and Lebel, 1999], [Kleiber et al., 2012] among others). The theoretical aspects of developing such a spatial stochastic precipitation generator are described in detail in a document which is attached to this deliverable (Appendix A). Here, we provide a summary of the main stages in the model development and present some validation results.

The second part, the modelling component, involves the identification of extremal episodes from the 12-km decadal predictions (2015-2024) and their subsequent downscaling to high-resolution (0.02°/2.2-km) in order to produce ensembles of extreme precipitation events for all research sites for the period 2015-2024. The 12-km predictions were previously downscaled from a coarser model as part of D2.2. As in D2.4, the high-resolution ensembles are produced by selectively dynamically downscaling the 12-km simulations. This selective downscaling utilises the methodology for selecting extremal episodes developed for D2.4.

Note also that, as with D2.4, the original incarnation of this deliverable also incorporated the creation of medium-resolution (7-km) extremal episode ensembles. As was explained and justified in the 1st Periodic BINGO Technical Report (Part B), the medium-resolution downscaling of the 12 km simulations to 7 km has been skipped. After a review of this step and discussions with experienced regional climate modellers, the FUB team decided that dynamically downscaling the 12-km simulations to a

resolution of 7-km offers little additional benefit and does not represent optimal usage of limited computing resources. For a resolution increase of only ~40%, the computational expense would be quite considerable. In addition to this, a horizontal resolution of 7-km is at the border of where dynamical models can and cannot explicitly resolve convective processes, the so-called “grey zone”. It may be sufficient to resolve some large-scale convection, though is insufficient for all other convection. For this reason, it is generally considered good practice to avoid running models relying on convective parameterization schemes at this type of resolution.

Instead, we directly downscale the 0.11°/12-km simulations to a horizontal resolution of 0.02°/2.2-km. This represents a downscaling ratio of 5.5 to 1, which is well within the range of what is considered acceptable in the regional climate modelling literature [Denis et al., 2003]. The resources previously allocated to this task have been re-allocated (i) towards producing a greater number of ensemble members for the 0.02° simulations, resulting in more robust results, and (ii) to research leading to more options and thus an enhanced flexibility in the methodology for the identification of extremal periods. As with previous deliverables, all simulation data are provided via the DECO plugin.

## 2. Spatial stochastic precipitation generator

For BINGO Research Sites where station data are available, we aim to develop a spatial stochastic precipitation generator additionally to RCM simulations. Such a spatial model allows generation of large ensembles of precipitation sequences with limited computational time and expense, and thus allows uncertainty assessment. Compared to RCMs, stochastic generators generally have low computational cost and provide relatively fast simulations. The development of this model is described in Appendix A, the following section outlines the main features of the model and gives a summary of the results for the Wupper catchment as an example.

### 2.1 Stochastic model

The underlying idea of the spatial stochastic precipitation generator is to use a latent Gaussian field to model the occurrence process and to transform the same latent field to model the positive rainfall amounts. The use of the same Gaussian field to simulate both occurrence and intensity is preferred for two main reasons. First, the number of parameters is reduced; second, it allows avoiding the edge effect, in which unrealistically large intensities can be generated near the boundary of dry areas when the occurrence and intensity processes are modelled separately (see [Bardossy and Plate, 1992], [Wilks, 1998] ). In this model, we assume the precipitation process is driven by a latent Gaussian random field such that wet days correspond to positive values of the Gaussian field and dry days to negative or null values. To fully characterize the model, three components are required: (i) the marginal model of rainfall amount; (ii) the mean and (iii) the covariance functions of the Gaussian field.

#### **Marginal model**

We use a gamma distribution to model the precipitation amount at individual stations. To account for seasonality and dependence on broad scale atmospheric conditions, the shape and scale parameters of the gamma distribution are modelled by linear functions of sine and cosine waves and large scale climate covariates. The large-scale climate covariates are from ERA-interim reanalysis (Dee et al., 2011). Variables that are used are CAPE (convective available potential energy, an instability metric), wind, temperature, dewpoint temperature, geopotential, specific humidity, relative humidity, vertical velocity. Covariates that lead to the “best” model are selected according to Akaike information criterion (AIC) .

### Mean function

The mean function of the Gaussian field is estimated by inverting the wet day probability at each location. The wet day probability is estimated through a logistic regression with multiple covariates. As for the marginal model, we also use covariates harmonic functions and climate covariates from ERA-interim.

### Covariance function

We first estimate the empirical covariance function on the transformed rainfall data for wetdays using the maximum likelihood approach (see [Glasbey and Nevison, 1997]). We then fit a parametric spatio-temporal covariance function, the non-separable Gneiting covariance function [Gneiting, 2002], assuming the covariance function only depends on the distance between two locations and the time lag.

Note that direct estimates of the marginal model and the mean function parameters are available only at locations with observed data. In order to generate a complete precipitation field across the entire simulation domain, we spatially interpolated the model parameters over the simulation domain using kriging.

## 2.2 Case of study: spatial stochastic precipitation generator for Wupper catchment

We consider a network of 22 rainfall stations, located in and around the Wupper catchment as shown in Figure 2.1. The RS Wupper is chosen because of the excellent data availability. The data consist of accumulated daily rainfall from 1980 to 2015.

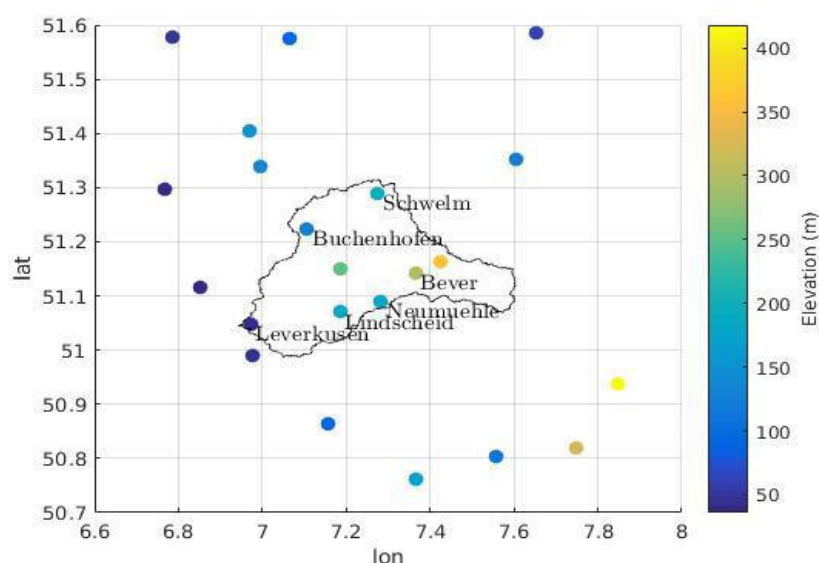


Figure 2.1 - The locations of the stations

We generate 100 ensembles of precipitation sequences for the period covered by observations (1980-2015) and then examine the results.

### Temporal features

- Frequency of wet days

We estimate the frequency of wet days from observations and then compare it to the ones obtained from simulations. The result is gathered in Figure 2.2. The model replicates well the occurrence of the wet days.

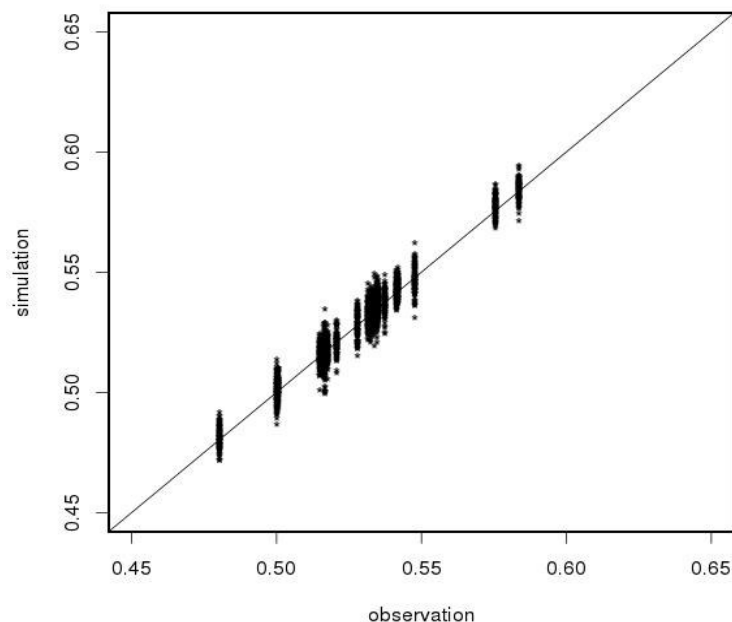


Figure 2.2 : Frequency of wet days

- Dry day spells length

An essential feature that a stochastic precipitation generator attempts to replicate is the length of dry and wet spells. Figure 2.3 displays the log frequency of dry spells at six stations within the Wupper catchment, with the mean and 90% pointwise confidence intervals based on the 100 simulations overlaid. The stochastic generator shows a good characterization of dry spell behaviour at individual locations. The same results (see Appendix A) are obtained for the length of wet spells.

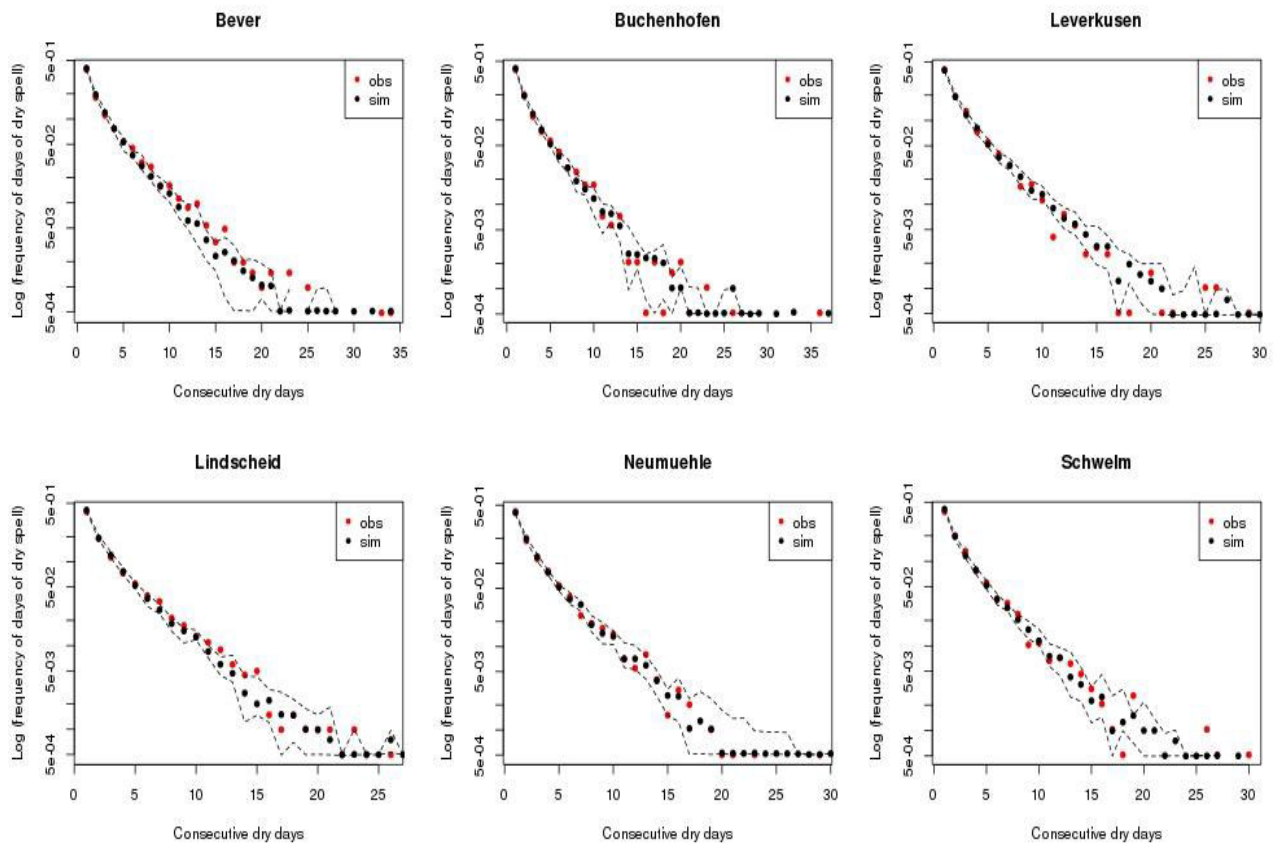


Figure 2.3 : Log frequency of empirical dry spell (red dot) and the mean of the ones based on simulations (black dot) with a 90% confidence interval superimposed.

- Seasonal cycle

We turn next to the ability of the stochastic generator to well reproduce the seasonal cycle of precipitation. In Figure 2.4, the monthly mean of precipitation obtained from observations and the ones obtained from simulations are shown for six stations. The model reproduces very well the seasonal cycle of precipitation. The root mean square errors (RMSE) between the observed monthly mean and the mean of simulation monthly means are very low (less than 0.15 mm) and the correlations between them are high (greater than 0.9).

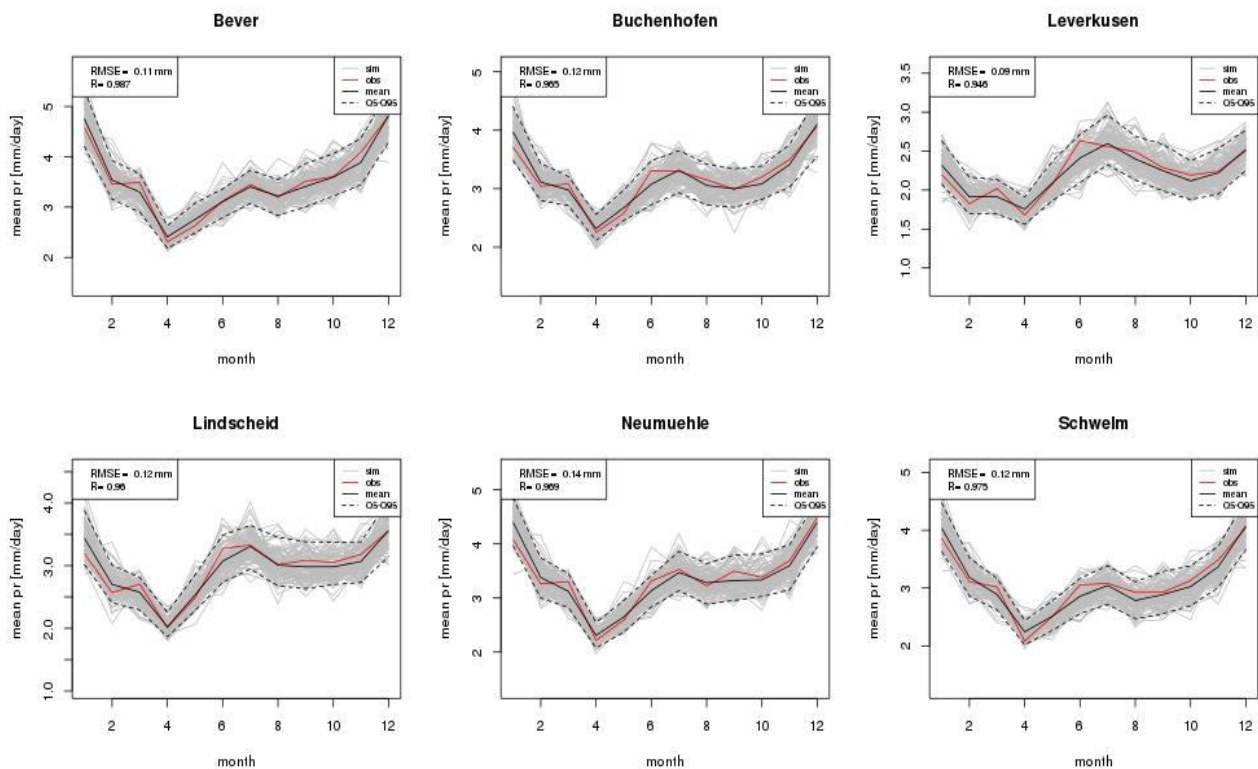


Figure 2.4 : Monthly mean of precipitation: observation (red line), 100 simulations (grey lines), mean of the 100 simulations (black line) and 90% confidence interval (dashed black line)

### Spatial features

- Joint occurrence probability

A positive feature of the defined stochastic model is correlating wet and dry occurrences across space. In order to illustrate the fit of the spatial dependence, we consider all pairs of stations and then compute, for observations and simulations, the probability of simultaneously wet, simultaneously dry, and the probability that one site is wet and the second is dry. The results are shown in Figure 2.5. The stochastic generator seems to replicate the joint probability of precipitation occurrence state quite well; the correlations between the observed joint probabilities and the simulated ones are most of the time greater than 0.9.

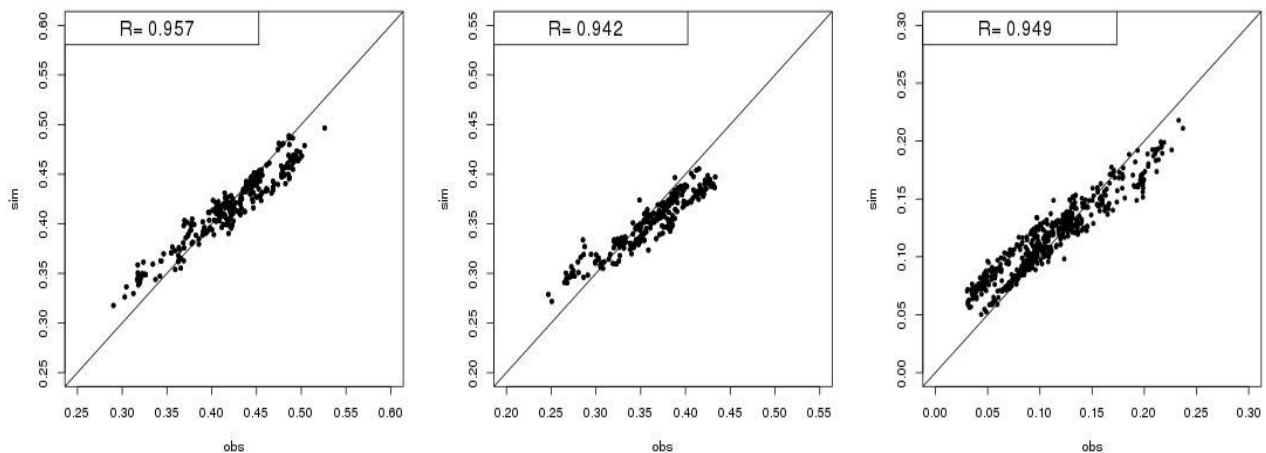


Figure 2.5 :Joint probability of precipitation occurrence state of two sites:  
simultaneously wet (left), simultaneously dry (middle), one site is wet and the second is  
dry (right)

- Spatial correlation

Figure 2.6 shows the spatial correlation of rainfall for observations (red line) and for simulations (grey lines) and 90% pointwise confidence intervals based on the 100 simulations overlaid for each season. Except for summer (JJA), there is little underestimation of the correlation. It could be attributed to the maximum likelihood approach used to estimate the empirical correlation of the Gaussian field. This underestimation may slightly reduce the variability in the simulated precipitation.

Additional results may be found in Appendix A, in particular cross-validation results. In general, we feel that the model replicates quite well the spatial aspects of precipitation process. Some propositions to further improve the model may also be found in Appendix A.

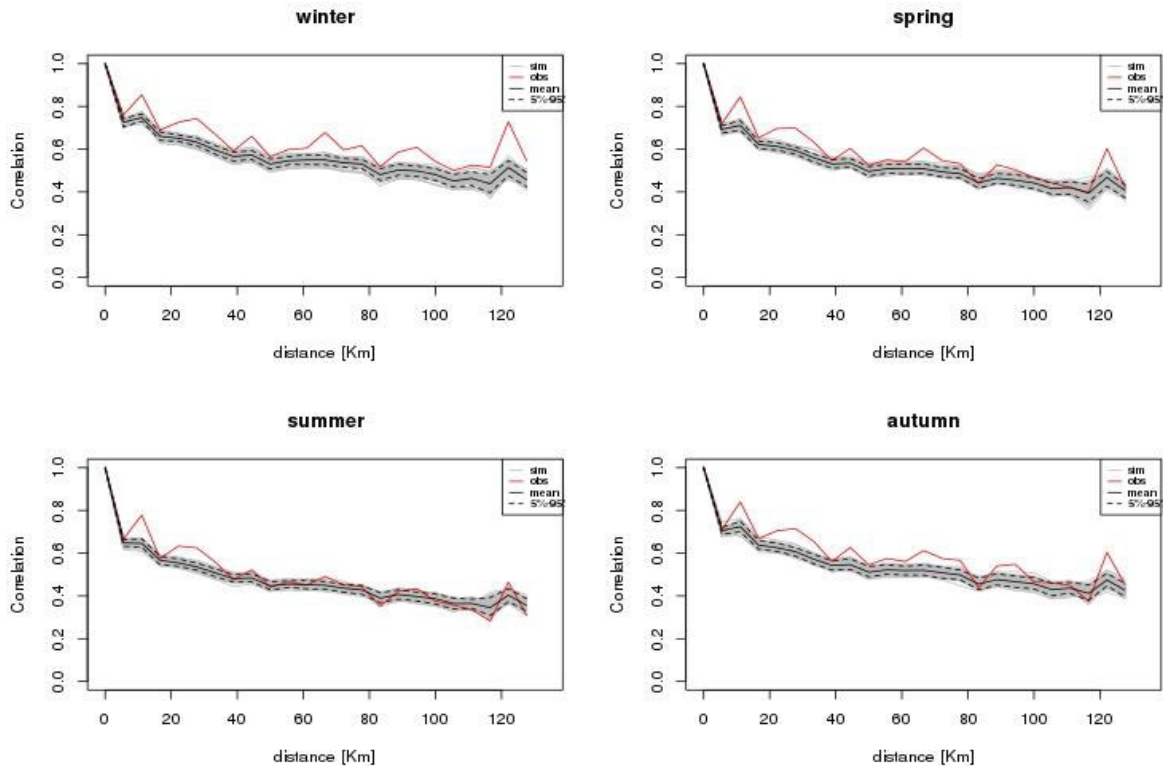


Figure 2.6 : Spatial correlation of precipitation: observation (red line), 100 simulations (grey line), mean of the 100 simulations (black line) and 90% confidence interval (dashed black line)

### 2.3 Adaptation of the stochastic model to other BINGO Research Sites

The spatial stochastic precipitation generator we developed is applicable for the other BINGO Research Sites as well. This requires two main analyses :

1. Identification of sets of covariates that give the « best » model for the precipitation amount and the mean function of the Gaussian field at the location of stations.
2. Identification of a class of parametric covariance function that fits best the empirical covariance.

### 2.3 Disaggregation of daily precipitations into hourly values

At this stage of developpement, the stochastic generator is able to simulate daily rainfall. Some sites requested a disaggregation of the daily precipitation into hourly values. To do so, we will adopt an approach similar to the one proposed by [Wheater et

D2.6 Spatial stochastic precipitation generator.  
High-resolution downscaling of predictions.  
April 2017.



al., 2005] and also used by [Segond et al., 2007]. First, the daily precipitation field is simulated with the Gaussian field based stochastic generator described above. Then a disaggregation at each location is done using a single site generator based on Poisson-cluster processes (see [Koutsoyiannis and Onof, 2001] ).

## **Dynamical downscaling of extremal episodes from decadal predictions**

This section discusses the numerical modelling component of D2.6. A brief recap of decadal predictions and the MiKlip system is first presented. This is then followed by a discussion of the high-resolution downscaling carried out for D2.6.

Worth noting is that the high-resolution downscaling procedure for D2.6 is almost identical to the procedure followed in D2.4, where extremal episodes for the present climate period were downscaled. For concision, aspects already described in detail in D2.4 are not repeated in such detail here. Instead, more space is given to describe the (few) aspects which differ from the procedure described in D2.4.

### **3.1 MiKlip decadal predictions**

As a brief recap of Deliverable 2.2, the relatively new field of decadal climate prediction, e.g. Smith et al. [2007], aims to simulate both the climate response to future anthropogenic forcing and the future evolution (from the present) of the climate due to internal climate variability [Marotzke et al., 2016]. This differs from the approach taken in climate projections, e.g. the CMIP5 project [Taylor et al., 2012], where the focus is on the response of the climate to anthropogenic forcing and the impacts of internal climate variability are (supposed to be) nullified via multi-decadal climate model integrations. The earth system models (ESMs), e.g. in our case the MiKlip system, used in decadal prediction systems are initialized with an observed state of the climate system, i.e. ocean, atmosphere, soil, ice, etc. Skill in predicting internal climate variability on a decadal scale is derived from the long-term memory (i.e. sensitivity to the initial state) of certain components of the climate system, predominantly the ocean. As such, decadal predictions (unlike climate projections) are reliant on a high-quality initialization of the ESM for those components which exhibit long-term memory.

In Deliverable 2.2, coarser realizations of the MiKlip 2015-2024 predictions were dynamically downscaled to 12-km resolution for two sub-European domains. The first covering Iberia (i.e. the sites at Badalona and Tagus) and the second covering north-west Europe (i.e. the sites at Bergen, Wupper and Veluwe). This was carried out for 10 realizations of the 2015-2024 predictions, creating 100-years of data in total. These simulations serve as the basis for the high-resolution downscaling of the present deliverable.

### 3.2 High-resolution ensembles of extremal episodes from decadal-predictions (2015-2024).

The objective of this downscaling step is to create high-resolution ensembles of extremal episodes for each research site.

#### Theory and Model Setup

As discussed in detail in D2.4, the added value provided by downscaling extremes to high-resolution is invaluable to hydrological modellers. The more detailed fine-scale data enhance the accuracy of the modelling. Added value is also provided through a better simulation of convective processes and a more realistic diurnal convection cycle [e.g. Prein et al., 2015]. In addition to this, high-resolution simulations can also modify the local climate change signal of large-scale models [e.g. Kendon et al., 2014].

The modelling setup and underlying theory behind the downscaling of the extremal episodes to 0.02° (2.2-km) resolution is *identical* to that described in D2.4, which the reader is referred to for further details. Indeed, for consistency between the evaluation period and the decadal predictions, model settings were specifically kept constant.

#### Identification and Downscaling of Extremal Episodes

A methodology for identifying extremal episodes has been developed and presented in detail in D2.4, where it was applied to the 12-km evaluation period simulations. For D2.6, the *exact same* method is used again and applied *identically* to the 12-km decadal prediction simulations. For a detailed discussion of the method, the reader is referred to D2.4. For a briefer recap, the method is now summarised.

The methodology for identifying extremal episodes can be summarised as follows:

1. Observational precipitation data are used to identify days with extreme precipitation
2. The observed (i.e. reanalysis) large-scale circulation patterns on all of these days are taken and are subject to a cluster analysis
3. The resulting clusters are used as reference extremal weather patterns, against which all days from the 12-km CCLM simulation are compared. Individual days are either rejected or passed based on their similarity to an extremal pattern.

4. If a day is passed, local-scale meteorological factors conducive to heavy precipitation events are then analysed. If these meteorological factors suggest a chance of heavy precipitation in the vicinity of the research site, then that day is selected for downscaling.

#### Differences to the downscaling approach of D2.4 (evaluation period)

As described in Section 3.1 and Deliverable 2.2, ten realizations of the MiKlip decadal predictions (2015-2024) at 12-km resolution had previously been produced, giving a total of 100 years of data. This is about 2.7-times greater than the evaluation runs (1979-2015). As such, applying the same methodology to the MiKlip runs results in a similarly greater number of extreme days being identified, which are then split into events to be downscaled. To maintain consistency between the 0.02° MiKlip and evaluation simulations, the number **N** of MiKlip events downscaled for each site to 0.02° is kept the same as the number **N** of events downscaled from the evaluation runs. These **N** events were selected at random from all of the identified MiKlip events.

### 3.3 Accessing the extremal episode data

As with all simulation data produced by FUB, the data are available to download via the DECO plugin. The procedure is basically the same as for the evaluation period extremal episodes, except that users just need to select the extremal-episodes-miklip experiment. With the download, users will be presented with one data file for each extremal episode that has been downloaded. Accompanying these data files will be a text file listing all days that were identified as being potentially extreme.

### **3. Summary**

The developed spatial stochastic precipitation generator relies on a latent Gaussian random field that drives both the rainfall occurrence and the rainfall intensity processes. The model was used to simulate spatially correlated fields across the Wupper catchment covering the time period from 1980 to 2015. We have shown that the stochastic generator is able to reproduce not only the occurrence and the amount of rainfall at individual stations, but also the wet- and dry-day sequences as well as the correlation between different stations quite well.

The method developed for identifying extremal episodes for D2.4 (evaluation period) has been applied to the decadal predictions (2015-2024). This has been used to identify potentially extreme days for the sites at Badalona, Bergen, Tagus, and Wupper, which were then downscaled to high resolution (2.2-km), with an hourly time resolution. The data fields required by the hydrological modelling groups are available via the DECO plugin of the FreVa web platform in the specific formats requested by each research site.

## 4. References

- Bardossy, A. and Plate, E. J. (1992). Space-time model for daily rainfall using atmospheric circulation patterns. *Water Resources Research*, 28(5):1247–1259.
- Dee et al. (2011), The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q.J.R. Meteorol. Soc.*, 137: 553–597. doi:10.1002/qj.828
- Denis B., et al. (2003). Sensitivity of a regional climate model to the resolution of the lateral boundary conditions. *Climate Dynamics*, 20(2-3), 107-126.
- Glasbey, C. and Nevison, I. (1997). Rainfall modelling using a latent gaussian variable. In *Modelling Longitudinal and Spatially Correlated Data*, pages 233–242. Springer.
- Gneiting, T. (2002). Nonseparable, stationary covariance functions for space–time data. *Journal of the American Statistical Association*, 97(458):590–600.
- Guillot, G. and Lebel, T. (1999). Approximation of sahelian rainfall fields with metagaussian random functions. *Stochastic Environmental Research and Risk Assessment*, 13(1-2):113–130.
- Kendon E. J., et al. (2014). Heavier summer downpours with climate change revealed by weather forecast resolution model. *Nature Climate Change*, 4(7):570-576.
- Kleiber, W., Katz, R. W., and Rajagopalan, B. (2012). Daily spatiotemporal precipitation simulation using latent and transformed gaussian processes. *Water Resources Research*, 48(1).
- Koutsoyiannis, D. and Onof, C. (2001). Rainfall disaggregation using adjusting procedures on a poisson cluster model. *Journal of Hydrology*, 246(1):109–122.
- Marotzke J, et al. (2016). Miklip – a national research project on decadal climate prediction. *Bulletin of the American Meteorological Society*, 2016.
- Prein A. F., et al. (2015). A review on regional convection-permitting climate modeling: demonstrations, prospects, and challenges. *Reviews of Geophysics*, 53, 323-361.
- Segond, M.-L., Neokleous, N., Makropoulos, C., Onof, C., and Maksimovic, C. (2007). Simulation and spatio-temporal disaggregation of multi-site rainfall data for urban drainage applications. *Hydrological Sciences Journal*, 52(5):917–935.

Smith D. M., et al. (2007). Improved surface temperature prediction for the coming decade from a global climate model. *Science*, 317(5839):796-799.

Taylor K. E., et al. (2012). An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.*, 94(4):485-498.

Wheater, H., Chandler, R., Onof, C., Isham, V., Bellone, E., Yang, C., Lekkas, D., Lourmas, G., and Segond, M.-L. (2005). Spatial-temporal rainfall modelling for flood risk estimation. *Stochastic Environmental Research and Risk Assessment*, 19(6):403–416.

Wilks, D. (1998). Multisite generalization of a daily stochastic precipitation generation model. *Journal of Hydrology*, 210(1):178–191.

D2.6 Spatial stochastic precipitation generator.  
High-resolution downscaling of predictions.  
April 2017.



## **Appendix A:**

**Daily spatio-temporal stochastic precipitation generator based on a censored and transformed latent Gaussian field.**

# Appendix A

—

## Daily spatio-temporal stochastic precipitation generator based on a censored and transformed latent Gaussian field

Komlan A. Kpogo-Nuwoklo

April 26, 2017

### 1 Introduction

Since recent decades, there is an increasing demand for high-resolution precipitation data for impact studies (hydrological, economic, agricultural, etc). Rainfall data are generally collected at coarser scales than required, hence the need for a spatio-temporal model which allows simulation of rainfall at fine scales, yet consistent with observations. Stochastic precipitation generators have been key features of climate impact studies for a number of decades. They are able to reproduce accurately the spatio-temporal dynamics as well as the natural variability of the rainfall distribution. In the context of the HORIZON 2020 project BINGO (Bringing INnovation to onGOing water management), we aim to develop a spatial stochastic precipitation generator for research sites where observations are available. Stochastic generators generally have a low computational cost and provide relatively fast simulations. They allow simulation of large ensembles of precipitation sequences and thus allow uncertainty assessment.

A large number of spatio-temporal rainfall models have been suggested in the literature. Overviews were provided by Wilks and Wilby [1999], Srikanthan and McMahon [2001], Mehrotra et al. [2006], Wilks [2010], and more recently by Ailliot et al. [2015]. Point process models are one class of precipitation stochastic generators. This class seeks to represent the physical rainfall process with a hierarchical structure of "rain cells" clustering within "storms" living inside a "rain event", albeit in a highly idealized fashion (see Onof et al. [2000], Chandler et al. [2006]).

Another type of approach is to define a latent Gaussian variable for which dry conditions correspond to censored values below a given threshold (see Allcroft and Glasbey [2003], Ailliot et al. [2009], Kleiber et al. [2012], Allard and Bourotte [2015], Baxevani and Lennartsson [2015]). Kleiber et al. [2012] used two different latent Gaussian fields: one for the occurrence process of rainfall and another for its intensity. Some authors used the same latent Gaussian field for occurrence and intensity (Allcroft and Glasbey [2003], Allard and Bourotte [2015], Baxevani and Lennartsson [2015]). The last option

is preferred for two main reasons. First, the number of parameters is reduced and, second, to avoid the edge effect, in which unrealistically large intensities can be generated near the boundary of dry areas when the occurrence and intensity processes are modelled separately (see Bardossy and Plate [1992], Wilks [1998], Baxevani and Lennartsson [2015]).

We choose in this work to use a censored latent Gaussian field to model the occurrence process and to transform the same latent field to model the positive rainfall amounts. The censored, i.e. positive Gaussian field can not be considered as a good representation of daily rainfall amounts. Thus, for wet days, a gamma distribution is used to transform Gaussian values into rainfall amounts. The proposed model takes into account the seasonality and the dependence of rainfall on large-scale atmospheric conditions.

This document is organized as follows: the proposed model and estimation parameters are described in section 2. In section 3, the validity of the model is shown. The Wupper catchment, which is one of the six BINGO research sites, is used as a showcase. The validation of the model is based on its ability to reproduce a series of temporal and spatial dependence.

## 2 Model

The proposed model uses a latent Gaussian field to model the occurrence process and a transformation of the same latent field to model the positive rainfall amounts, see Fig. 1. We assume the precipitation process  $Y(\mathbf{s}, t)$  at location  $\mathbf{s}$  and time  $t$  depends on a latent Gaussian random field  $Z(\mathbf{s}, t)$  such that a wet day corresponds to a positive value of  $Z(\mathbf{s}, t)$  and a dry day to negative or null value of  $Z(\mathbf{s}, t)$ :

$$Y(\mathbf{s}, t) = \begin{cases} \psi(Z(\mathbf{s}, t)), & \text{if } Z(\mathbf{s}, t) > 0 \\ 0, & \text{if } Z(\mathbf{s}, t) \leq 0 \end{cases} \quad (1)$$

where  $\psi(Z(\mathbf{s}, t))$  is referred to as meta-Gaussian with  $\psi$  a non-decreasing function referred to as anamorphosis. There are two main advantages of using a meta-Gaussian random field. First, it is compatible with any marginal distribution (see Guillot and Lebel [1999]). Second, the Gaussian random field is fully specified by its mean and its covariance function while exhibiting remarkable flexibility. The model is fully defined by three components:

- the anamorphosis  $\psi$
- the mean of the Gaussian random field,  $\mu(\mathbf{s}, t) = \mathbb{E}[Z(\mathbf{s}, t)]$
- the covariance function of the Gaussian random field.

### 2.1 Anamorphosis

The role of  $\psi$  is to transform the Gaussian values into precipitation amounts for wet days. A precipitation amounts model at a single location has received considerable attention

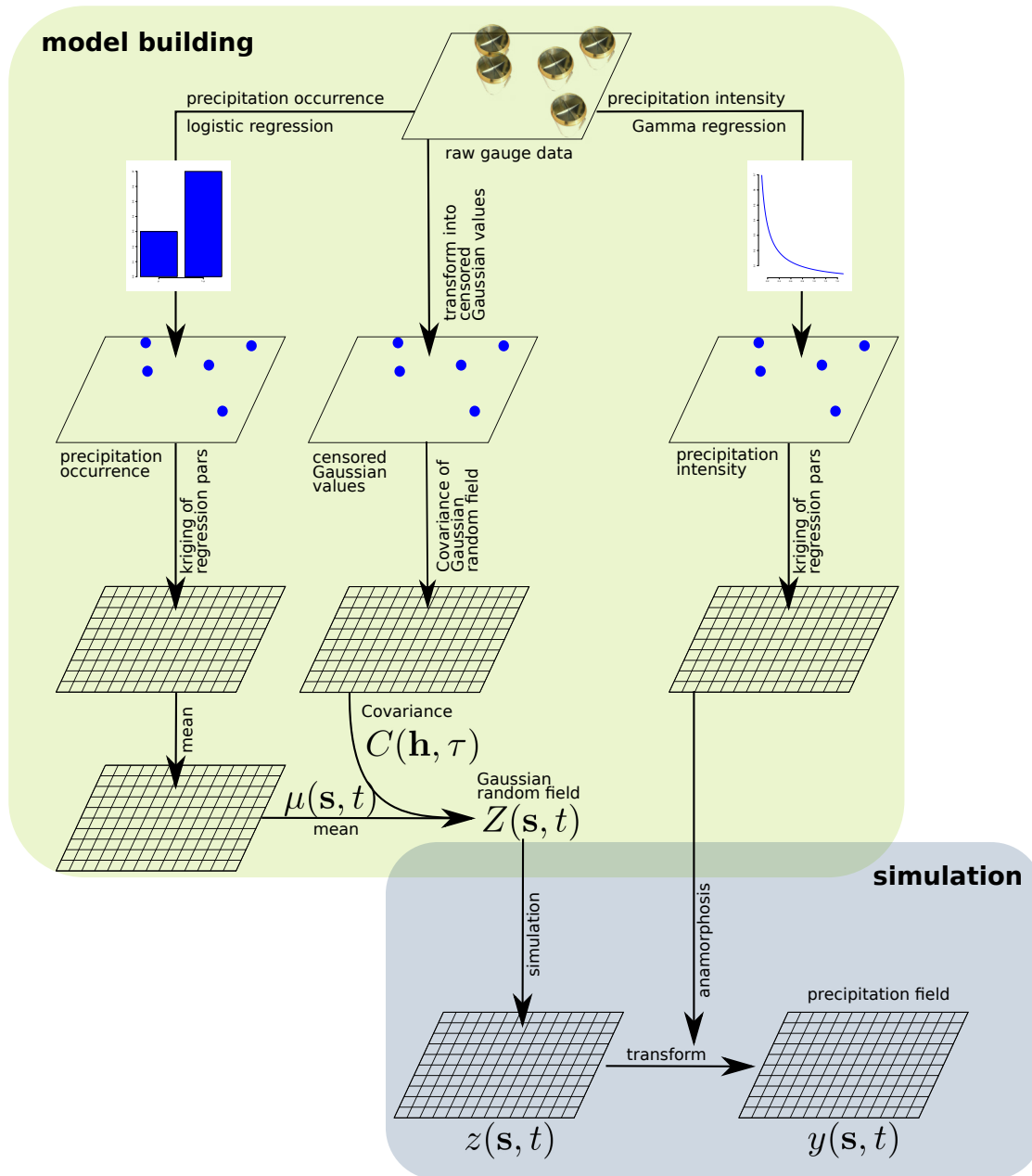


Figure 1: Schematic plot of the stochastic precipitation generator from gauge based precipitation measurements to simulated precipitation fields.

over the past few years. Models that are traditionally used are: Box-Cox transformation Box and Cox [1964], mixtures of exponential distribution Wilks [1998], a quadratic power transformation Glasbey and Nevison [1997], or a power-exponential transformation Allard and Bourotte [2015]. In Kleiber et al. [2012], a gamma distribution was used, while Lennartsson et al. [2008] and Serinaldi and Kilsby [2014] used a generalized Pareto distribution to model heavy rainfall above a threshold. Baxevani and Lennartsson [2015] used a composite model of a gamma distribution for precipitation amounts below some threshold with a generalized Pareto distribution modelling the excesses above the threshold.

In this work, the precipitation amount is modelled by a gamma distribution and the anamorphosis is written as follows:

$$\psi(\cdot) = G_{\mathbf{s},t}^{-1} \circ \Phi_{\mu}(\cdot), \quad (2)$$

where  $G_{\mathbf{s},t}$  is the cumulative distribution function (cdf) of a gamma distribution with shape parameter  $\alpha(\mathbf{s}, t)$  and scale parameter  $\gamma(\mathbf{s}, t)$  at location  $\mathbf{s}$  and time  $t$  and  $\Phi_{\mu}$  is the cdf of a censored normal random variable with mean value  $\mu$  and unit variance. Therefore, the anamorphosis  $\psi$  is fully parametrized by the shape and scale parameters of the gamma distribution. To account for seasonality and dependence on large-scale atmospheric conditions, the shape and scale parameters of the gamma distribution are modelled by linear functions of sine and cosine waves, and large-scale climate covariates,

$$\log(\alpha(\mathbf{s}, t)) = \beta'_{\alpha}(\mathbf{s})X(\mathbf{s}, t) \quad (3)$$

$$\log(\gamma(\mathbf{s}, t)) = \beta'_{\gamma}(\mathbf{s})X(\mathbf{s}, t) \quad (4)$$

where  $X(\mathbf{s}, t)$  are the covariates and  $\beta_{\alpha}(\mathbf{s})$  and  $\beta_{\gamma}(\mathbf{s})$  are regression parameters. The estimation of the parameters are made by using the R package VGAM Yee [2015].

## 2.2 Mean function

The mean of the censored Gaussian field is linked to the precipitation data through the probability of wet days. If  $p(\mathbf{s}, t)$  denotes the probability that a location  $\mathbf{s}$  is wet at day  $t$ ,

$$p(\mathbf{s}, t) = P(Y(\mathbf{s}, t) > 0) = P(Z(\mathbf{s}, t) > 0) = \Phi(\mu(\mathbf{s}, t)) \quad (5)$$

where  $\Phi$  is the cdf of the standard normal random variable. Estimation of the mean value  $\hat{\mu}(\mathbf{s}, t)$  is therefore straightforward by simply inverting (5):

$$\hat{\mu}(\mathbf{s}, t) = \Phi^{-1}(\hat{p}_{\mathbf{s},t}) \quad (6)$$

where  $\hat{p}_{\mathbf{s},t}$  is the estimation of the probability of wet days at location  $\mathbf{s}$  and day  $t$ .  $\hat{p}_{\mathbf{s},t}$  is estimated through a logistic regression using a probit model. As in the case of the marginal model, we take into account the seasonality and dependence on large-scale atmospheric conditions,

$$\hat{\mu}(\mathbf{s}, t) = \beta'_{\mu}(\mathbf{s})X_{\mu}(\mathbf{s}, t) \quad (7)$$

where  $X_{\mu}(\mathbf{s}, t)$  are the covariates and  $\beta_{\mu}(\mathbf{s})$  are regression parameters.

## 2.3 Covariance structure

The covariance function of the Gaussian field,  $\text{Cov}(Z(\mathbf{s}_1, t_1), Z(\mathbf{s}_2, t_2)) = C(\mathbf{s}_1, \mathbf{s}_2, t_1, t_2)$  is the component of the model that represents the dependence over space and time. We follow the classical paradigm in geostatistic by first estimate empirical covariance and then fit a parametric covariance function to it.

### 2.3.1 Estimation of empirical covariance

Estimation of the covariance function of the latent Gaussian field  $Z$  is challenging, as for dry locations the Gaussian variable is censored. We can only transform the precipitation data into Gaussian values for wet days,

$$Z^+ = \Phi_{\hat{\mu}}^{-1} \circ F_{emp}(Y)$$

where  $Z_i^+ = \max(Z_i, 0)$  and  $F_{emp}(\cdot)$  the empirical cdf. Note that the variance of  $\Phi_{\hat{\mu}}$  is unity and therefore the covariance of  $Z^+$  coincides with its correlation.

Some existing approaches for correlation estimation of a censored Gaussian variable can be found in the literature. Glasbey and Nevison [1997] used a method based on a likelihood maximization of the observed bivariate histogram of the censored latent variable, while Baxevani and Lennartsson [2015] used an alternative method of moments approach by inverting the theoretical expression for the mean of the censored cross product. Guillot and Lebel [1999] proposed a method of estimating the covariance function of the latent field which consists of computing the empirical covariance of raw data, then fitting a positive definite function to it, computing the inverse of this function through the Hermite polynomial expansion of the anamorphosis function, fitting a positive definite function to it, and then reversing it again with the use of the same Hermite polynomial expansion of the anamorphosis function to finally obtain an estimate of the desired covariance. To the knowledge of the authors, there is no study showing the superiority of one approach over others. In this study, we choose the approach based on likelihood maximization. The correlation at lag  $(k, l, \tau)$  denoted  $\hat{\rho}_{k,l,\tau}$  is obtained by maximization of  $\sum_{ijt} \log p(Z_{ijt}, Z_{i-k,j-l,t-\tau})$ . The pairwise probabilities,  $p$ , take one of three forms depending on whether, one or both of locations is dry:

$$p(Z_{ijt}, Z_{i-k,j-l,t-\tau}) = \begin{cases} \Phi_2(0, 0; \rho), & \text{if } Z_{ijt} = Z_{i-k,j-l,t-\tau} = 0 \\ \phi(0, 0; \rho) \Phi\left(\frac{z_0 - \rho Z_{ijt}}{\sqrt{1 - \rho^2}}\right), & \text{if only } Z_{i-k,j-l,t-\tau} = 0 \\ \phi_2(Z_{ijt}, Z_{i-k,j-l,t-\tau}; \rho), & \text{otherwise.} \end{cases} \quad (8)$$

where  $\phi$  and  $\Phi$  are the probability density function and the cumulative distribution function, respectively, for a standard univariate Gaussian distribution, and  $\phi_2$  and  $\Phi_2$  are the probability density function and the cumulative distribution function, respectively, for a standard bivariate Gaussian distribution (see Glasbey and Nevison [1997]).

### 2.3.2 Parametric covariance function

Our goal here is to find a parametric covariance function that fits well to the empirical covariance. We follow the classical assumption in geostatistics by assuming that the covariance function depends only on the distance between two locations  $\mathbf{h}$  and the time lag  $\tau$ .

$$\text{Cov}(Z(\mathbf{s}_1, t_1), Z(\mathbf{s}_2, t_2)) = C(\mathbf{s}_1, \mathbf{s}_2, t_1, t_2) = C(\mathbf{h}, \tau) \quad (9)$$

where  $\mathbf{h} = \mathbf{s}_2 - \mathbf{s}_1$  and where  $\tau = t_2 - t_1$ .

Several parametric functions specifying explicitly the joint space-time covariance structure have been fitted. For a collection of valid spatiotemporal covariances, see [Cressie and Huang, 1999, Ma, 2003, Stein, 2005] among others. The covariance structure that gave the best fit in terms of the Weighted Least Squares (WLS) criterion is the non-separable Gneiting covariance function Gneiting [2002],

$$C(\mathbf{h}, \tau) = \eta + \sigma^2 \left[ 1 + \left( \frac{\tau}{\theta_t} \right)^\lambda \right]^{-1} \exp \left[ - \frac{\left( \frac{\mathbf{h}}{\theta_s} \right)^\nu}{\left( 1 + \left( \frac{\tau}{\theta_t} \right)^\lambda \right)^{0.5\gamma\nu}} \right] \quad (10)$$

where  $\tau$  is the nugget,  $\sigma^2$  the sill,  $\theta_s$  the spatial scale parameter,  $\theta_t$  the temporal scale parameter,  $\nu$  the spatial power parameter ( $0 \leq \nu \leq 2$ ),  $\lambda$  the temporal power parameter ( $0 \leq \lambda \leq 2$ ) and  $\gamma$  the nonseparability parameter ( $0 \leq \gamma \leq 1$ ). Similar to the technique proposed by Cressie [1993] and Cressie and Huang [1999], we use a weighted-least-squares method to estimate the parameters  $\Theta = \{\tau, \sigma^2, \theta_s, \theta_t, \nu, \lambda, \gamma\}$  by numerical optimization of

$$\Theta = \underset{\Theta}{\text{argmin}} \sum_{\mathbf{h}} \sum_{\tau} \left( \frac{\hat{C}(\mathbf{h}, \tau) - C(\mathbf{h}, \tau|\Theta)}{1 - C(\mathbf{h}, \tau|\Theta)} \right)^2 \quad (11)$$

where  $\hat{C}(\mathbf{h}, \tau)$  is the empirical covariance function.

## 2.4 Smoothing anamorphosis and mean function model parameters in space

Direct estimates of the model parameters  $(\beta_\alpha, \beta_\gamma, \beta_\mu)$  for the anamorphosis (Gamma-regression) and mean function (logistic regression), denoted hereafter by  $\beta$ , are available only at locations with observed data, but in order to generate a complete precipitation field across the entire simulation domain ( $D$ ), we require a method that yields estimates of  $\beta$  between stations.

We use ordinary kriging [Cressie, 1993] to spatially interpolate  $\beta$  which are first estimated at the location of stations. The kriging predictor is the best linear unbiased predictor (in the sense of quadratic loss), and coincides with conditional expectation for normally distributed variables. We assume the geostatistical process  $\{\beta(\mathbf{s}) : \mathbf{s} \in D\}$  follows an intrinsically stationary process, i.e. having constant unknown mean  $m$ , known spatial covariance function  $\text{Cov}(\mathbf{h})$  for spatial lags  $\mathbf{h} = \mathbf{s}_i - \mathbf{s}_j$ , and can be written as

$$\beta(\mathbf{s}) = m + \epsilon(\mathbf{s}). \quad (12)$$

We typically want to predict values of the process at unobserved locations,  $\mathbf{s}_0 \in D$ . Kriging is a method that enables prediction of a spatial process based on a weighted average of the observations,

$$\hat{\beta}(\mathbf{s}_0) = \sum_{i=1}^n \omega_i \beta(\mathbf{s}_i). \quad (13)$$

The weights are found such that the predictor is unbiased,  $E[\hat{\beta}(\mathbf{s}_0) - \beta(\mathbf{s}_0)] = 0$  and by minimizing the variance of the interpolation error,  $\text{var}[\hat{\beta}(\mathbf{s}_0) - \beta(\mathbf{s}_0)]$ . Note that  $\hat{\beta}(\mathbf{s}_0)$  is the estimation of  $\beta$  at location  $\mathbf{s}_0$  and  $\beta(\mathbf{s}_0)$  the true value. We are dealing here with a minimization problem with a constraint, which can be solved using Lagrange multipliers. The solution is given by

$$\begin{cases} \sum_{i=1}^n \omega_i = 1 \\ \sum_{j=1}^n \omega_j \text{Cov}[\beta(\mathbf{s}_i), \beta(\mathbf{s}_j)] + \lambda = \text{Cov}[\beta(\mathbf{s}_0), \beta(\mathbf{s}_i)], \forall i = 1 \dots n \end{cases} \quad (14)$$

where  $\lambda$  is the Lagrange multiplier.

It is possible to include in the kriging spatial covariates (latitude and longitude) and elevation. More preferably, the two steps in the estimation procedure (estimation of  $\beta$  at location of stations + spatial interpolation) could probably be combined in one by building a composite likelihood or by including in the estimation of  $\beta$  spatial variations, modelled with Legendre polynomials for longitude, latitude and altitude. These options are not explored at this stage of the model development and will be investigated in future work.

### 3 Test case: Wupper catchment

#### 3.1 Data

We consider a network of 22 rainfall stations, located in and around the Wupper catchment, as shown in Fig. 2. The data consist of accumulated daily rainfall from 1980 to 2015. Between stations, 12 contain missing values. However, for those stations the proportion of missing values is less than 1% for half and less than 5% for the other half.

In addition to the station data, we also use as covariates climate variables from ERA-Interim reanalysis (grid  $0.5 \times 0.5$  degree). Climate variables are CAPE, wind, temperature, dewpoint temperature, mean sea level pressure, and geopotential, specific humidity, relative humidity and vertical velocity at 500 *hPa*.

#### 3.2 Estimation of the model parameters

##### 3.2.1 Marginal model for precipitation amounts

We use a gamma distribution to model the precipitation amount at each station. For the shape and scale parameter of the gamma distribution, the seasonality and the dependence on large-scale atmospheric conditions is taken into account as specified in (3) and

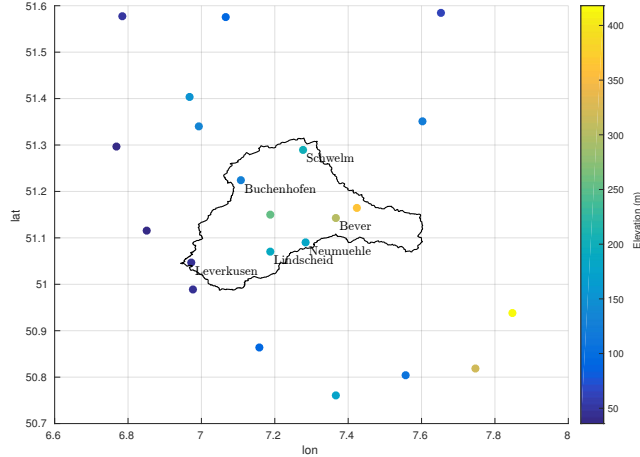


Figure 2: Locations of the gauge stations with color encoding the altitude.

(4). Covariates that lead to the "best" model according to Akaike information criterion (AIC) are sine and cosine waves (for temporal seasonality), CAPE, wind, temperature, dewpoint temperature, mean sea level pressure, potential vorticity.

### 3.2.2 Mean function to model precipitation occurrence

To estimate the mean function of the Gaussian field, we first estimate the probability that a day is wet through logistic regression and then use (6). The seasonality and the dependence on large-scale atmospheric conditions is also taken into account in the logistic regression. Covariates that lead to the best model according to AIC criterion are sine and cosine waves, CAPE, wind, temperature, dewpoint temperature, geopotential, specific humidity, relative humidity, vertical velocity.

### 3.2.3 Smoothing parameters over the simulation domain

We use ordinary kriging to spatially interpolate the model parameters  $\beta$  over space as described in section 2.4. Figure 3 shows, as an example, the result of the spatial interpolation of  $\beta$  associated to the wind in the amount model. In the left panel, the estimates of  $\beta$  at the location of stations is showed and in the right panel the smoothing of  $\beta$  over the simulation domain.

### 3.2.4 Covariance function

First, we estimate an empirical covariance function using the likelihood approach described in section 2.3.1 for each season separately. We then fit a Gneiting covariance function (10) to the empirical covariance. To illustrate the fitting of the spatio-temporal covariance function, Fig. 4 shows for winter (DJF) the fitted spatial covariance function

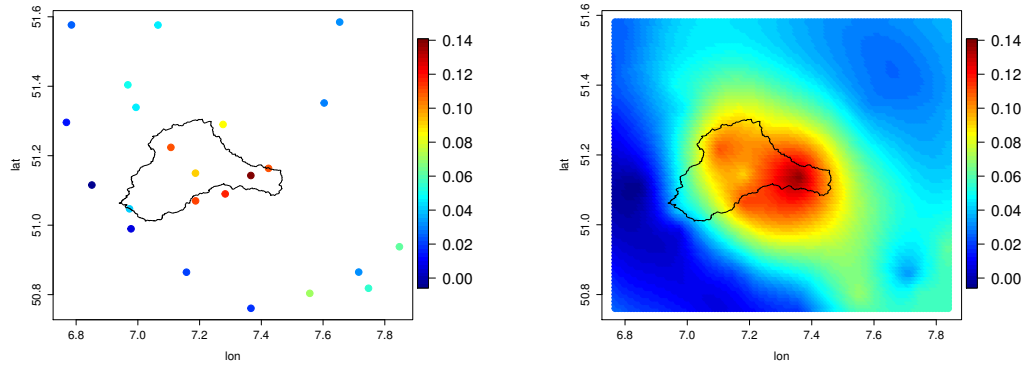


Figure 3: Parameter  $\beta$  associated with the covariate wind in the amount model: estimation at the location of stations (left); smoothing  $\beta$  over space (right).

$C(\mathbf{h}, \tau = 0)$  in the left panel and the fitted temporal covariance function  $C(\mathbf{h} = 0, \tau)$  in the right panel. There is good agreement between the fitted covariance function (red line) and the empirical covariance function (black dot).

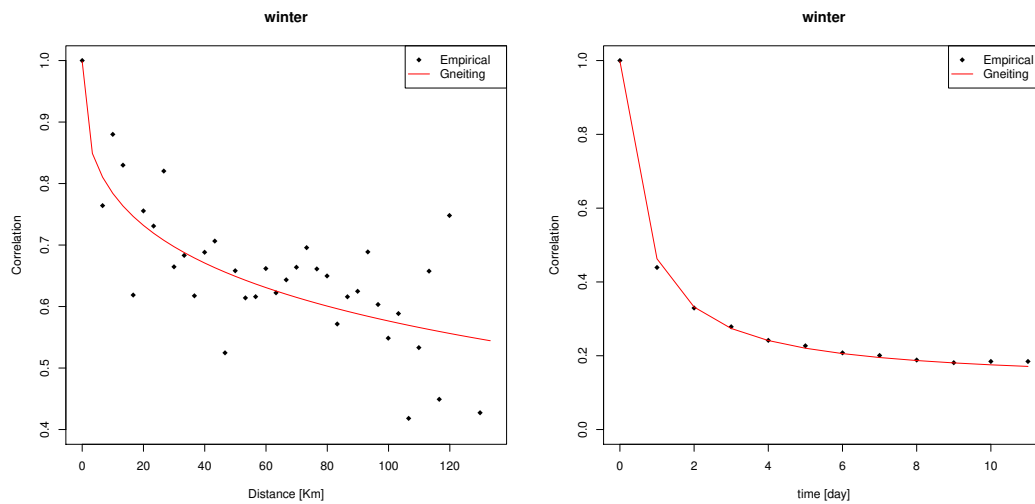


Figure 4: Spatial covariance function (left) and temporal covariance function (right)

### 3.3 Simulations

To simulate precipitation data at site  $\mathbf{s}$  on day  $t$ , the following steps are performed

1. A realization of a Gaussian random field  $Z(\mathbf{s}, t)$  with mean function  $\hat{\mu}$  and covariance  $\hat{C}$  is generated.

2. For every location and day, there is zero precipitation if  $Z(\mathbf{s}, t) \leq 0$ .
3. For location and day with positive precipitation, the simulated intensity is set to  $Y(\mathbf{s}, t) = \psi(Z(\mathbf{s}, t))$ .

### 3.4 Do the simulations reproduce characteristics of the data?

We check in this section the performance of the model in reproducing the temporal and the spatial variability of rainfall fields. For this, we generate 100 times series of precipitation for the period covered by observations (1980-2015) and then examine the results.

#### 3.4.1 Temporal model

**Wet day frequency** We estimate the frequency of wet days from observations and then compare it to the ones obtained from simulations. The result is shown in Fig. 5. The model replicates well the occurrence of the wet days.

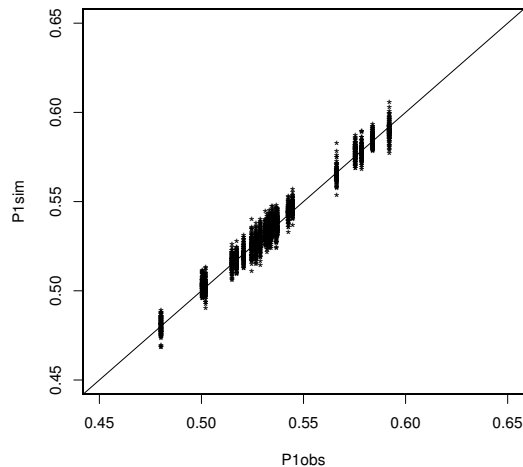


Figure 5: Frequency of wet days from simulations (ordinate) against observations (abscissa).

**Dry and wet spells length** An essential feature that stochastic precipitation generators attempt to replicate is the length of dry and wet spells. Figures 6 and 7 show respectively the log frequency of dry spells and the log frequency of wet spells at six stations within the Wupper watershed, with the mean and 90% pointwise confidence intervals based on the 100 simulations overlaid. The stochastic generator shows good characterization of both dry spell and wet spell behaviour at individual locations.

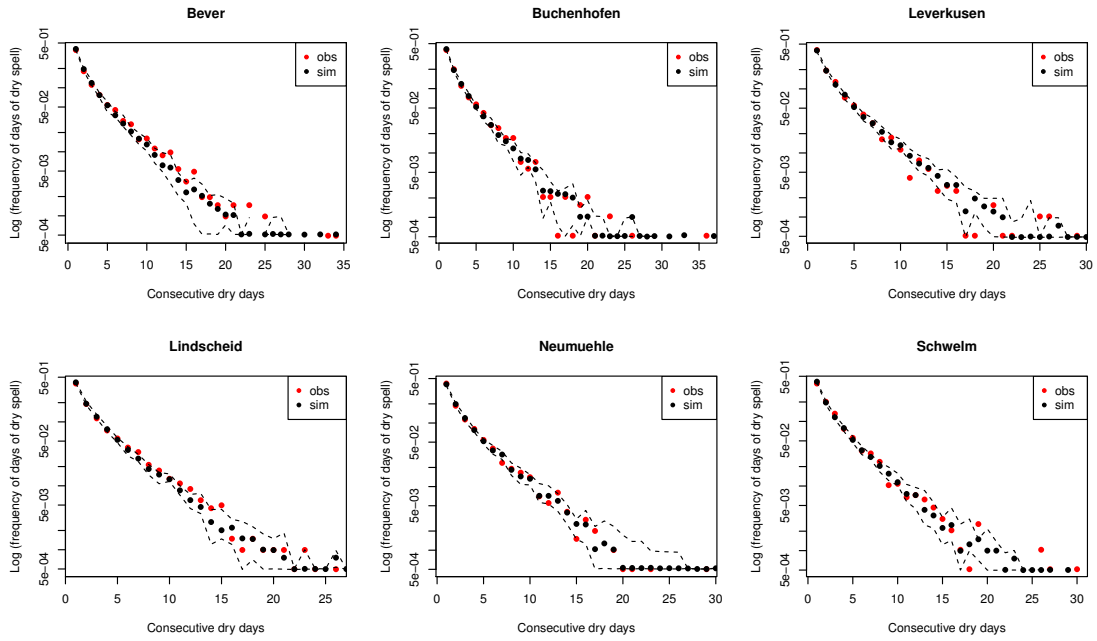


Figure 6: Dry spell frequency from observations (red dot) and simulations (black dot, mean with a 90% confidence interval superimposed).

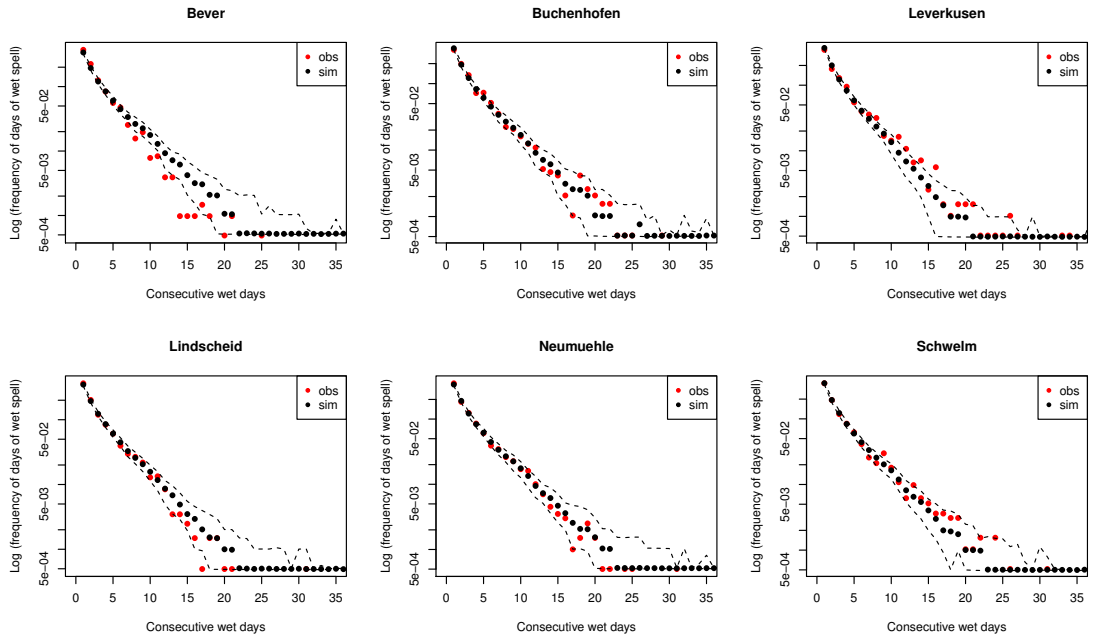


Figure 7: Same as Fig. 6 but for wet spells.

**Seasonal cycle** We turn next to the ability of the stochastic generator to reproduce the seasonal cycle of precipitation at individual station. In Figure 8 the monthly mean of precipitation obtained from observations and those obtained from simulations are shown for six stations. The model reproduces very well the seasonal cycle of precipitation. The root mean square errors (RMSE) between the observation’s monthly mean and the mean of simulation’s monthly mean are very low (less than 0.15 mm) and the correlations between them are high (greater than 0.9).

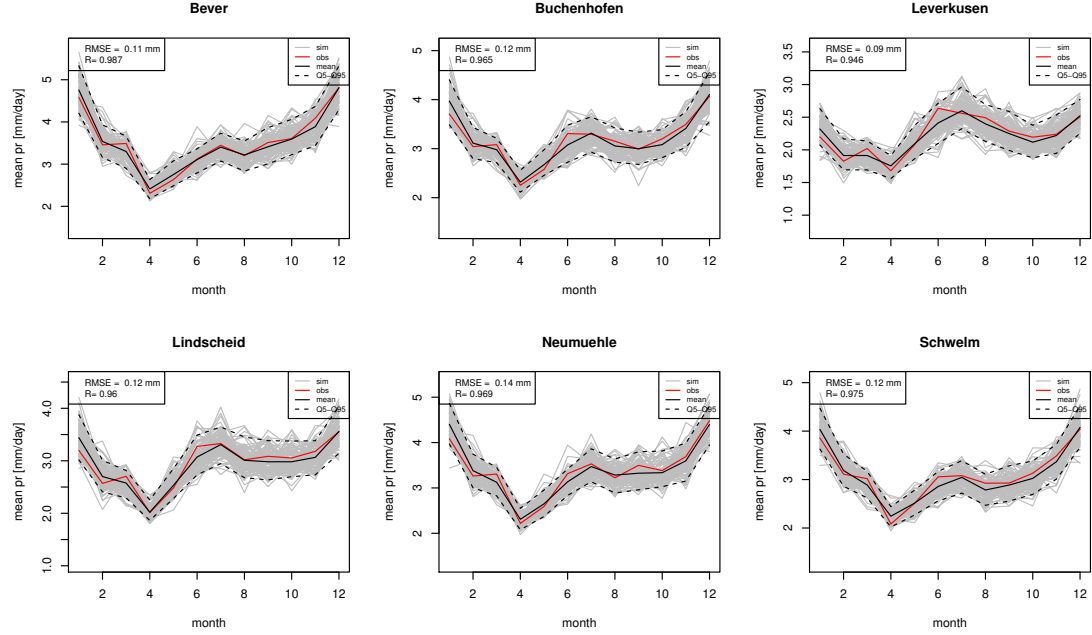


Figure 8: Monthly mean of precipitation: observation (red line), 100 simulations (grey line), mean of the 100 simulations (black line) and 90% confidence interval (dashed black line).

**Interannual variability** We also examine the interannual variability. Figure 9 shows the annual sum of precipitation for six stations. Even if there are some small mismatches for some years, the observed annual precipitation sum (red line) is found in the confidence interval (dashed black line) for all stations. The model is thus able to well represent the interannual feature of the precipitation and thus suited as a conditional weather generator for downscaling decadal prediction experiments. Potentially, using indices related to El Niño-Southern Oscillation (ENSO) or the North Atlantic Oscillation (NAO) as covariates in the model could further improve the interannual variability in the simulations.

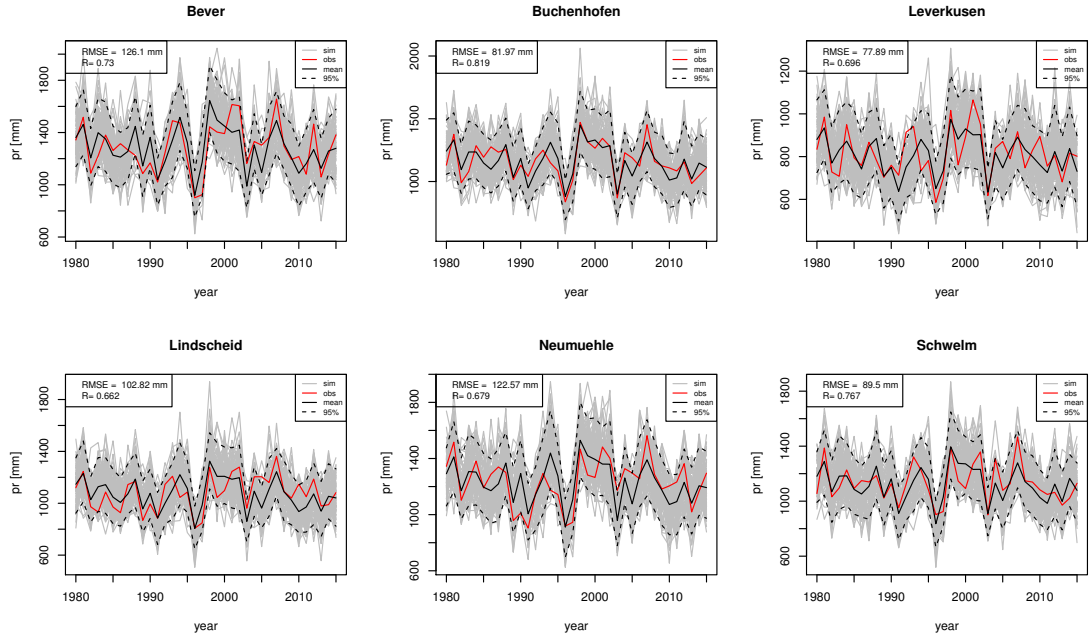


Figure 9: Annual precipitation sum: observations (red line), 100 simulations (grey line), mean of the 100 simulations (black line) and 90% confidence interval (dashed black line).

### 3.4.2 Spatial model

The stochastic precipitation generator satisfactorily simulates the occurrence behavior and intensity of rainfall at individual locations. Now, we examine the spatial features of the stochastic generator.

**Spatial correlation** Figure 3.4.2 shows, for wet days, the spatial correlation of rainfall for observations (red line) and the one for simulations (grey lines) and 90% pointwise confidence intervals based on the 100 simulations overlaid for each season. Excepted summer (JJA), there is a little underestimation of the correlation. It could be attributed to the maximum likelihood approach used to estimate the empirical correlation of the Gaussian field. This underestimation may slightly reduces the variability in the simulated precipitation field.

**Joint occurrence probability** A positive feature of the defined stochastic model is in correlating wet and dry occurrences across space. In order to illustrate the fit of the spatial dependence, we consider all pairs of stations and then compute, for observations and simulations, the probability of simultaneously wet  $P[Y(\mathbf{s}_i, t) > 0, Y(\mathbf{s}_j, t) > 0]$ , simultaneously dry  $P[Y(\mathbf{s}_i, t) = 0, Y(\mathbf{s}_j, t) = 0]$  and the probability that one site is wet and the other dry  $P[Y(\mathbf{s}_i, t) > 0, Y(\mathbf{s}_j, t) = 0]$ . The results are assembled in Figure 11. The stochastic generator seems to replicate quite well the joint probability of precipita-

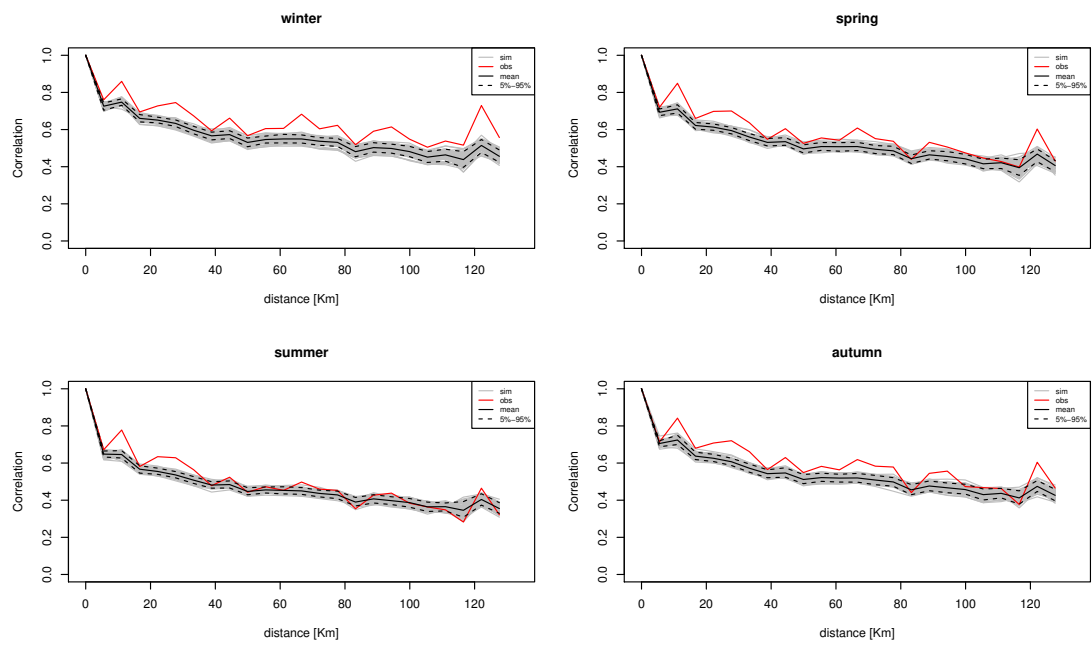


Figure 10: Spatial correlation of rainfall: observations (red line), 100 simulations (grey line), mean of the 100 simulations (black line) and 90% confidence interval (dashed black line)

tion occurrence state with correlations between the observed joint probabilities and the simulated ones greater than 0.9. In general, we feel that the model replicates spatial aspect of precipitation processes well.

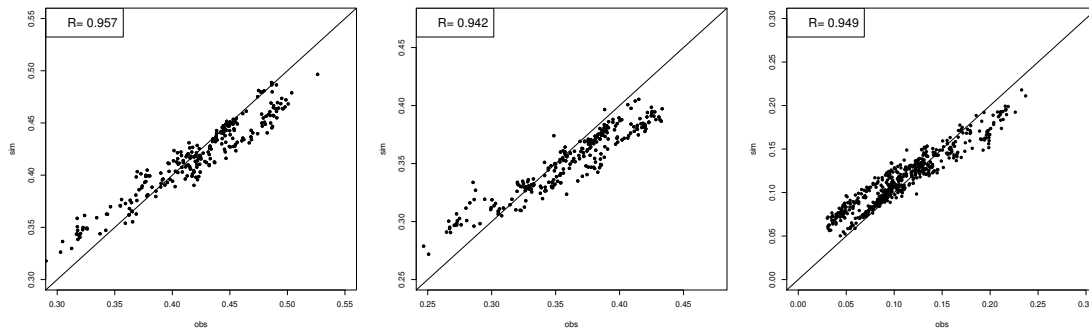


Figure 11: Joint probability of precipitation occurrence state of two sites: simultaneously wet (left), simultaneously dry (middle), one site wet and the other dry (right).

### 3.5 Out-of-sample validation

In this section we examine the ability of the stochastic generator to reproduce the precipitation behavior at locations with no available data. For this reason, we exclude two stations (Buchenhofen and Neumuehle) from the data. We re-estimate the model parameters without these two stations. We then simulate 40 time series from 1980 to 2015 and check how the model performs at these two stations.

We start by examining the frequency of wet day. Table 1 shows the wet-day frequency of observations and a mean wet-day frequency over from 40 simulations together with their 5% and 95% quantile at Buchenhofen and Neumuehle. For both stations, the frequency of wet days is well in between  $Q5\%$  and  $Q95\%$  and very close to the mean of the wet day frequency estimated with the 40 simulations. The seasonal cycle and the interannual variability are also well replicated by the model, as displayed in Figs. 12 and 13, respectively. Overall, we feel that our stochastic generator captures the important aspects of precipitation data when interpolated to locations with no available information.

Table 1: Wet-day frequency at Buchenhofen and Neumuehle in an out-of-sample experiment.

	Observation	Simulation		
	-	$Q5\%$	mean	$Q95\%$
Buchenhofen	0.5448	0.5437	0.5499	0.5567
Neumuehle	0.5428	0.5398	0.5437	0.5488

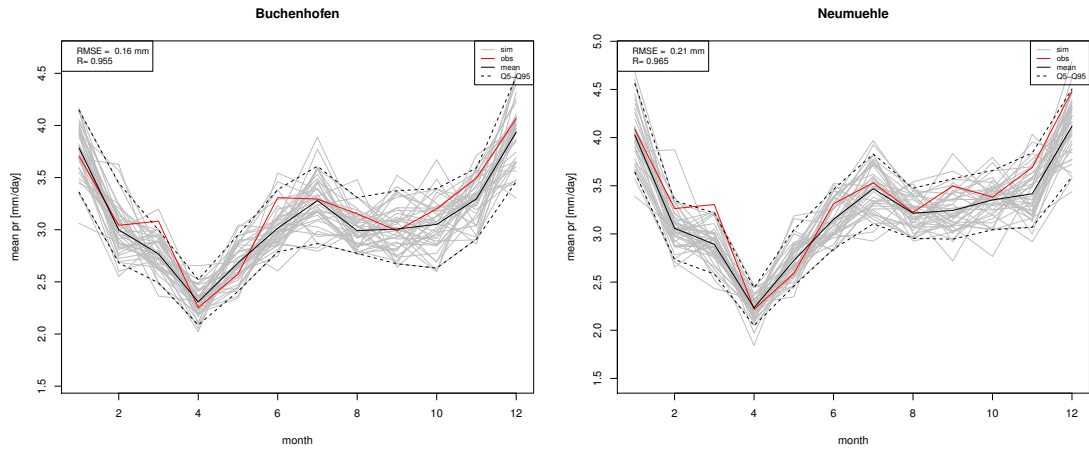


Figure 12: Monthly mean of precipitation: observations (red line), 40 simulations (grey line), mean of the 40 simulations (black line) and 90% confidence interval (dashed black line).

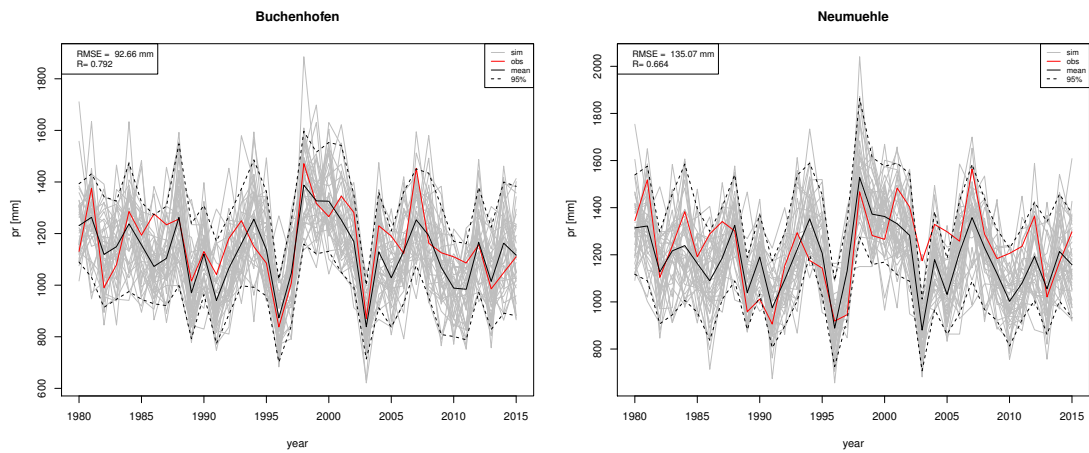


Figure 13: Yearly precipitation sum: observations (red line), 40 simulations (grey line), mean of the 40 simulations (black line) and 90% confidence interval (dashed black line).

## 4 Conclusions

We developed a stochastic generator which is able to simulate spatially correlated fields of daily precipitation. The generator is based on a latent Gaussian random field that drives both the rainfall occurrence and the rainfall intensity processes. In our model, the precipitation amount was modelled by a gamma distribution, with the scale and shape parameters being linear models of seasonal and atmospheric covariates. The mean function of the Gaussian field was estimated by inverting the probability of wet days at each locations. The parameters of both, the amounts model and mean function were interpolated over space, allowing the generation of a complete precipitation field across the entire simulation domain. The model was used to simulate spatially a correlated field across the Wupper catchment covering the time period from 1980 to 2015. We have shown that the stochastic generator is able to reproduce not only the occurrence and the amount of rainfall at individual stations, but also the dependence between different stations quite well. Additionally, out-of-sample experiment shows that the model is well be able to reproduce the precipitation behavior at locations with no available data.

The amounts model and the mean function parameters are first estimated at station locations and then interpolated using kriging. The possibility to combine these two steps into one, for instance, by using spatial covariates directly in the estimation, will potentially be investigated in the future. Other approaches to estimate empirical correlation of a censored Gaussian field should also be explored. Additionally, non-isotropic spatial covariance structures might be a further interesting field to be considered. Finally, a mixture of gamma and GPD distributions might further improve the daily precipitation extremes

In summary, the generator developed is well suitable to simulate spatially and temporally consistent precipitation fields for BINGO. With its ability to reproduce the inter-annual summary statistics it has proven adequate for the decadal prediction experiments used.

## References

- Pierre Ailliot, Craig Thompson, and Peter Thomson. Space–time modelling of precipitation by using a hidden markov model and censored gaussian distributions. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 58(3):405–426, 2009.
- Pierre Ailliot, Denis Allard, Valérie Monbet, and Philippe Naveau. Stochastic weather generators: an overview of weather type models. *Journal de la Société Française de Statistique*, 156(1):101–113, 2015.
- Denis Allard and Marc Bourotte. Disaggregating daily precipitations into hourly values with a transformed censored latent gaussian process. *Stochastic Environmental Research and Risk Assessment*, 29(2):453–462, 2015.
- David J Allcroft and Chris A Glasbey. A latent gaussian markov random-field model for

- spatiotemporal rainfall disaggregation. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 52(4):487–498, 2003.
- Andras Bardossy and Erich J Plate. Space-time model for daily rainfall using atmospheric circulation patterns. *Water Resources Research*, 28(5):1247–1259, 1992.
- Anastassia Baxevani and Jan Lennartsson. A spatiotemporal precipitation generator based on a censored latent gaussian field. *Water Resources Research*, 51(6):4338–4358, 2015.
- George EP Box and David R Cox. An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 211–252, 1964.
- Richard E Chandler, Valerie Isham, Enrica Bellone, Chi Yang, and Paul Northrop. Space-time modelling of rainfall for continuous simulation. *MONOGRAPHS ON STATISTICS AND APPLIED PROBABILITY*, 107:177, 2006.
- Noel Cressie and Hsin-Cheng Huang. Classes of nonseparable, spatio-temporal stationary covariance functions. *Journal of the American Statistical Association*, 94(448):1330–1339, 1999.
- Noel AC Cressie. *Statistics for spatial data: Wiley series in probability and mathematical statistics. Find this article online*, 1993.
- CA Glasbey and IM Nevison. Rainfall modelling using a latent gaussian variable. In *Modelling Longitudinal and Spatially Correlated Data*, pages 233–242. Springer, 1997.
- Tilmann Gneiting. Nonseparable, stationary covariance functions for space–time data. *Journal of the American Statistical Association*, 97(458):590–600, 2002.
- Gilles Guillot and T Lebel. Approximation of sahelian rainfall fields with meta-gaussian random functions. *Stochastic Environmental Research and Risk Assessment*, 13(1-2): 113–130, 1999.
- William Kleiber, Richard W Katz, and Balaji Rajagopalan. Daily spatiotemporal precipitation simulation using latent and transformed gaussian processes. *Water Resources Research*, 48(1), 2012.
- Jan Lennartsson, Anastassia Baxevani, and Deliang Chen. Modelling precipitation in sweden using multiple step markov chains and a composite model. *Journal of Hydrology*, 363(1):42–59, 2008.
- Chunsheng Ma. Families of spatio-temporal stationary covariance models. *Journal of Statistical Planning and Inference*, 116(2):489–501, 2003.
- R Mehrotra, R Srikanthan, and Ashish Sharma. A comparison of three stochastic multi-site precipitation occurrence generators. *Journal of Hydrology*, 331(1):280–292, 2006.

- C Onof, RE Chandler, A Kakou, P Northrop, HS Wheeler, and V Isham. Rainfall modelling using poisson-cluster processes: a review of developments. *Stochastic Environmental Research and Risk Assessment*, 14(6):384–411, 2000.
- Francesco Serinaldi and Chris G Kilsby. Simulating daily rainfall fields over large areas for collective risk estimation. *Journal of Hydrology*, 512:285–302, 2014.
- R Srikanthan and TA McMahon. Stochastic generation of annual, monthly and daily climate data: A review. *Hydrology and Earth System Sciences Discussions*, 5(4): 653–670, 2001.
- Michael L Stein. Space–time covariance functions. *Journal of the American Statistical Association*, 100(469):310–321, 2005.
- Daniel S Wilks. Use of stochastic weathergenerators for precipitation downscaling. *Wiley Interdisciplinary Reviews: Climate Change*, 1(6):898–907, 2010.
- Daniel S Wilks and Robert L Wilby. The weather generation game: a review of stochastic weather models. *Progress in physical geography*, 23(3):329–357, 1999.
- DS Wilks. Multisite generalization of a daily stochastic precipitation generation model. *Journal of Hydrology*, 210(1):178–191, 1998.
- T. W. Yee. *Vector Generalized Linear and Additive Models: With an Implementation in R*. New York, USA: Springer, 2015.