Dear Prof. Dr. Axel Bronstert

We have revised the manuscript according to your comments and the comments from both reviewers.

Concerning the questions raised by Reviewer 1, we have already addresses these issues in the first revision round, but it appears that we failed to report this thoroughly in the manuscript and thereby created confusion. We hope that the clarifications now made solve these issues, please see our detailed response to Reviewer #1 for explanations.

As suggested by yourself and Reviewer 2 we have added an electronic appendix, which supplements the methodology and results sections and gives a more in-depth description of how to work with the weather generator and how to report tables with parameter estimates for both present and future conditions and calculate change factors. Note that the Figure in the appendix (Figure S1) is taken directly from Burton et al. (2010a) and that we have not yet asked for permission to use it. This is the best figure to describe the current methodology, and we will be happy to write the author/publisher for permission once our manuscript has been accepted.

With respect to the other questions raised by reviewer 2, we have addressed them all mostly by making modifications in the manuscript – and refer to the detailed response to Reviewer 2 for explanations.

We would like to express our sincere thanks to the reviewers for their comments and identification of areas in the manuscript which needed clarification. We furthermore thank you for allowing us to resubmit a revised version of the manuscript and look forward to hear from you again.

All the best,

Hjalte Jomo Danielsen Sørup

Reply to reviewer 1

We would like to thank the reviewer for re-reviewing the manuscript.

In the re-revision three questions are raised relating to the hourly RCM data used and we address them here and in the manuscript and hope the issues are clarified accordingly:

1) What exactly means "out of a research project" and how were the data thus obtained validated?

I believe this question refers to the sentence on page 5 lines 29-30. In that case the correct citation is "*The simulations were carried out as part of the research project RiskChange (www.riskchange.dhigroup.com)*", which was primarily meant as a means of referencing the overall project in which the simulations had taken place (and where no proper paper reference exist). Two of the co-authors are officially affiliated with the mentioned project which is why we have access to data and the validation of data is done for the ERA-driven runs in Mayer et al. (2015) (as already mentioned by the reviewer). The sentence has been replaced with a reference to a technical report (Fox Maule et al., 2014) stating exactly which model runs that have been run.

2) Are the 2 in-house simulations for present climate also driven by ERA interim?

The high resolution RCM runs used in the present study are all forced by GCMs (EC-EARTH and NorESM) as mentioned on page 5 lines 25-28 and table 2. The RCM runs for present conditions are run using the same GCMs as forcing. We have further highlighted this information in the revised manuscript. Further, reference to the background report, Fox Maule et al. (2014), has been added to enable full access to the technical details of the RCM runs used.

3) For any simulation of the future, say by using GCM G and RCM R, are the corresponding simulations of the present, also done by G and R, available?

Yes, and that is what has been used in all cases (for both high resolution RCMs and ENSEMBLES RCMs). We have clarified this in the manuscript as mentioned above.

Reference

Fox Maule, C., Mayer, S., Sobolowski, S. and Christensen, O.B. (2014) Background information on the RiskChange simulations by BCCR and DMI. Danish Climate Centre Report 14-05. The Danish Meteorological Institute, Copenhagen, Denmark. Available at: http://www.dmi.dk/fileadmin/user_upload/Rapporter/DKC/2014/RiskChange_simulations_report.pdf.

Reply to reviewer 2

We would first of all like to thank the reviewer for the thorough review of the manuscript. In the following we are answering the review section by section.

1 General Comments

In their manuscript "Downscaling future precipitation extremes to urban hydrology scales using a spatio-temporal Neymann-Scott weather generator" H.J.D. Sorup, O.B. Christensen, K. Arnbjerg-Nielsen and P.S. Mikkelsen use a spatio-temporal precipitation generator of the Neymann-Scott-type to to study climate change signals in high resolution precipitation fields. They consider also extreme precipitation events. These types of precipitation generators are very useful for the hydrology community using the output of these models as input to hydrological models. Particular for those settings where a high spatio-temporal resolution is important (e.g. urban drainage or small catchments), there is not yet a feasible alternative to precipitation generators. Running RCMs at this resolution is numerically very costly and suffer from problems as biases. I thus consider the topic of this manuscript as very important.

The manuscript is relatively well organised and some parts are also well written. However, even as somebody who is also working with spatio-temporal Poisson-cluster models (as the Neymann-Scott), I could not follow at several stages of the document and would not be able to reproduce the results. It is obvious that the authors have already thought a lot on balancing the level of details with the readability. While it is working well in some parts, it does not in others. One solution to this would be to split the manuscript into two: one desribing the setting up of the WG for the chosen region, including parameter estimation and model validation for the present climate (only observations); the other manuscript would deal with the climate change study and has then more room to discuss the results found. Another solution would be to have an extensive appendix with all the details or suplementary material (not a Video, rather clear describtions of the approaches with equations and more detailed validation results). At the current stage, the manuscript does not seem to be convincing enough as a) methods being somewhat too simple have been used (e.g. Eq. 3), b) a too small ensemble to estimate the uncertainty (ensemble size 10) and c) the description is not clear enough to be reproducible.

I recomment publishing this work but with this manuscript undergoing a major revision. A suggestion would be to split the manuscrip or to transfer the model building (parameter estimation and validation) in large parts into an appendix to allow for more details without negative effects on readability.

The reviewer notes the high relevance of the study and highlights three general comments in the introduction that support his/hers overall recommendation of major revisions:

a) methods used is being somewhat too simple
b) a too small ensemble of weather generator runs to estimate the uncertainty
c) the description is not clear enough to be reproducible

His/hers recommendation is to add details to the paper and consider splitting the paper in two. At the same time the reviewer recognizes the difficulty in balancing accuracy, readability, and length of paper. The recommendation of splitting the paper into two is not supported by the editor.

Our study does not contribute to weather generator development in general, but on applying a weather generator at a different scale than what has previously been reported in literature.

We have tried to improve clarity and readability and supplement the description on how we have used the weather generator by highlighting citations to existing literature to guide the reader further. Further, we have revised the manuscript based on the specific comments and added supplementary material to support the understanding on how the weather generator was used and on the parameter estimation. We hope that the reviewer will agree that it has helped making the paper more clear and concise.

2 Specific Comments

The authors mention very frequently that they generate "realistic" rainfall. This leaves the impression that they want to convince by repeating the word "realistic". I'd further-more not use this in a scientific article as it does not have a well defined meaning.

I'd miss a sketch (either verbal or as a figure) describing how the study is carried out and what assumptions are behind that. The RCMs show changes in summray statistics obtained from a grid of a few kilometers. The WG actually works in continuous space and the reseult is discretized later (which is actually never talked about). Does it make sense to simply scale the summary statistics ontained on the coarse grid and aplly the same scaling to the statistics from gauges (if this is what is done, which I suppose but I am not sure).

We will rephrase the relevant sentences. "Realistic" was introduced to the manuscript based on the comments from an earlier reviewer to highlight that we do not (necessarily) expect an excellent fit at all points, but rather a rainfall field that has the same properties over the full model domain as the observations.

An overview of how the work with the weather generator is carried out is given in Burton et al. (2008) and a proper reference is put in the paper to guide the reader to this. As we have not developed the weather generator we prefer to guide the reader to the original literature instead of repeating it.

It is true that the WG works in continuous space and that it is not discussed in this manuscript. This topic is covered well in the literature (Burton et al., 2008; 2010a) and was originally left out for that reason. The same is the case for the suitability of using change factors from the RCMs for perturbation of the WG which is covered in Burton et al. (2010b). For both topics, they are mentioned in the supplement with relevant references.

2.1 Abstract

"used to perturbe the WG" is a strange description of what (I think) is actually done. WG parameters are estimated from summary statistics scaled by change factors.

It is true that the individual summary statistics are scaled by change factors but the weather generator, as such, appear perturbed by the change factors as the full response of all the change factors is not linear.

2.2 Introduction

p2 l32 what is "one hour or higher"? Is that two hours or a half?

The sentence "one hour or higher" has been changed for something more clear in the manuscript.

2.3 Data and weather generator

2.3.1 Data representing present conditions

p5 l10ff I don't think the unexperienced reader can follow why you introduce a third set of data which are ten realisations of the WG. What is meant by "refitting and rerunning"? Why does it corroborate the results? I think a different motivation for this is needed.

The introduction of the ten realizations of the weather generator and the "re-fitting" of the model to these data was done as a way of evaluating the fit of the original model. This approach was proposed by an earlier reviewer as a way of

1. testing how representative the individual model runs are, which is also closely linked to the other aspect which is
2. giving some sort of rough estimate of the uncertainty of the weather generator output

By "re-fitting" the model to these data sets we can also get an idea about which of the fitting parameters are well preserved and which are less well preserved. Based on our findings we doubt that adding more simulations will give higher certainty in our interpretation of parameter identifiability.

2.3.2 Weather generator

In this chapter, a figure sketching the idea of the WG would be good. The reader not familiar with the NSRP will not get an idea. You need to give the model parameters as they are needed later. Why not using a figure to explain them?

An excellent figure explaining the model in detail is given in both Burton et al. (2008) and Burton et al. (2010a) and as we haven't changed the actual model structure we prefer to refer to one of these. As such we have added the figure from Burton et al. (2010a) to the supplement in the part that detailed explain how to operate the weather generator. The model parameters are (already) listed and discussed in this paragraph and by splitting the Data and Weather Generator Sections (as proposed in the next comment) the readability has increased.

2.4 Methodology

2.4.1 Fitting of the weather generator

For me it is strange to see that the parameter estimation is not just in another subsection but in a completely new section. Why not havig a section on data, a section on the WG (as it is central here) with subsection on describing the model and another subsection on describing the parameter estimation. Then have a section on "Methods" and describe the way you validate the model and obtain change factors
"Fitting of the weather generator" is not a nice title. Why not using "Parameter estimation"?
I find it strange that you don't specify the objective function minimized to obtain estimates but you specify the name of the minimization algorithm. Why is that important? I guess it somehow avoids local minima but not completely otherwise you would not have to run it three times ("thrice" is not so frequently used).
I suggest to show the objective function as the construction has probably a larger influence on the result than the minimization algorithm.
the auto correlation function is estimated by the lag-1 and then an exponential decay is used, thus you have a parametric form for it (I supppose). What about the cross correlation function? Looking at Fig.9, I think this is the same: an exponential decay and the rate is estimated. If this is the case, you could also draw lines in Fig. 9 for the WGs instead of re-estimating it.
p9 l3 if you introduce the weighing scheme, you could also show the objective function such that one can see where the weights go in practice.

The structure of the paper has been changed as proposed and the new structure is: 2 Data, 3 Weather Generator and 4 Methodology, with the present sections 2.3, 3.1 and 3.3 gone into the new section 3.

"Fitting" is the term generally used by Burton et al. and Cowpertwait et al. but we agree that "parameter estimation" is a proper term that better describes what actually takes place. Accordingly, the relevant parts of the manuscript have been re-phrased.

The optimization function was something that a reviewer specifically asked for earlier on, but it is (as pointed out by the reviewer) really too specific information for a paper like this, as this is well documented in Burton et al. (2008a). The same is the case for the objective function (straightforward, well documented). About the three reruns of the algorithm it is not us who rerun it three times; it is the software RainSim that by default does this. The whole section describing the weather generator has been re-phrased and re-structured with thorough citations to the specific objective function used as well as the other specifics described in the background papers (Burton et al., 2008; 2010a; 2010b).

In Figure 9 there are already drawn exponential decay lines but the supporting text has been altered and clarified to further guide the reader.

The objective function, model parameters and the weighing scheme are all interlinked and the restructuring of the manuscript as described already brings the relevant sections together and makes the link clearer.

2.4.2 Evaluation of simulated time series

p8 l24ff I thought the idea of generating an enesemble is to obtain an estimate for the uncertainty of parameters, statistics, simulations. I was confused by the sentence "to evaluate if the realisation is representative". My idea of a WG is that it is relatively easy to generate lots of realisations to propagate uncertainty to hydrological models. I would not be looking for "one" representative realisation.
p8 l29ff I wonder why this procedure takes so long. A central advantage of WGs is that it is easy to generate simulations. Refitting of the simulations could be sped up by giving the parameter values used for simulation. The algorithm is than already in the local minimum you are looking for.
Eq.1 is not clear to me. Are the Y_{WGS} sampled over space? The notation needs to be clearer.

The idea of ensembles is generally to quantify uncertainty and identifiability. The data sets generated by the weather generator are very large. Our focus has been on evaluating whether the features we test can be considered generic for all realizations we could do. Hence, the extra information we could extract by generating more realizations is very limited as we want to look at properties of a single data set; this is what likely will be used by someone who would like to use these data in hydrological modelling (but, Yes, we could give better estimates of the uncertainty in some aspects).

The procedure is cumbersome for practical reasons: because the output is written to a lot of text files and because we have a lot of grid points. This also limits the length of the time series as the total data set cannot take up more space than 2 GB before the program will crash because of limitations of the RainSim software package that we have employed. By not starting the re-fitting from the original values we secure that the optimization routine is robust; and the time invested in this process is well spent compared to the increased trust in robustness we get from it.

The weather generator is sampled over space and the notation in the equations has been be clarified.

2.4.3 Perturbation of the weather generator with climate change signals

p9 l8 I suggest to introduce the notation of the statistics $Y_{i;j;k}$ on page 7

The notation has been adopted.

2.4.4 Evaluation of Extremes

Unfortunately, the structure of this section does not reveal itself to me.

p9 l23 “spatial correlation distance” needs to be defined

p9 l24 “simulated data set” ... “better than RCM” Also RCM data is “simulated”.

p10 l1 the description of the concept used does not become clear to me. You talk about “maximum average intensities” which sounds like a block maxima approach. Shortly later you a POT approach is mentioned.

Extreme event statistics

I do not see why you are first estimating return periods by plotting positions and later by a POT approach?

Seasonality of extreme events

Here you do not consider magnitudes, only the number of events over a threshold. This is a weaker criterion than the magnitudes but I would not expect the WG to reproduce the magnitudes. This is OK but need to be made transparent.

You might cite Wilks [2011] for the $\chi 2$-test.

Unconditional spatial correlation extremes

I do not understand the concept “unconditional” here. What would be conditional?

I do not understand the notation in Eq. 8. The expectation values E{ZA | U} and E{ZB | U} are constant values (as all expectation values) and thus I do not see how a covariance is obtained between them

The section on extremes evaluation has been reformulated to avoid ambiguous use of words.

The interplay between extraction of extremes, “maximum average intensities” and the POT approach has been clarified to secure that it is clear that the return period of the extremes are assigned at event level, the intensities are found using “maximum average intensities” and that only the extremes above a certain threshold (POT) is used in the evaluation of extremes.

Depending on how you construct your GDP model you need to have an idea about how long your time series are and either how frequent your threshold is exceeded or what your threshold should be to get a certain exceedance. The POT approach used here is simply a way of selecting the events to which the GDP model is fitted.

We only look at the number of extremes for seasonality, but it is the same events extracted using the POT approach and the magnitude is as such already covered in the GDP assessment. You can do a similar analysis on monthly or seasonal basis, but the data basis would be very low and the increased understanding limited. Wilks has been cited for the χ^2-test.

The Unconditional Spatial Correlation is a notation introduced by Mikkelsen et al. (1996) and it is considered unconditional because it takes into account the situations where extremes are monitored at both stations or at only one or the other station. Mikkelsen et al. (1996) explains the approach in detail.

The notation in equation 8 is explained (or outlined) in Mikkelsen et al. (1996) clarifying that $E\{Z_A|U\}$ is actually two values: one for U=0 and one for U=1. As such a covariance can be obtained. This has also been clarified in the manuscript.

2.5 Results and discussion

2.5.1 Fitting the weather generator

I prefer the term “estimation of parameters” over “fitting”. But that is a personal preference.
p12 l19 “The WG converges to an optimum” I suggest to be more precise with the concepts here. The WG is a stochastic model. The parameters are estimated by minimzing an objective function. The minimization algorithm can converge to an optimum, not the WG.
p12 l20 what is a “realistic” rainfall field?
p14 l3 what are the “features expected to have the highest influence on the produced extremes”?

The suggestion of changing “fitting” to “parameter estimation” has been integrated generally in the manuscript and the other suggestion for clarifications has been changed or further specified in the manuscript.

2.5.2 Evaluation of extremes for present climate conditions

p14 l10 Why are you choosing an 68% intervall? 95% is more common. How is that obtained?

The 68% confidence interval corresponds to the one standard deviation envelope and it was introduced in the manuscript based on the earlier review to give an idea of how far off the extremes in the data sets from the weather generator are compared to the uncertainty we have on the observations. A 95% confidence interval could have been used but would appear as a weaker criterion in this case. The supporting text has been altered to highlight this.

2.5.3 Perturbation of the WG with climate change signals from RCMs

p15 l8 I do not understand this sentence. Why a 30-year realisation?

30-year realizations are generated for two reasons. The data sets become very large for longer realizations (the limit is just over 50 years before the program crashes) and by generating time series of approximately the same length as the observations we generate data sets where the uncertainty of the extremes are expected to be at the same level as for the observations.

3 Technical comments and typos

p18 l11 “Neuman” change to Neyman

The typo has been corrected.

References

D. S. Wilks. Statistical methods in the atmospheric sciences. Academic Press, San Diego, CA, 3rd edition, 2011.

Downscaling future precipitation extremes to urban hydrology scales using a spatio-temporal Neyman-Scott weather generator

H. J. D. Sørup[1,2], O. B. Christensen[2], K. Arnbjerg-Nielsen[1] and P. S. Mikkelsen[1]

[1]{Urban Water ~~Engineering~~ Systems Section, Department of Environmental Engineering, Technical University of Denmark, Lyngby, Denmark}

[2]{Section for Climate and Arctic, Danish Meteorological Institute, Copenhagen, Denmark}

Correspondence to: H. J. D. Sørup (hjds@env.dtu.dk)

Abstract

Spatio-temporal precipitation is modelled for urban application at 1-hour temporal resolution on a 2 km grid using a Spatio-Temporal Neyman-Scott Rectangular Pulses weather generator (WG). Precipitation time series ~~for fitting~~use as input to the ~~model~~ WG are obtained from a network of 60 tipping-bucket rain gauges irregularly placed in a 40 by 60 km model domain. The ~~model~~ WG simulates precipitation time series that are comparable to the observations with respect to extreme precipitation statistics. The WG is used for downscaling climate change signals from Regional Climate Models (RCMs) with spatial resolutions of 25 km and 8 km respectively. Six different RCM simulation pairs~~s~~ are used to perturb the WG with climate change signals resulting in six very different perturbation schemes. All perturbed WGs result in more extreme precipitation at the sub-daily to multi-daily level and these extremes exhibit a much more realistic spatial pattern than what is observed in RCM precipitation output. The WG seems to correlate increased extreme intensities with an increased spatial extent of the extremes meaning that the climate-change-perturbed extremes have a larger spatial extent than those of the present climate. Overall, the WG produces robust results and is seen as a reliable procedure for downscaling RCM precipitation output for use in urban hydrology.

1 Introduction

Formatted: Justified, Space Before: 6 pt, After: 0 pt, Add space between paragraphs of the same style, Line spacing: 1,5 lines, Don't keep lines together, Hyphenate, Tab stops: Not at 0,75 cm

Pluvial flooding of urban areas is often caused by very local extreme precipitation at sub-daily temporal scale (Berndtsson and Niemczynowicz, 1988, Schilling, 1991). Traditionally, historical gauge measurements of precipitation at minute-scale temporal resolution are thus used as input to design and analysis of urban water infrastructure (Mikkelsen et al., 1998, Madsen et al., 2009, Arnbjerg-Nielsen et al., 2013). Climate change is, however, expected to change the occurrence rate and magnitude of extreme events causing urban pluvial flooding (Fowler and Hennessy, 1995; Larsen et al., 2009; Olsson et al., 2009, Sunyer et al., 2014), and high-resolution input time series representing future climates are therefore needed. Even though the overall qualitative features of precipitation are reproduced realistically by regional climate models (RCMs) (Christensen and Christensen, 2007) they are~~, however,~~ not able to capture the very fine-scale spatio-temporal features of precipitation satisfactorily and yield output that is too spatially correlated (Tebaldi and Knutti, 2007; Gregersen et al., 2013). To overcome this, either dynamic downscaling with climate models has to operate at much finer scales in order to properly describe convective precipitation dynamics (Kendon et al., 2014: Mayer et al., 2015) or further statistical downscaling of the climate model output has to be performed (Olsson and Burlando, 2002; Wood et al., 2004; Cowpertwait, 2006; Molnar and Burlando, 2008; Willems et al., 2012; Sunyer et al., 2012; Arnbjerg-Nielsen et al., 2013). Fine scale dynamic downscaling is computationally extremely expensive and statistical downscaling is therefore often favoured (Maraun et al., 2010). Several approaches exist within statistical downscaling, each with its pros and cons (Wilks and Wilby, 1999; Willems et al., 2012; Arnbjerg-Nielsen et al., 2013). In the present study a stochastic weather generator (WG) is used for statistical downscaling.

WGs can take different forms (Vrac et al., 2007; Burton et al., 2008; Arnbjerg-Nielsen and Onof, 2009; Chen et al., 2010; Cowpertwait et al., 2006; 2013) but they generally work by analysing observed precipitation (and possibly other weather related variables) and use the obtained statistics to create artificial stochastic precipitation (or weather) time series that replicate the behaviour of the observations with respect to these statistics (Maraun et al, 2010, Sunyer et al., 2012). Perturbation of the WG to yield output time series representing future climates is then possible by application of climate change factors calculated from output from RCMs (operation at too large space-time scales) to relevant parameters of the WG (that operates at the right space-time scale).

Several WGs exist that model precipitation as a stochastic point process where the given observations are considered single realisations of an underlying precipitation process (Waymire and Gupta 1981). Rodríguez-Iturbe et al. (1987a,b) developed the stochastic point process models in a way to better characterise and describe the precipitation process at the event level. Implementations of the stochastic point process models for spatio-temporal precipitation seem to work satisfactorily at ~~a~~ temporal resolutions ~~of~~ down to one hour ~~or higher~~ (Cowpertwait and O'Connell, 1997; Burton et al., 2008; 2010a; Cowpertwait et al., 2006: 2013). Also, downscaling to finer resolution than one hour is inherently problematic as the scaling properties change below this point (Nguyen et al., 2002; Molnar and Burlando, 2008). Thus, for downscaling of extreme precipitation at sub-daily level and subsequent application of climate change signals from climate models, stochastic weather generators implementing stochastic point process models seem useful (Cowpertwait, 1998; Furrer and Katz, 2008; Hundecha et al., 2009; Verhoest et al., 2010; Sunyer et al., 2012). The trade-off is that the models do not involve rainfall movement and, hence that the spatio-temporal scale of the model has to be such that rainfall movement is not the main descriptor of the spatial rainfall pattern.

Formatted: Font: 12 pt

At the daily level, the Neyman-Scott Rectangular Pulses (NSRP) and the Spatio-Temporal Neyman-Scott Rectangular Pulses (STNSRP) models (Burton et al., 2008; 2010a; 2010b; Cowpertwait et al., 2013) have shown good skill in downscaling point precipitation extremes. This applies for individual gauges (Sunyer et al., 2012) as well as for spatially averaged precipitation covering large areas considered as having a uniform climate described by relatively few gauges (e.g. 5 gauges for a 4000 km^2 basin in the Pyrenees (Burton et al., 2010a) and 3 gauges used to calibrate a regional model covering a catchment of 342 km^2 in the Basque Country (Cowpertwait et al., 2013)). This is however inadequate in urban hydrology where the rainfall dynamics causing effects under study occur on much smaller time and space scales.

In the present study, the STNSRP weather generator (WG) in the form of the software package RainSim (version 3.1.1, Burton et al., (2008)) is used in a new, urban hydrology context focusing on much smaller space and time scales than what has been done in previous studies. Due to the limitations in scalability of both RCM model output and precipitation measurements discussed above a temporal resolution of 1 hour is adopted, even though a higher resolution would be preferable from an urban hydrology perspective. ~~It is fitted to hourly~~ Hourly data from 60 rain gauges from a dense rain gauge network in Denmark are used to estimate parameters for the WG ~~and~~which is used to generate synthetic precipitation data series on a~~n~~ ~~equally~~ regular dense grid covering approximately 2400 km^2. The synthetic precipitation data is then evaluated with respect to its applicability for urban hydrological purposes. A 1-hour temporal resolution on a 2 km grid is chosen as realistic and sufficient performance scales of the model for fine-scale precipitation data in urban hydrology. The evaluation of the WG is done from an engineering perspective with respect to its ability to reproduce rainfall features relevant for urban hydrological modelling. We will thus focus on:

- the WG's ability to produce realistic extreme event intensities at point scale
- the WG's ability to reproduce the seasonal distribution of extreme events at point scale
- the WG's ability to reproduce small scale spatio-temporal correlation structures of the extreme events

This study uses the presented WG to analyse climate change in precipitation at scales comparable to the observational data sets traditionally used today for urban water infrastructure design and analysis. The WG is perturbed with climate change information obtained from a collection of temporal high resolution RCMs. Six RCM runs using three different RCMs, driven by three different GCMs and covering three different emission scenarios (ranging from average to very high emissions) are included in the analysis and four of the RCM runs are run as high resolution models at an 8 km grid. Finally, climate change at urban scale is assessed based on the perturbed WG output.

2 Data ~~and weather generator~~

2.1 ~~Data representing present conditions~~Observational data

The model area is a 40 by 60 km region covering the North-Eastern part of Zealand (Denmark) including Copenhagen, see Figure 1. This study uses two different observational data sets; Table 1 summarises their main characteristics.

The area is highly urbanised and has a dense but irregular network of rain gauges designed and used for urban hydrology applications. The main observational precipitation data set, SVK (abbreviation for *Spildevandskomiteen*, the Water Pollution Committee of the Society of Danish Engineers) is obtained from this dense network of high-resolution tipping bucket rain gauges (Jørgensen et al., 1998; Sunyer et al., 2013). Data from 60 stations that have been active between 2 and 34 years in the period 1979 and 2012 are included in the analysis; see Figure 1 for locations within the study area. Figure 2 shows (top) the temporal development of ~~(top)~~ the number of active stations, ~~and~~ (middle) the average distance between nearest neighbouring stations through the measuring period, and ~~Figure 2~~ (bottom) shows the

Formatted: Justified, Space Before: 6 pt, After: 0 pt, Add space between paragraphs of the same style, Line spacing: 1,5 lines, Outline numbered + Level: 1 + Numbering Style: 1, 2, 3, … + Start at: 2 + Alignment: Left + Aligned at: 0 cm + Tab after: 0,76 cm + Indent at: 0,76 cm, Don't keep lines together, Hyphenate distribution of record lengths by 2012. Generally, there has been an increase in the number of stations and a densification of the network over the years. Some studies impose a minimum length of the time series to be included in regionalisation studies, e.g. Madsen et al. (2009), but in this study the cross-correlation is of key interest and hence all gauges are included in the analysis regardless of their record length. The original data resolution is 1 min and 0.2 mm but for the present study, data is aggregated to hourly time series. This data set is used to estimate ~~(or calibrate, or fit)~~ most of the parameters of the WG.

The second observational data set included in the analysis is referred to as the Climate Grid Denmark (CGD) (Scharling 2012). It comprises spatially averaged daily data in a uniform 10 km grid for all of Denmark from 1989 to 2010 inclusive, cf. Figure 1. This data is generated based on a national network of gauges with 27 gauges within the study site (Scharling 1999) and is only used to estimate the spatial component in the WG.

~~Finally, a third data set is the output from the applied weather generator (WG). A total of 10 data sets comprising sets of 50 years' time series in the 2 km grid (as shown on Figure 1) are simulated as output from the WG. These data sets are used to corroborate the WG by refitting and rerunning it, evaluating the output variability and comparing the output statistics to those of observations.~~

2.2 Regional climate model data

Formatted: Space Before: 18 pt, After: 6 pt, Add space between paragraphs of the same style, Line spacing: 1,5 lines, Outline numbered + Level: 2 + Numbering Style: 1, 2, 3, … + Start at: 1 + Alignment: Left + Aligned at: 0 cm + Tab after: 1,02 cm + Indent at: 1,02 cm, Don't keep lines together, Hyphenate, Tab stops: Not at 1,25 cm

Precipitation output from ~~six~~ twelve different RCM runs representing present and future condition is used in this study, see Table 2. ~~Two~~ Four of the model runs are identical to the ones used by Gregersen et al. (2013), namely the two SRES A1B scenarios ~~driving~~ forcing the RCM RACMO (version 2.1, Meijgaard et al., 2008) and the RCM HIRHAM (version 5, Christensen et al., 2006) and their present counterparts.~~,~~ ~~which are both~~All RCM runs are driven by the GCM ECHAM5 (Roeckner et al., 2003) and are part on the ENSEMBLES project (van der Linden and Mitchell, 2009). ~~Both~~ All have a spatial resolution of 25 km and a temporal output resolution of 1 hour. These were the ~~two~~ ENSEMBLES runs we had available through personal contacts for the present study at ~~true~~ 1-hour resolution. The more generally available data series with only daily maximum 1-hour intensity are not sufficient for the employed downscaling procedure. The ~~four~~ other eight simulations used in this study are run with the RCM HIRHAM driven by the GCM EC-EARTH (Hazeleger et al., 2012) and the RCM WRF (Skamarock et al., 2005) driven by the GCM NorESM (Bentsen et al., 2013). The four future simulations use the RCP 4.5 and RCP 8.5 scenarios (van Vuuren et al., 2011), see Table 2. The spatial resolution of these simulations is 8 km and the output frequency is again 1 hour (Fox Maule et al., 2014; Mayer et al., 2015). ~~The simulations were carried out as part of the research project RiskChange (www.riskchange.dhigroup.com).~~ The SRES A1B and RCP 4.5 scenarios are considered comparable moderate forcing scenarios whereas the RCP 8.5 scenario is a very strong forcing scenario. All future RCM runs are related to RCM runs driven by the same GCM for present conditions when climate factors are calculated (Table 2).

As in Gregersen et al. (2013), climate change is considered uniform for all land cells over Denmark; this results in 87 considered grid cells for the ENSEMBLES SRES A1B simulations and 648 for the RiskChange RCP 4.5 and 8.5 simulations.

Formatted: Heading 2

2.3 Weather generator data

~~Finally, a~~The last ~~third~~ data set is the output from the applied weather generator (WG, described in Section 3). A total of 10 data sets comprising sets of 50 years' time series in the 2 km grid (as shown on Figure 1) are simulated as output from the WG. These data sets are used to corroborate the WG by refitting and rerunning it, evaluating the output variability and comparing the output statistics to those of observations.

Formatted: Heading 1, Space Before: 0 pt, After: 0 pt, Add space between paragraphs of the same style, Line spacing: single, No bullets or numbering, Don't keep lines together, Hyphenate, Tab stops: Not at 1,25 cm

3 Weather generator

Burton et al. (2008) gives a thorough description of the weather generator and its components, Burton et al. (2010a) an introduction to the application of the model and Burton et al. (2010b) an introduction to incorporation of climate change into the WG; the remainder of this section is, thus, only giving a brief introduction to the WG and a more in depth description of the workflow associated with working with the model is given in the supplement. Generally, the approach by Burton et al. (2010a) is followed with inclusion of climate change as described by Burton et al. (2010b) using the software presented by Burton et al. (2008).

~~2.3~~3.1 Parameters

The RainSim WG (version 3, Burton et al. (2008)) ~~WG~~ describes the spatio-temporal rain field as discs of rain (rain cells) with uniform rain intensity that temporarily occur and overlap in space and time to produce output that realistically describe the statistical properties of precipitation. As the calibration data set consists of point observations, the time series from the simulations are not grid cell averages but strictly comparable to what a gauge would have measured if present in a grid point. ~~(see Burton et al. (2010a) for a thorough description of the weather generator).~~

Formatted: French (France)

The WG parameters and their meaning and interactions are described in-depth in Burton et al. (2008) where a schematic representation of the WG is also found (Burton et al., 2008: Fig. 1). ~~As the calibration data set consists of point observations, the time series from the simulations are not grid cell averages but strictly comparable to what a gauge would have measured if present in a grid point.~~

A uniform Poisson process governed by λ describes the storm occurrences. For each storm a random number of rain cells are produced, which occur at independent time intervals after the storm origin and where the time intervals follow an exponential distribution with parameter β. A uniform spatial Poisson process governed by ρ describes the density of the rain cells in space. The cell radii are randomly drawn from an exponential distribution described by γ, and the duration and intensity of each rain cell is independent and follows an exponential distribution with parameters η and ξ, respectively. The rain intensity at a given point is therefore the sum of all overlapping rain cell intensities at a given time. In all, seven parameters describe the WG: ~~(Burton et al., 2010a):~~

- λ^{-1}, the mean waiting time between storm origins (in hours)
- β^{-1}, the mean waiting time for rain cell origins after storm origin (in hours)
- η^{-1}, the mean duration of rain cells (in hours)
- ρ, the spatial density of rainfall cell centres (cells per km^2)
- ξ^{-1}, the mean intensity of the rain cells (in mm/h)
- γ^{-1}, the mean radius of the rain cells (in km)
- Φ, the non-homogeneous intensity scaling field describing how the mean monthly rainfall intensity varies in space within the model area (-)

The non-homogeneous intensity scaling field, Φ, is a proxy for the spatial variation of mean monthly precipitation and is used for relative scaling of the precipitation in space; for this study it is interpolated from the CGD data set using inverse distance weighting. Regional modelling of short-duration extreme precipitation for Denmark using the SVK data set has shown that the only significant parameter that can explain the geographical variation of point extremes statistically is the corresponding mean annual precipitation (Madsen et al., 2002; 2009). Thus, taking Φ as the only spatially varying parameter in the WG, and as such the only parameter describing spatial differences within the WG, is considered to be an acceptable approximation. The actual spatial variation of mean monthly precipitation calculated from the CGD data set is considerable (see Figure 3), even though the model area is small in size and relatively flat. Especially in June and July there is a clear North-South gradient with 75-80 mm/month in the North of the area and 55-60 mm/month in the South.

3.2 ~~Fitting of the weather generator~~Parameter estimation

The parameters for RainSim (see Section 3.1) ~~is~~are ~~fitted to~~ estimated based on daily and hourly statistics for each calendar month from the observed time series (SVK). The objective function is adopted from Burton et al. (2010b: Eq. (2)) and the weighs are chosen ~~-~~to best reproduce features at both hourly and daily levels, as described by Burton et al. (2008; 2010a~~,~~; 2010b). ~~A~~The custom weighing scheme ~~is~~ used is constructed to support the features of rainfall that are important in the context of the present study (i.e. the higher order moments are assigned more weight to secure a realistic fit for the extremes, see Table 3)~~. RainSim uses the Shuffled Complex Evolution fitting algorithm in combination with an objective function that normalises the fitting statistics (to avoid bias) for optimisation; furthermore, the algorithm is run thrice to avoid sub-optima (Burton et al., 2008)~~. The statistics used for fitting the WG are:

- The mean daily precipitation intensity from the individual gauges (24 hour mean)
- The variance of the intensity of the daily and hourly observations from the individual gauges (1 hour and 24 hour variance)
- The skewness of the intensity of the daily and hourly observations from the individual gauges (1 hour and 24 hour skewness)

- The probability of dry days and of dry hours based on the observations from the individual gauges and with thresholds of 1.0 and 0.1 mm respectively as suggested by Burton et al. (2008).
- The lag-1 auto-correlation of the hourly precipitation intensity calculated from the observations at the individual gauges
- The cross-correlation between observations of hourly precipitation intensity at the individual gauges

~~A weighing scheme is created from general knowledge on rainfall and urban hydrology, which prioritizes rainfall features relevant for the present study.~~ The chosen weighing scheme ~~(see Table 3)~~ favours the higher order moment statistics, variance and skewness, over the mean as the extreme characteristics of the simulated precipitation is prioritised. Furthermore, the cross-correlation and auto-correlation are given high weights to ensure a realistic representation of the spatio-temporal extent of the simulated precipitation. The different observation time series are furthermore weighted relative to each other according to the effective length of the time series to give more weight to longer time series. This is done to increase the data basis for cross-correlation analysis, utilising that a great deal of the short time series are from recent years and thus overlap in time, see Figure 2.

The standard fitting bounds suggested by Burton et al. (2008) are applied in the fitting procedure to ensure that the WG is fitted with values that are considered realistic by the model developers for a Northern European climate.

~~2.1~~3.3 Perturbation of the weather generator with climate change signals

The ~~fitted~~ WG is perturbed with climate change signals by application of change factors, $\alpha_{i,j,k}$'s, to the statistics, $Y_{i,j,k}^{Present}$'s, calculated from the SVK data set and used ~~to fit the~~for the original ~~WG~~parameter estimation for ~~the~~ present climate. In this manner new statistics are produced for ~~the~~ future climate, $Y_{i,j,k}^{Future}$'s, as (Fowler et al., 2007, Burton et al., 2010b):

$$Y_{i,j,k}^{Future} = \alpha_{i,j,k} Y_{i,j,k}^{Present} \qquad (~~2~~1)$$

where one climate change factor, $\alpha_{i,j,k}$, is calculated for each statistic, i, for each month, j, for each RCM, k. The change factors are calculated using the methodology introduced by Burton et al. (2010b: Eq. 1-3) which includes transformations that ensure that the bounded statistics (probabilities of dry days and hours and auto-correlation) stays within their prescribed boundaries (further described in the supplement). No change factor is calculated for the cross-correlation as this statistic is described poorly by the RCMs (Gregersen et al., 2013).

~~3~~4 Methodology

Formatted: Justified, Space Before: 6 pt, After: 0 pt, Add space between paragraphs of the same style, Line spacing: 1,5 lines, Outline numbered + Level: 1 + Numbering Style: 1, 2, 3, … + Start at: 2 + Alignment: Left + Aligned at: 0 cm + Tab after: 0,76 cm + Indent at: 0,76 cm, Don't keep lines together, Hyphenate

~~3.12.1 Fitting of the weather generator~~

~~RainSim is fitted to daily and hourly statistics for each calendar month from the observed time series (SVK) to best reproduce features at both hourly and daily levels, as described by Burton et al. (2008; 2010a,b). A custom weighing scheme is used to support the features of rainfall that are important in the context of the present study. RainSim uses the Shuffled Complex Evolution fitting algorithm in combination with an objective function that normalises the fitting statistics (to avoid bias) for optimisation; furthermore, the algorithm is run thrice to avoid sub-optima (Burton et al., 2008). The statistics used for fitting the WG are:~~

- ~~The mean daily precipitation intensity from the individual gauges (24 hour mean)~~
- ~~The variance of the intensity of the daily and hourly observations from the individual gauges (1 hour and 24 hour variance)~~
- ~~The skewness of the intensity of the daily and hourly observations from the individual gauges (1 hour and 24 hour skewness)~~
- ~~The probability of dry days and of dry hours based on the observations from the individual gauges and with thresholds of 1.0 and 0.1 mm respectively as suggested by Burton et al. (2008).~~
- ~~The lag-1 auto-correlation of the hourly precipitation intensity calculated from the observations at the individual gauges~~
- ~~The cross-correlation between observations of hourly precipitation intensity at the individual gauges~~

~~A weighing scheme is created from general knowledge on rainfall and urban hydrology, which prioritizes rainfall features relevant for the present study. The chosen weighing scheme (see Table 3) favours the higher order moment statistics, variance and skewness, over the mean as the extreme characteristics of the simulated precipitation is prioritised. Furthermore, the cross-correlation and auto-correlation are given high weights to ensure a realistic representation of the spatio-temporal extent of the simulated precipitation. The different~~

~~observation time series are furthermore weighted relative to each other according to the effective length of the time series to give more weight to longer time series. This is done to increase the data basis for cross-correlation analysis, utilising that a great deal of the short time series are from recent years and thus overlap in time, see Figure 2.~~

~~The standard fitting bounds suggested by Burton et al. (2008) are applied in the fitting procedure to ensure that the WG is fitted with values that are considered realistic by the model developers for a North European climate.~~

~~3.2~~4.1 Evaluation of simulated time series

Formatted: Space Before: 18 pt, After: 6 pt, Add space between paragraphs of the same style, Line spacing: 1,5 lines, Outline numbered + Level: 2 + Numbering Style: 1, 2, 3, … + Start at: 1 + Alignment: Left + Aligned at: 0 cm + Tab after: 1,02 cm + Indent at: 1,02 cm, Don't keep lines together, Hyphenate, Tab stops: Not at 1,25 cm

The evaluation of the simulated time series will be in line with previous studies such as Olsson and Burlando (2002), Cowpertwait (2006) and Molnar and Burlando (2008). This implies that simulated time series are not evaluated against the observations with the expectation of a perfect fit; the expectation is rather that the simulated series should resemble measured precipitation. In practise this is achieved by analysis of the statistics used in the fitting procedure and through analysis of statistics which are independent of the fitting statistics as will be outlined in Section ~~3.4~~4.2.

For evaluation of all realisations of the WG the 60 grid cells closest to the observational gauges are extracted and evaluated point-wise with respect to all the fitting statistics as recommended by Burton et al. (2008). Furthermore, the WG is refitted to the simulated data sets to evaluate if the realisation is representative and results in model parameters that are comparable to the parameters estimated from the SVK observational data set.

Ten realisations of the WG, named WG1 to WG10, are used in this study. The actual simulation time is very short, but the process of writing data to text files for the complete grid takes long time.~~ Also, the refitting of the WG data sets takes a long time to complete~~, making it a rather cumbersome approach, which limits the number of realisations evaluated in this study.

The refitted WG data is evaluated with respect to the fitting statistics, $Y_{i,j,k}^{WG}$, through discussion of the density plots for the normalized error against the SVK data set:

$$\epsilon = \frac{Y_{i,j,k}^{WG} - Y_{i,j,k}^{SVK}}{Y_{i,j,k}^{SVK}} \qquad (\cancel{1}2)$$

~~3.3~~4.2 Evaluation of extremes

Gregersen et al. (2013) compare extreme precipitation observations with RCM output. One issue is the difference in absolute magnitude of the extremes, which can partly be explained by the inherent difference between gridded data and point observations; another issue is the spatial correlation structure of the extremes, where extremes calculated from RCM output ~~have much too large spatial correlation distances~~are much more spatially correlated at the sub-daily time scale. In this study, a ~~simulated~~ data set simulated with a WG will be considered better than using RCM data directly for the specified purpose if it better resembles the observations with respect to both the absolute magnitude and the spatial correlation structure of the extremes.

The statistics used in this study to evaluate the WG's performance with respect to simulating extreme precipitation are based on the identification of independent rainfall events, as done when estimating intensity-frequency-duration relationships, see e.g. Madsen et al. (2002). Individual events are separated by dry periods equal to or longer than the chosen event duration (i.e. 1-hour events have at least 1 hour of dry weather between them and 24-hour events have at least 24 hours of dry weather between them) and the maximum averaged event intensities over the chosen durations are ~~noted~~calculated. Furthermore, the Peak over Threshold (POT) approach from Mikkelsen et al. (1996) and Madsen et al. (2002) is adopted with a global constant intensity threshold (i.e. Type I censoring) to ~~derive~~ define the extreme events ~~intensities~~ for each gauge/grid point. In this study, extreme precipitation events are evaluated for 11 distinct durations of 1, 2, 3, 4, 6, 8, 12, 24, 48, 72 and 120 hours with thresholds ranging (approximately log-linearly) from 7.6 to 0.34 mm/hour (the same as used by Gregersen et al. (2013) for the SVK data set). Three different event-based indices of extreme precipitation are evaluated as explained below.

~~2.1.1~~4.2.1 ~~Extreme event statistics~~ Magnitude of extreme events

To evaluate the magnitude of the extreme events intensity-duration-frequency relationships are calculated for all data sets. First, t~~T~~he return periods of extreme events extracted from an observed or simulated rainfall time series is calculated using the California plotting position formula:

$$T_m = \frac{T_{obs}}{m} \quad (3)$$

Formatted: Space Before: 18 pt, After: 6 pt, Add space between paragraphs of the same style, Line spacing: 1,5 lines, Outline numbered + Level: 2 + Numbering Style: 1, 2, 3, … + Start at: 1 + Alignment: Left + Aligned at: 0 cm + Tab after: 1,02 cm + Indent at: 1,02 cm, Don't keep lines together, Hyphenate, Tab stops: Not at 1,25 cm where T_m is the return period of the event (years) with rank m and T_{obs} is the observation period (years) of the time series. T_m is obviously affected by sampling variability and is biased, especially for large return periods. There are more elaborate methods to estimate T_m than Eq. (~~2~~3), but we use Eq. (~~2~~3) here because it allows for comparing extreme value curves from multiple sites (including sampling variability and spatial variability) in a straightforward way.

Secondly, a~~A~~ Generalised Pareto Distribution is fitted to extremes from every single time series:

$$z_T = z_0 + \mu \frac{1+\kappa}{\kappa}\left(1 - \left(\frac{1}{\lambda T}\right)^{\kappa}\right) \quad (4)$$

where:

- z_T is the intensity for extreme event with return period T
- z_0 is the threshold
- μ is the mean intensity of the extreme events
- λ is the mean number of extremes per year
- κ is the shape parameter
- T is the return period

Finally, b~~B~~ased on $z(T)$'s intensity-duration-frequency curves are calculated for each data set.

For the climate change scenarios, climate factors for the intensity of the extreme events are calculated as a function of the return period for different T-year event durations. This is done as a simple ratio between the present and future levels for a given return period as

$$CF_T = \frac{z(T)^{future}}{z(T)^{present}} \quad (5)$$

~~2.1.2~~4.2.2 Seasonality of extreme events

The seasonality of the extreme events is determined to further evaluate the realism of the behaviour of the WG. This is done to evaluate whether the WG data set constructed with individual monthly model parameters results in a realistic distribution of the extremes throughout the year. The same extreme events used in the evaluation of the magnitude are used in this analysis.

The determination is in practice performed by counting the number of extremes from the POT analysis that occur within each month for the SVK and WG data sets. These are then normalised and compared with a ~~χ^2-~~χ^2-test (Wilks, 2011) where the normalised counts C for the SVK data act as the expected values for the WG data set and where the summation is done over months giving a test statistic x:

$$x = \sum_{i=1}^{12} \frac{\left(C_i^{WG} - C_i^{SVK}\right)^2}{C_i^{SVK}} \tag{6}$$

x then follows a χ^2-distribution with (12-1)(2-1) = 11 degrees of freedom.

~~2.1.3~~4.2.3 Unconditional ~~spatial~~ Spatial ~~correlation~~ Correlation of extremes

The ~~unconditional~~ Unconditional ~~spatial~~ Spatial ~~correlation~~Correlation (Mikkelsen et al., 1996), ρ, between the intensities of extreme events that are considered concurrent at different sites A and B is estimated. The methodology follows Mikkelsen et al. (1996) with the i'th extreme intensity Z_{Ai} measured at site A being concurrent with the j'th extreme event Z_{Bj} measured at site B if Eq. 7 is fulfilled. In this framework the precipitation process is considered to generate random occurrences of precipitation that are treated as correlated random variables, Z_A and Z_B, and two events are considered concurrent if they are overlapping in time or at most separated by a lag time Δt, which is introduced to account for the travel time of rain storms between sites.

$$\{Z_{Ai}, Z_{Bj}\}: \left[t_{si} - \frac{\Delta t}{2}, t_{ei} + \frac{\Delta t}{2}\right]_A \cap \left[t_{sj} - \frac{\Delta t}{2}, t_{ej} + \frac{\Delta t}{2}\right]_B \neq \emptyset \tag{7}$$

Here t_s's ~~is~~ are the start times of the events and t_e's ~~is~~ are the end times of events. A lag time of Δt = 11 hours + the duration of the event is adopted in accordance with Gregersen et al. (2013). The introduction of this lag time, in combination with lack of knowledge of the movement direction of precipitation, implies that an individual event at one site can be correlated to more than one event at another site.

The unconditional covariance is then estimated by also accounting for non-concurrent extreme events at the two sites as:

$$\operatorname{cov}\{Z_A, Z_B\} = \operatorname{cov}\{\mathrm{E}\{Z_A|U\}, \mathrm{E}\{Z_B|U\}\} + \mathrm{E}\{\operatorname{cov}\{Z_A, Z_B|U\}\} \tag{8}$$

with U being a boolean operator taking the value of $U = 1$ if events are concurrent and $U = 0$ otherwise. Thus, $\mathrm{E}\{Z|U\}$'s are not a single values, but two values for $U = 0$ and $U = 1$ respectively and a covariance between them can be calculated.

Formatted: Font: Italic

Formatted: Font: Italic

Finally, the ~~unconditional~~ Unconditional Spatial ~~correlation~~ Correlation is obtained by division of Eq. (8) with the sample standard deviations of the two sites (Mikkelsen et al., 1996):

$$\rho_{AB} = \frac{\mathrm{cov}\{Z_A,Z_B\}}{\sqrt{\mathrm{var}\{Z_A\}\,\mathrm{var}\{Z_B\}}} \qquad (9)$$

The ~~unconditional~~ Unconditional Spatial ~~correlation~~ Correlation values are grouped together in bins where the distance between the points considered are approximately the same, and an exponential model is fitted to describe the ~~unconditional~~ Unconditional Spatial ~~correlation's~~ Correlation's dependence on distance between sites using the e-folding distance measure as proposed by Gregersen et al. (2013).

~~4~~5 Results and discussion

Formatted: Justified, Space Before: 6 pt, After: 0 pt, Add space between paragraphs of the same style, Line spacing: 1,5 lines, Outline numbered + Level: 1 + Numbering Style: 1, 2, 3, … + Start at: 2 + Alignment: Left + Aligned at: 0 cm + Tab after: 0,76 cm + Indent at: 0,76 cm, Don't keep lines together, Hyphenate

~~4.1~~5.1 ~~Fitting the w~~Weather generator parameter estimation

~~The WG converges to an optimum fit for the SVK and CGD data for all calendar months, resulting in a WG that is able to simulate realistic rainfall fields all year round.~~ The parameter estimates (cf. Section ~~2.3~~3.2) for the model fitted to SVK data, the parameter estimates for the model refitted to the 10 realisations of the WG (WG1 – WG10) and the used boundary values are given in Figure 4. All parameters values shown in Figure 4 are given in the supplement. All parameters vary over the course of the year, some more smoothly than others. Note that the β parameter (the parameter controlling the arrival time of cells after a storm origin) is constrained at its prescribed minimum value for four months (February, September, October and December). However, rain events can easily last for several days at these times of the year in Denmark, and this fitting artefact is therefore considered to have limited influence on those features of rainfall, which are of interest for this study. Figure 4 shows that all the refitted values are different and especially the β parameter does not always seem to follow the same structural pattern as for the SVK data set. As β^{-1} controls the arrival time of cells after storm origin it will be heavily dependent on the actual realisation of weather from the WG and this is not considered to be important for the realised extreme events. The ξ parameter seems to be slightly biased in the same direction for all WGs. ξ^{-1} controls the mean intensity of the rain cells and the difference in fit suggests that the rain in the WG data sets are slightly more intense during summer than what is seen in the SVK data set. Generally, the WG data sets however represent the SVK data set well.

The fitting statistics (cf. section 3.1) resulting from the direct analysis of the observations (SVK data set) and the simulations (WG data sets that are simulated based on fitting the WG to the SVK and CGD data) are compared in Figure 5 through the normalized error (Eq. ~~1~~2) and directly in Table 4. Generally, the fit seems reasonable for all variables with a mean of the normalized errors close to zero. For the moment statistics the WG data sets seem to have a slight positive bias, and the variance and skewness distributions are also slightly positively skewed (Figure 5a-e). However the WG fit are still within the bounds reported for the SVK data set in Table 4. The lag-1 auto-correlation and the probabilities of dry hours seem to be fitted well even though the probability of dry days also seem to have some skewness in the error distribution. The probability of dry days is the only parameter that seems to differ between observations and WGs, indicate that the WG concentrates the precipitation on too few days. Also, it seems that none of the WG realisations performs differently than the others with respect to reproduction of the fitting statistics. Hence the discrepancies observed in Figure 4 do not seem to impede the use of the WGs as good proxies for observed precipitation..

The cross correlation of the 1-hour intensities is shown in Figure 6 for each month of the year. The 10 WG data sets seem to reflect the overall behaviour of the SVK data set very well and also capture most of the variability seen in the SVK data set. The very low correlations observed in the SVK data set for some “traces” of points, especially in March, October and November, are due to some time series only overlapping for very short time periods in recent years where the number of stations has increased dramatically (see Figure 2); hence the correlation is depending on only very few precipitation events. There is no evidence of a systematic pattern in these readings. Again, the difference between different WG realisations is very limited.

From Figures 5 and 6 the WG fit is considered satisfactory given the complex data set used and the purpose of this study. For analysis of extremes at event level this WG reproduces the higher order moment statistics, which are the features expected to have the highest influence on the produced extremes, well.

~~4.2~~5.2 Evaluation of extremes for present climate conditions

For durations of 1 to 120 hours the extreme events are extracted from the SVK data set at each gauge and from the WG data sets in each grid cell closest to the SVK observation points

Formatted: Space Before: 18 pt, After: 6 pt, Add space between paragraphs of the same style, Line spacing: 1,5 lines, Outline numbered + Level: 2 + Numbering Style: 1, 2, 3, … + Start at: 1 + Alignment: Left + Aligned at: 0 cm + Tab after: 1,02 cm + Indent at: 1,02 cm, Don't keep lines together, Hyphenate, Tab stops: Not at 1,25 cm and ranked according to return period (Eq. 3). Figure 7 shows intensity-duration-frequency curves estimated for WG realisation along with the SVK data set. For both 100 and 10-year events the WG data sets result in comparable extreme intensity values for all considered durations well within the shown 68% confidence interval (corresponding to a one standard deviation envelope) for the SVK IDF curve.

Figure 8 shows that the seasonal distribution of these extreme events is captured very well by the considered grids from the simulated WG data sets for all considered event durations. The χ^2 tests furthermore confirm that there are no significant differences between distributions for the WG and the SVK data sets for all event durations.

Figure 9 shows the unconditional spatial correlation for the SVK and for the selected WG grid points calculated according to Eq. (9) and grouped in selected bins. Table 5 furthermore compares the e-folding distances based on the fitted exponential models with a set of values calculated from RCM data representing a slightly larger area, taken from Gregersen et al. (2013).

Gregersen et al. (2013) show, using data from the whole of Denmark (range 0-350 km), that the spatial correlation pattern is not the same when considering output from climate models compared to SVK data as the climate model output maintains too long spatial correlation lengths at scales below approximately 150 km and 12 hours (see Table 5). Both Figure 9~~Figure 9~~ and Table 5 indicate that the WG better reproduces the spatial correlation pattern of the SVK data within the spatial range (0-60 km) covered by the observations included in this study. The e-folding distances computed in this study for the SVK data set are somewhat lower than the ones calculated by Gregersen et al. (2013). This is a consequence of inclusion of fewer gauges and, most importantly, that the time series in the SVK data set for this study have been aggregated into hourly time series prior to the smoothing and POT analysis. Gregersen et al. (2013) conducted the smoothing and POT analysis directly on the original time series that have a one-minute resolution. The WG data sets represent the space-time features of precipitation of crucial importance for urban hydrology applications much better than the climate model output; the WG data set is considered realistic at this small-scale spatio-temporal resolution.

Overall, the results show that the WG is able to realistically simulate extreme precipitation statistics down to the hourly scale at a 2x2 km spatial resolution.

Formatted: English (U.K.)

~~4.3~~5.3 Perturbation of the weather generator with climate change signals from RCMs

As the different realisations of the WG produce very similar ~~weather~~output, only one 30-years realisation is ~~used for~~generated for each perturbation with climate change signals from each of the RCMs. Furthermore, all grid cells are used for both present and future evaluations as no comparisons are made to the observational data.

For each RCM run and each statistic the change factors, $\alpha_{i,j,k}$'s, are calculated. All change factors and all parameters values for WG's representing future climate are given in the supplement. They are primarily above 1 for the moment derived statistics (Figure 10~~Figure 10~~a-e) but the different RCM runs appear different. For the 24 hour mean (Figure 10~~Figure 10~~a) the $\alpha_{i,j,k}$'s are mostly above 1 with all RCM runs showing some months with values below 1 in an unsystematic pattern. For both the 24 and 1 hour variances (Figure 10~~Figure 10~~b and d) the number of RCM runs and months that show a decrease is very limited and in general the variance will increase for all seasons. The HIRHAM RCP 8.5 simulation differs from the other RCM runs with very high $\alpha_{i,j,k}$'s for the summer months. The 24 and 1 hour skewness (Figure 10~~Figure 10~~c and e) show more clear seasonality than the mean and variance with higher $\alpha_{i,j,k}$'s from May to September for all RCM runs clearly indicating a shift in the distribution of precipitation intensities towards more extremes. Again the HIRHAM RCP 8.5 run stands out with very high $\alpha_{i,j,k}$'s for the 1 hour skewness for most of the year. This means that the extreme precipitation intensities are expected to be higher during summer and especially the sub-daily extremes for the HIRHAM RCP 8.5 perturbation could have very high intensities as a combination of a large increase in both variance and skewness will result in many severe precipitation events with a high mean intensity.

For the lag-1 hour auto-correlation (Figure 10~~Figure 10~~h) the $\alpha_{i,j,k}$ are mostly below 1 indicating more variations from one hour to the next and thus a possibility of more abrupt changes in the rainfall at the hourly level. For the probability of dry days and dry hours (Figure 10~~Figure 10~~f and g) the pattern is less clear. The RCM simulations show some variation around 1 (approximately between 0.7 and 1.7) but do not agree with respect to season of these changes or their relative magnitude. This suggests that future rainfall will follow the same overall patterns as today but as all RCM runs have months with $\alpha_{i,j,k}$ below 1 there will also be more severe periods since the precipitation is concentrated on fewer days and hours. For instance, the peaks for the WRF RCP 8.5 perturbation in August for both

Formatted: Space Before: 18 pt, After: 6 pt, Add space between paragraphs of the same style, Line spacing: 1,5 lines, Outline numbered + Level: 2 + Numbering Style: 1, 2, 3, ... + Start at: 1 + Alignment: Left + Aligned at: 0 cm + Tab after: 1,02 cm + Indent at: 1,02 cm, Don't keep lines together, Hyphenate, Tab stops: Not at 1,25 cm

Formatted: English (U.K.)

Formatted: English (U.K.)

Formatted: English (U.K.)

Formatted: English (U.K.)

Formatted: English (U.K.)

Formatted: English (U.K.)

probability of dry days and hours (Figure 10~~Figure 10~~f and g) in combination with the increases in variance and skewness (Figure 10~~Figure 10~~b to e) are expected to result in very severe extremes as the increased rainfall amount is expected to occur on fewer days. All in all, the $\alpha_{i,j,k}$'s indicate that for all RCM runs there will be more rainfall on average and it will be more variable resulting in more (and more severe) extremes events. This is in accordance with general findings from studies based on direct output from RCMs (Christensen and Christensen, 2007; Sunyer et al., 2014).

Formatted: English (U.K.)

~~4.4~~5.4 Changes in climate changed extremes from the weather generator

Formatted: Space Before: 18 pt, After: 6 pt, Add space between paragraphs of the same style, Line spacing: 1,5 lines, Outline numbered + Level: 2 + Numbering Style: 1, 2, 3, … + Start at: 1 + Alignment: Left + Aligned at: 0 cm + Tab after: 1,02 cm + Indent at: 1,02 cm, Don't keep lines together, Hyphenate, Tab stops: Not at 1,25 cm

Calculating the climate factors, *CF*'s(Eq. 5), from the perturbed and original WG using the *T*-year event estimates calculated with Eq. 4 shows that despite the differences observed in the $\alpha_{i,j,k}$ for the input statistics (Figure 10~~Figure 10~~), the perturbation schemes based on RCM simulations modelling comparable climate change (HIRHAM SRES A1B, RACMO SRES A1B, HIRHAM RCP 4.5 and WRF RCP 4.5) result in similar changes to extremes after downscaling with the WG (Figure 11~~Figure 11~~). Clearly, and as expected from the results in Figure 10~~Figure 10~~, the HIRHAM RCP 8.5 perturbed WG results in a much more severe change in extreme precipitation than the other perturbation schemes for both 10 and 100 year return periods. It is interesting that the WG perturbed with HIRHAM SRES A1B results in a rather stable *CF* in the range 1.35-1.55 with seemingly little dependence on return period and event duration, The WGs perturbed with RACMO SRES A1B, HIRHAM RCP 4.5 and WRF RCP 4.5 show similar *CF* values that are higher for 100-year extremes than for 10-year extremes but still not depend significantly on the event duration.

Formatted: English (U.K.)

Formatted: English (U.K.)

Formatted: English (U.K.)

Both the HIRHAM RCP 8.5 and WRF RCP 8.5 perturbed WGs yield *CF* values that depend on the event duration with higher *CF* for short duration precipitation extremes. This indicates that this high-end scenario is changing the climate more drastically than the more moderate scenarios (SRES A1B and RCP 4.5) and that the observed extreme effects are not linearly scalable from moderate to high end scenarios. For event durations above 48 hours the different WGs yield similar *CF*'s, but surprisingly the high-end scenario WRF RCP 8.5 perturbation scheme results in the smallest *CF* for the long duration events. This may indicate that the direct output from the RCMs underestimate the changes occurring at high spatio-temporal resolutions.

Despite the observed differences between WGs perturbed with different RCM runs and different forcing scenarios the results show an upwards change for all event durations (see Figure 11). The change seems to increase with the return period with a projected change factor in the order of 1.2-1.3 for T=10 years and 1.4-1.5 for T=100 years for the moderate scenarios (SRES A1B and RCP 4.5). Furthermore, the RCP 8.5 scenario perturbed WG runs suggest that short duration extreme events become relatively more severe compared to the WG runs perturbed with the other, moderate forcing scenarios.

~~4.5~~5.5 Unconditional spatial correlation of climate changed *T*-year events

All the perturbed WG runs produce T-year precipitation events with reasonable spatial correlation structure (Figure 12~~Figure 12~~, Table 6) includes calculated e-folding distances and it is noteworthy that the e-folding distance for present conditions is somewhat shorter for the full WG data set compared to the sub sets closest to the observations shown in Figure 9. The HIRHAM RCM and WRF RCM perturbed WG runs present similar results for all event durations whereas the RACMO SRES A1B perturbed WG run yield slightly larger correlations lengths for the very short durations (Figure 12~~Figure 12~~a). Generally, all the perturbed WG runs have larger correlation lengths than for the present climate, suggesting that the WG implicitly expects that more severe events on average also results in events with a larger spatial extent. This behaviour has recently been observed by Kendon et al. (2014) using a high resolution regional climate model (1.5 km resolution). This difference, however, is limited, and in general the WG produces extremes with a spatial extent much closer to that of observations than RCMs. Online Resource 1 includes an animation of extreme precipitation events generated directly as output from the 25 km resolution RCM HIRHAM SRES A1B, the 8 km resolution RCM HIRHAM RCP 4.5 and the 2 km WG evaluated in this study. From these it is clear that the small-scale variability is much more pronounced for the WG output than for the output of the RCMs, but also that the WG output lacks rainfall movement. At the hourly scale this is not a problem for a catchment of the size presented in the Online Resource (same as shown in Figure 1).

Only few apparent effects are observed with respect to choice of RCM, GCM and RCM spatial resolution and it is not possible to detect any systematic patterns. The WG seems to produce robust results with respect to change in extreme precipitation due to climate change that are similar for similar climate forcing scenarios.

Formatted: Space Before: 18 pt, After: 6 pt, Add space between paragraphs of the same style, Line spacing: 1,5 lines, Outline numbered + Level: 2 + Numbering Style: 1, 2, 3, … + Start at: 1 + Alignment: Left + Aligned at: 0 cm + Tab after: 1,02 cm + Indent at: 1,02 cm, Don't keep lines together, Hyphenate, Tab stops: Not at 1,25 cm

Formatted: English (U.K.)

~~5~~6 Conclusions

Precipitation time series based on high-resolution gauge measurements are presently used as input to design and analysis of urban water infrastructure, and time series representing future climates are needed in the future. Current RCMs operating at 25 and even 8 km spatial scales however yield too spatially correlated output that poorly represents the fine-scale precipitation features relevant for urban hydrology. The study indicate that statistical downscaling of precipitation output from RCMs using a stochastic weather generator (WG) is therefore a better solution.

This study demonstrates that the chosen Spatio-Temporal ~~Neuman~~Neyman-Scott Rectangular Pulses weather generator (WG) fitted to a dense network of 60 rain gauges in a 40 by 60 km region simulates realistic extreme precipitation of relevance to urban hydrology. Output is generated at the 1 hour temporal scale at a 2 km spatial grid, which is finer than what previous studies using this WG have focused on. Even though urban hydrology literature claims that rain data are ideally needed at a time scale of minutes, the hourly scale chosen here can still be of much use when assessing climate change impacts in urban hydrology as it is much finer than what regional climate models can currently provide.

The WG generally reproduces statistics of the observations such as mean, variance and skewness of the rainfall intensity distribution well at both the hourly and daily levels. It also produces realistic levels of lag-1 auto-correlation, cross-correlation between output at different grid points and probabilities of dry days and hours. Evaluating the WG from an urban hydrological engineering perspective yields the following conclusions:

- The extreme events of the simulated time series show realistic levels of intensity as well as a reasonable spatial variability for the full 60x40 km model area. Thus, the WG handles the large data set of spatially distributed observational input in a robust manner.
- The seasonal distribution of the extremes are not significantly different in the generated WG data sets compared to the observed SVK data set, implying that the applied procedure of individual monthly model fits results in a realistic seasonal behaviour of the WG.
- The spatial extent of the extreme events in the WG data set, as evidenced by the unconditional spatial correlation of extremes, is close to that of the observational SVK data set with e-folding distances in the same order of

Formatted: Justified, Space Before: 6 pt, After: 0 pt, Add space between paragraphs of the same style, Line spacing: 1,5 lines, Outline numbered + Level: 1 + Numbering Style: 1, 2, 3, … + Start at: 2 + Alignment: Left + Aligned at: 0 cm + Tab after: 0,76 cm + Indent at: 0,76 cm, Don't keep lines together, Hyphenate magnitude. This is much better than what is observed for Regional Climate Model (RCM) output at 25 and 8 km grid scale in previous studies.

This indicates that the WG is a good way to downscale spatio-temporal precipitation output from RCMs to relevant urban scales and that the simulated output can be used directly as input to urban hydrological models.

Output from six different RCM runs representing average to high emission scenarios are used to perturb the WG for different possible future climate scenarios. Two have a 25 by 25 km spatial resolution and four have a very high 8 by 8 km spatial resolution, and all RCM data sets are available at hourly temporal resolution. A clear increase in the magnitude of extreme precipitation is observed for all climate change perturbations of the WG.

This study highlights that different RCMs run with the same greenhouse gas emission scenario can result in different precipitation output and hence different CFs for perturbation of the WG. Despite these observed differences, downscaling with the WG results in similar extreme precipitation behaviour for similar emission scenarios.

Most perturbed WGs confirm that there is a more severe climate change signal for extreme events. The two WGs perturbed by the RCP 8.5 scenario show a more severe climate change signal for short-duration events. However, this finding is not shared by the other emission scenarios, suggesting that extreme precipitation at *T*-year event level is not scalable between emission scenarios. The spatial correlation structure of the WG output is slightly altered by the perturbation indicating a built-in correlation between intensity and spatial extent and suggesting that precipitation extremes in a future climate may have larger spatial extent than extremes in the present climate.

Acknowledgements

This work was carried out with the support of the Danish Council for Independent Research as part of the project “Reducing Uncertainty of Future Extreme Precipitation”, contract no. 09-067455. The observational SVK data set was provided by the Water Pollution Committee of the Society of Danish Engineers. The Climate Grid Denmark (CGD) is a commercial product made freely available for research by the Danish Meteorological Institute. The authors also thank the Royal Netherlands Meteorological Institute, KNMI, and Erik van Meijgaard who kindly provided the RACMO data in a temporal resolution of 1 h, although this was outside the agreement of the ENSEMBLES project. The high temporal resolution

HIRHAM/ECHAM data was provided by DMI. The high resolution regional climate model runs were carried out as part of the project RiskChange funded by the Danish Council for Strategic Research, contract no. 10-093894 (http://riskchange.dhigroup.com).

References

Arnbjerg-Nielsen, K. and Onof, C.: Quantification of anticipated future changes in high resolution design rainfall for urban areas. *Atmospheric Research*, 2(3) 350-363, doi: 10.1016/j.atmosres.2009.01.014. 2009

Arnbjerg-Nielsen, K., Willems, P., Olsson, J., Beecham, S., Pathirana. A., Gregersen, I.B., Madsen, H., Nguyen, V-T-V.: Impacts of climate change on rainfall extremes and urban drainage systems: a review. *Water Science and Technology*, 68(1), 16-28. doi: 10.2166/wst.2013.251. 2013.

Bentsen, M., Bethke, I., Debernard, J. B., Iversen, T., Kirkevåg, A., Seland, Ø., Drange, H., Roelandt, C., Seierstad, I. A., Hoose, C., and Kristjánsson, J. E.: The Norwegian Earth System Model, NorESM1-M – Part 1: Description and basic evaluation of the physical climate. *Geoscientific Model Development*, 6, 687-720. doi: 10.5194/gmd-6-687-2013. 2013.

Berndtsson, R. and Niemczynowicz, J.: Spatial and temporal scales in rainfall analysis: Some aspects and future perspectives. *Journal of Hydrology*, 100: 293-313. doi: 10.1016/0022-1694(88)90189-8. 1988

Burton, A., Kilsby, C. G., Fowler, H. J., Cowpertwait, P. S. P. and O'Connel, P. E.: RainSim: a spatial temporal stochastic rainfall modelling system. *Environmental Modelling and Software,* 23(12), 1356-1369. doi: 10.1016/j.envsoft.2008.04.003. 2008.

Burton, A., Fowler, H. J., Kilsby, C. G., and O'Connell, P. E.: A stochastic model for the spatial-temporal simulation of nonhomogeneous rainfall occurrence and amounts, *Water Resources Research,* 46(11). doi:10.1029/2009WR008884. 2010a.

Burton, A., Fowler, H.J., Blenkinsop, S., and Kilsby, C.G.: Downscaling transient climate change using a Neyman-Scott Rectangular Pulses stochastic rainfall model , *Journal of Hydrology*, 381 (1-2) 18-32, DOI: 10.1016/j.jhydrol.2009.10.031. 2010b.

Chen, J., Brissette, F. P., and Leconte, R.: A daily stochastic weather generator for preserving low-frequency of climate variability, *Journal of Hydrol*ogy, 388, 480–490. doi:10.1016/j.jhydrol.2010.05.032. 2010.

Cowpertwait, P. S. P.: A Poisson-cluster model of rainfall: high-order moments and extreme values. *Proceedings of the Royal* society *A,* 454, 885-898. doi: 10.1098/rspa.1998.0191. 1998.

Cowpertwait, P. S. P.: A spatial-temporal point process model of rainfall for the Thames catchment, UK. *Journal of Hydrology*, 330(3-4), 586–595. doi:10.1016/j.jhydrol.2006.04.043. 2006.

Cowpertwait, P. S. P. and O'Connell, P. E.: A Regionalised Neyman-Scott Model of Rainfall with Convective and Stratiform Cells. *Hydrology and Earth System Sciences*, 1(1), 71-80. doi: 10.5194/hess-1-71-1997. 1997.

Cowpertwait, P. S. P., Ocio, D., Collazos, G., de Cos, O. and Stocker, C.: Regionalised spatiotemporal rainfall and temperature models for flood studies in the Basque Country, Spain. *Hydrology and Earth System Sciences*, 17, 479–494. doi: 10.5194/hess-17-479-2013. 2013.

Christensen, O. B. and Christensen, J. H.: A summary of the PRUDENCE model projections of changes in European climate by the end of the century. *Climatic Change*, 81(1), 7-30. doi: 10.1007/s10584-006-9210-7. 2007.

Christensen, O. B., Drews, M., Christensen, J. H., Dethloff, K., Ketelsen, K., Hebestadt, I., Rinke, A.: The HIRHAM Regional Climate Model, version 5(β). Danish Meteorological Institute Technical report 06–17. 2006.

Fowler, A. M. and Hennessy, K.J.: Potential impacts of global warming on the frequency and magnitude of heavy precipitation. *Natural Hazards* 11:283–303. doi:10.1007/BF00613411. 1995.

Fowler, H. J., Blenkinsop, S. and Tebaldi, C.: Review linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling. *International Journal of Climatology* 27, 1547–1578. doi: 10.1002/joc.1556. 2007.

Fox Maule, C., Mayer, S., Sobolowski, S. and Christensen, O. B.: *Background information on the RiskChange simulations by BCCR and DMI.* Danish Climate Centre Report 14-05. The Danish Meteorological Institute, Copenhagen, Denmark. 2014.

Furrer, E. M. and Katz R. W.: Improving the simulation of extreme precipitation events by stochastic weather generators. *Water Resources Research,* 44(12). doi:10.1029/2008WR007316. 2008.

Gregersen I. B., Sørup H. J. D., Madsen H., Rosbjerg D., Mikkelsen P. S. and Arnbjerg-Nielsen K.: Assessing future climatic changes of rainfall extremes at small spatio-temporal scales. *Climatic Change*. 118(4), 783-797. doi: 10.1007/s10584-012-0669-0. 2013.

Hazeleger, W., Wang, X., Severijns, C., Ştefănescu, S., Bintanja, R., Sterl, A., Wyser, K., Semmler, T., Yang, S., van den Hurk, B., van Noije, T., van der Linden, E. and van der Wiel, K.: EC-Earth V2.2: description and validation of a new seamless earth system prediction model. *Climate Dynamics* 39(11), 2611-2629. doi: 10.1007/s00382-011-1228-5. 2012.

Hundecha, Y., Pahlow, M. and Schumann, A.: Modeling of daily precipitation at multiple locations using a mixture of distributions to characterize the extremes. *Water Resources Research,* 45(12). doi:10.1029/2008WR007453. 2009.

Formatted: Font: Italic

Formatted: English (U.S.)

Jørgensen, H. K., Rosenørn, S., Madsen, H. and Mikkelsen, P. S.: Quality control of rain data used for urban runoff systems. *Water Science and Technology*, 37(11), 113-120. doi: 10.1016/S0273-1223(98)00323-0. 1998.

Kendon, E. J., Roberts, N. M., Fowler, H. J., Roberts, M.J., Chan, S. C. and Senior, C.A. (2014) Heavier summer downpours with climate change revealed by weather forecast resolution model. Nature Climate Change, 4(7), 570-576.

Larsen, A. N., Gregersen, I. B., Christensen, O. B., Linde, J. J. and Mikkelsen, P. S.: Potential future increase in extreme precipitation events over Europe due to climate change. *Water Science and Technology*, 60(9), 2205-2216. doi: 10.2166/wst.2009.650. 2009.

Madsen, H., Mikkelsen, P. S., Rosbjerg, D. and Harremoes, P.: Regional estimation of rainfall intensity-duration-frequency curves using generalized least squares regression of partial duration series statistics. *Water Resources Research,* 38(11), 21-1-21-11. doi:10.1029/2001WR001125. 2002.

Madsen, H., Arnbjerg-Nielsen, K. and Mikkelsen, P. S.: Update of regional intensity-duration-frequency curves in Denmark: Tendency towards increased storm intensities. *Atmospheric Research* 92(3), 343-349. doi: 10.1016/j.atmosres.2009.01.013. 2009.

Maraun, D., Wetterhall, F., Ireson, A. M., Chandler, R. E., Kendon, E. J., Widmann, M., Brienen, S., Rust, H. W., Sauter, T., ThemeßI, M., Venema, V. K. C., Chun, K. P., Goodess, C. M., Jones, R. G., Onof, C., Vrac, M. and Thiele-Eich, I.:. Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user. *Reviews of Geophysics* 48(3). doi: 10.1029/2009RG000314. 2010.

Mayer, S., Maule, C., Sobolowski, S., Christensen, O., Sørup, H., Sunyer, M., Arnbjerg-Nielsen, K., and Barstad, I.: Identifying added value in high-resolution climate simulations over Scandinavia. *Tellus A*, 67. doi:http://dx.doi.org/10.3402/tellusa.v67.24941. 2015.Meijgaard, E. v, Ulft, L. H. v, Berg, W. J. v d, Bosveld, F. C., Hurk, B. J. J. M. v d, Lenderink, G., Siebesma, A. P.: The KNMI regional atmospheric climate model RACMO, version 2.1. Report no. 302. KNMI Technical Report. 2008.

Mikkelsen, P. S., Madsen, H., Rosbjerg, D. and Harremoes, P.: Properties of extreme point rainfall .3. Identification of spatial inter-site correlation structure. *Atmospheric Research,* 40(1), 77-98. doi:10.1016/0169-8095(95)00026-7. 1996.

Mikkelsen, P. S., Madsen, H., Arnbjerg-Nielsen, K., Jørgensen, H. K., Rosbjerg, D., and Harremoës, P.: A rationale for using local and regional point rainfall data for design and analysis of urban storm drainage systems. *Water Science and Technology*, 37(11), 7-14. 1998.

Molnar, P., and Burlando, P.: Variability in the scale properties of high-resolution precipitation data in the Alpine climate of Switzerland. *Water Resources Research*, 44(10), W10404. doi:10.1029/2007wr006142. 2008.

Nguyen, V.-T.-V., Nguyen, T-D. and Ashkar, F.: Regional frequency analysis of extreme rainfalls. *Water Science and Technology*, 45(2), 75-81. 2002.

Olsson, J., and Burlando, P.: Reproduction of temporal scaling by a rectangular pulses rainfall model. *Hydrological Processes*, 16(3), 611–630. doi:10.1002/hyp.307. 2002.

Olsson, J., Berggren, K., Olofsson, M., Viklander, M.: Applying climate model precipitation scenarios for urban hydrological assessment: a case study in Kalmar City, Sweden. *Atmospheric Research,* 92:364–375. doi:10.1016/j.atmosres.2009.01.015. 2009.

Roeckner, E., Bäuml, G., Bonaventura, L., Brokopf, R., Esch, M., Giorgetta, M., Hagemann, S., Kirchner, I., Kornblueh, L., Manzini, E., Rhodin, A., Schlese, U., Schulzweida, U. and Tompkins, A.: The atmospheric general circulation model ECHAM5: Model description. Max Planck Institute for Meteorology Rep. 349, 140 pp. 2003.

Rodriguez-Iturbe, I., Cox, D. R. and Isham, V.: Some models for rainfall based on stochastic point processes. *Proceedings of the Royal Society of London, Series* A 410, 269–288. doi: 10.1098/rspa.1987.0039. 1987a.

Rodriguez-Iturbe, I., Febres de Power, B. and Valdes, J. B.: Rectangular pulses point process models for rainfall: analysis of empirical data. *Journal of Geophysical Research*, 92(8), 9645–9656. doi: 10.1029/JD092iD08p09645. 1987b.

Scharling, M.: *klimagrid Danmark* nedbør 10∗10 km (ver.2) – metodebeskrivelse. Danish Meteorological Institute Technical report no 99-15. In Danish. 1999.

Scharling, M.: *Climate Grid Denmark*. Danish Meteorological Institute Technical report no 12-10. 2012.

Schilling, W.: Rainfall data for urban hydrology: what do we need? *Atmospheric Research* 27, 5–22. doi: 10.1016/0169-8095(91)90003-F. 1991.

Skamarock, W., Klemp, J., Dudhia, J., Gill, D. and Barker, D.:. A description of the Advanced Research WRF version 3. *NCAR Tech. Note NCAR/TN-475+ STR*, 113. 2005.

Sunyer, M. A., Gregersen, I. B., Rosbjerg, D., Madsen, H., Luchner, J., and Arnbjerg-Nielsen, K.: Comparison of different statistical downscaling methods to estimate changes in hourly extreme precipitation using RCM projections from ENSEMBLES. International Journal of Climatology. doi:10.1002/joc.4138. 2014.

Formatted: English (U.S.)

Sunyer, M. A., Madsen, H. and Ang, P. H.: A comparison of different regional climate models and statistical downscaling methods for extreme rainfall estimation under climate change. *Atmospheric Research,* 103. 129-128 doi:10.1016/j.atmosres.2011.06.011. 2012.

Sunyer, M. A., Madsen, H., Rosbjerg, D. and Arnbjerg-Nielsen, K.: A Bayesian Approach for Uncertainty Quantification of Extreme Precipitation Projections Including Climate Model Interdependency and Non-Stationary Bias. *Journal of Climate*, 27(18), 7113-7132 doi: 10.1175/JCLI-D-13-00589.1. 2014.

Sunyer, M. A., Sørup, H. J. D., Madsen, H., Rosbjerg, D., Christensen, O. B., Mikkelsen, P. S. and Arnbjerg-Nielsen K.: On the importance of observational data properties when assessing regional climate model performance of extreme precipitation. *Hydrological Earth System Science*. 17(11), 4323-4337. doi: 10.5194/hess-17-4323-2013. 2013.

Tebaldi, C., and Knutti, R.: The use of the multi-model ensemble in probabilistic climate projections. *Philosophical Transactions Series A, Mathematical, Physical, and Engineering Sciences,* 365, 2053-2075. doi: 10.1098/rsta.2007.2076. 2007.

van der Linden, P., Mitchell, J. F. B. (eds): ENSEMBLES: Climate Change and its Impacts: Summary of research and results from the ENSEMBLES project. Met Office Hadley Center, Exeter. 2009.

van Vuuren, D. P., Edmonton, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., Hurtt, G. C., Kram, T., Krey, V., Lamarque, J.-F., Masui, T., Meinshausen, M., Nakocenovic, N., Smith, S. J. and Rose, S. K.: The representative concentration pathways: an overview. *Climatic Change* 109(1-2), 5-31. doi: 10.1007/s10584-011-0148-z. 2011.

Verhoest, N. E. C., Vandenberghe, S., Cabus, P., Onof, C., Meca-Figueras, T. and Jameleddine, S.: Are Stochastic point rainfall models able to preserve extreme flood statistics? *Hydrological Processes* 24, 3439-3445. doi: 10.1002/hyp.7867. 2010.

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.

Waymire, E. and Gupta, V. K.: The mathematical structure of rainfall representations. I. A review of the stochastic rainfall models. *Water Resources Research*, 17(5), 1261-1272. doi:10.1029/WR017i005p01261. 1981.

Wilks, D. S.: *Statistical methods in the atmospheric sciences*. Academic Press, San Diego, CA, 3rd edition, 2011.

Formatted: Font: Italic

Wilks, D. S. and Wilby, R. L.: The Weather generator game: a review of stochastic weather models. *Progress in Physical Geography,* 23(3), 329-357. doi:10.1177/030913339902300302. 1999.

Willems, P., Arnbjerg-Nielsen, K., Olsson, J. and Nguyen, V.-T.-V.: Climate change impact assessment on urban rainfall extremes and urban drainage: methods and shortcomings. *Atmospheric Research,* 103. 106-118. doi:10.1016/j.atmosres.2011.04.003. 2012.

Wood, A. W., Leung, L. R., Sridhar, V. and Lettenmaier, D. P.: Hydrologic Implications of Dynamical and Statistical Approaches to Downscaling Climate Model Outputs. *Climatic Change*, 62(1-3) 189-216. doi:10.1023/B:CLIM.0000013685.99609.9e. 2004.

Table 1 Main characteristics of the two observational data sets used in this study.

	Type of data	Spatial data resolution	Temporal data resolution	Period
SVK	Point observations	60 stations	Minute data	1979-2012
CGD	Gridded data	10 km grid	Daily data	1989-2010

Table 2 Regional Climate Model (RCM) runs from which precipitation output is used to calculate perturbations schemes for the WG used in this study. All have a temporal resolution of 1 hour.

Name	RCM	GCM	Spatial resolution	Present period run	Future period run
HIRHAM SRES A1B	HIRHAM 5	ECHAM 5	25 km	1980-2009	2070-2099
RACMO SRES A1B	RACMO 2.1	ECHAM 5	25 km	1980-2009	2070-2099
HIRHAM rcp 4.5	HIRHAM 5	EC-EARTH	8 km	1981-2010	2071-2100
HIRHAM rcp 8.5	HIRHAM 5	EC-EARTH	8 km	1981-2010	2071-2100
WRF rcp 4.5	WRF 3	NorESM	8 km	1981-2010	2071-2100
WRF rcp 8.5	WRF 3	NorESM	8 km	1981-2010	2071-2100

Table 3 The relative weights used in the fitting procedure. *All the cross-correlations of a gauge have equal weights that sum up to the value shown.

Statistic	Relative weight
24 hour mean	1
24 hour variance	3
24 hour skewness	6
1 hour variance	3
1 hour skewness	6
1 hour auto-correlation	6
1 hour Cross-correlation	6*
Probability of dry day	1
Probability of dry hour	1

Table 4 Comparison between observational (SVK) data and the simulated (WGs) statistics. Data are averaged over the full course of the year and over the full model domain. For the SVK data set the 50th percentile is reported as well as the 16th to 84th percentiles interval to emulate the empirical standard deviation. For the WGs one central 50th percentile is reported across the ten simulations.

	24 hour mean (mm/day)	**24 hour variance (mm^2/day^2)**	**24 hour skewness (-)**	**1 hour variance ($mm^2/hour^2$)**	**1 hour skewness (-)**	**Probability of dry days (-)**	**Probability of dry hours (-)**	**Lag-1 hour auto-correlation (-)**
SVK (p50 (p16-p84))	1.67 (1.09-2.34)	12.6 (6.05-32.9)	3.56 (2.76-4.79	0.117 (0.0576-0.409)	8.93 (6.73-15.1)	0.718 (0.667-0.770)	0.934 (0.914-0.947)	0.572 (0.422 - 0.654)
WGs (p50)	1.60	14.9	4.04	0.151	10.4	0.812	0.945	0.578

Table 5 e-folding distances for the SVK and WG maximum averaged intensities of extremes for 1, 6, 12 and 24 hours duration, based on the fitted exponential models (cf. Figure 8) as well as for a regional climate model (HIRHAM/ECHAM) from the study by Gregersen et al. (2013) for comparison. *Values from Gregersen et al. (2013).

e-folding distance [km]	1 hour	6 hour	12 hour	24 hour
SVK	3.5	5.5	7.3	8.0
WGs	7.1 – 9.9	9.1 – 14	9.5 – 16	10 – 28
HIRHAM/ECHAM*	56	48	48	54

Table 6 e-folding distances for all aggregation periods for all WG output.

e-folding distance [km]	Aggregation period			
	1 hour	6 hour	12 hour	24 hour
WG – Present Climate	3.9	5.0	4.9	5.0
WG – HIRHAM SRES A1B	5.2	7.4	7.7	8.1
WG – RACMO SRES A1B	7.3	9.7	9.1	8.4
WG – HIRHAM rcp 4.5	5.2	8.4	8.7	8.8
WG – HIRHAM rcp 8.5	4.6	7.7	9.3	9.0
WG – WRF rcp 4.5	5.1	9.1	9.3	11.5
WG – WRF rcp 8.5	4.9	9.4	9.9	10.2

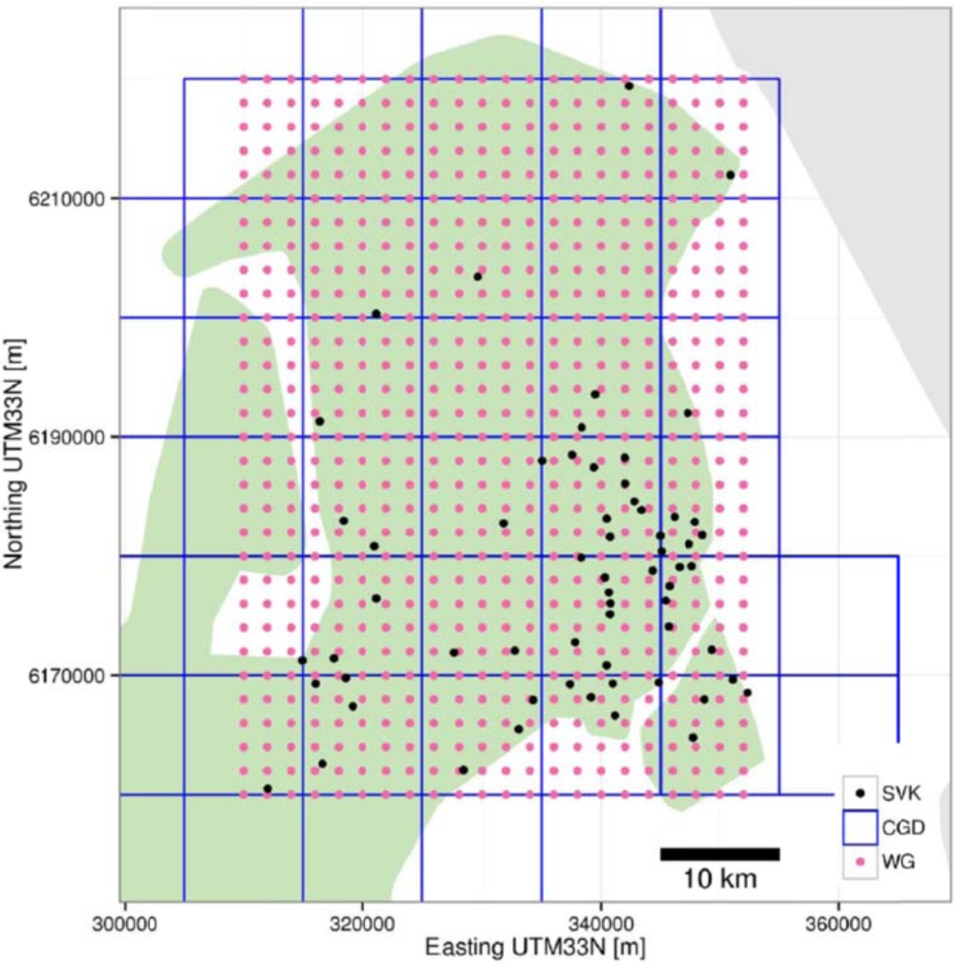


Figure 1 Locations of the rain gauges (SVK), the gridded data set (CGD) and extent of the modelled grid (WG) in the North-Eastern part of Zealand (Denmark) including Copenhagen in the South-Eastern part of the map where the concentration of SVK gauges is highest.

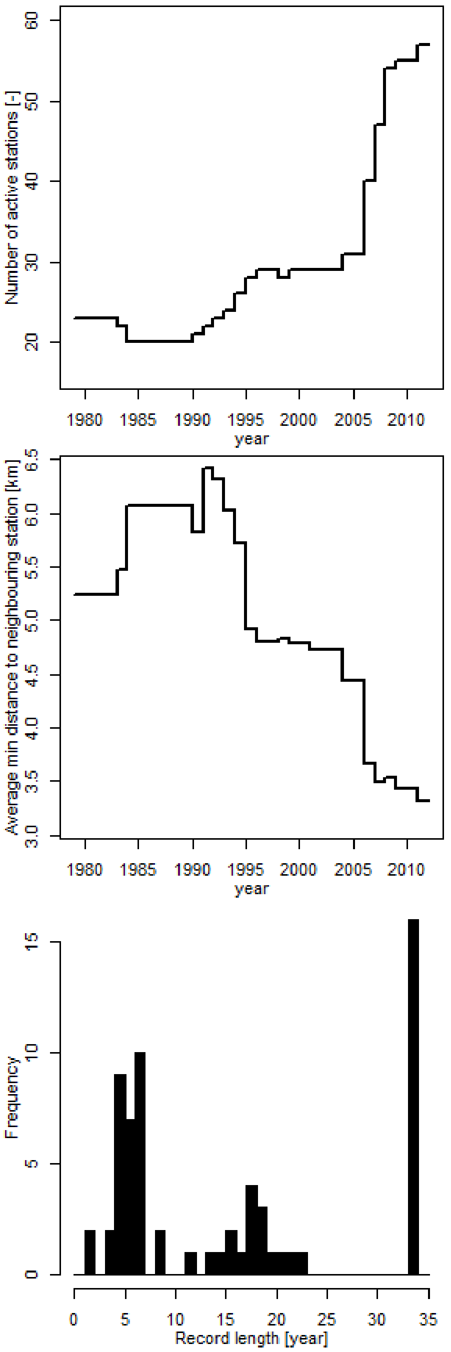


Figure 2 Temporal development in (top) the number of stations in the SVK data set and (middle) the average distance between closest neighbouring stations, and (bottom) the distribution of record lengths.

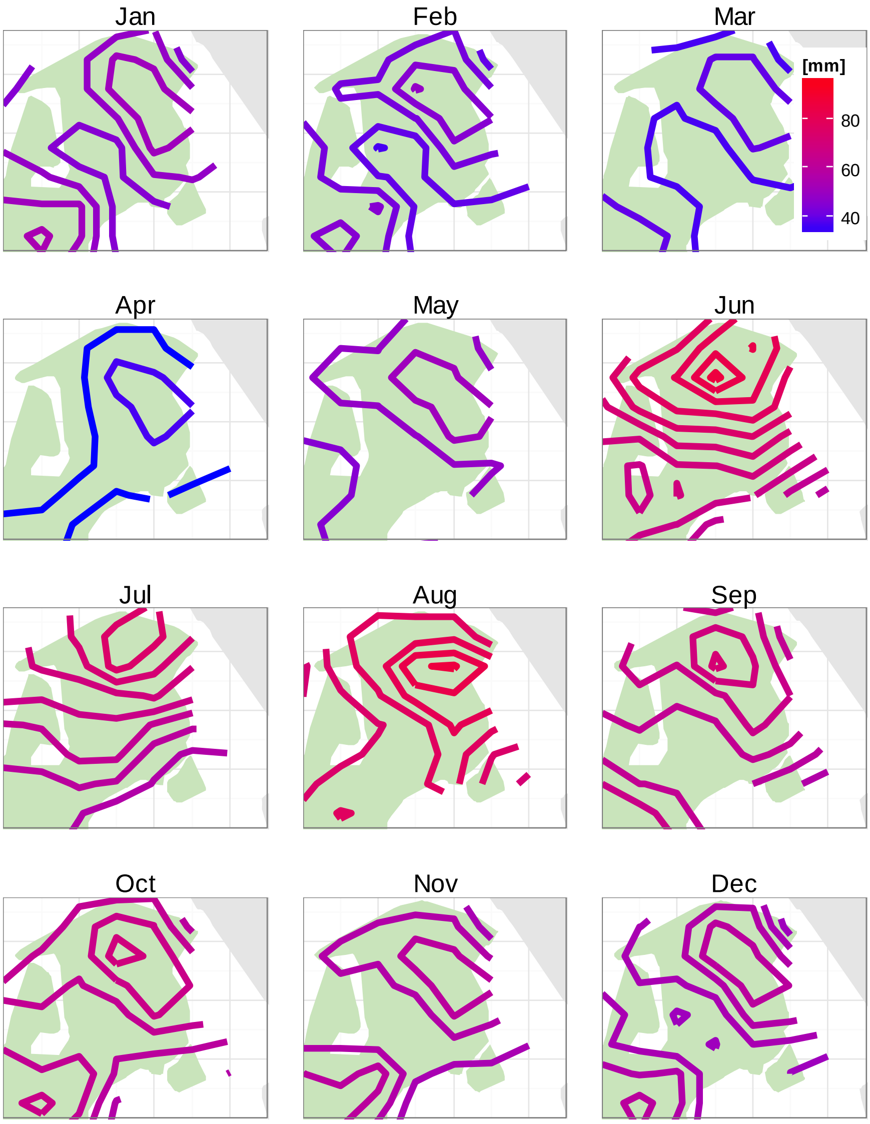
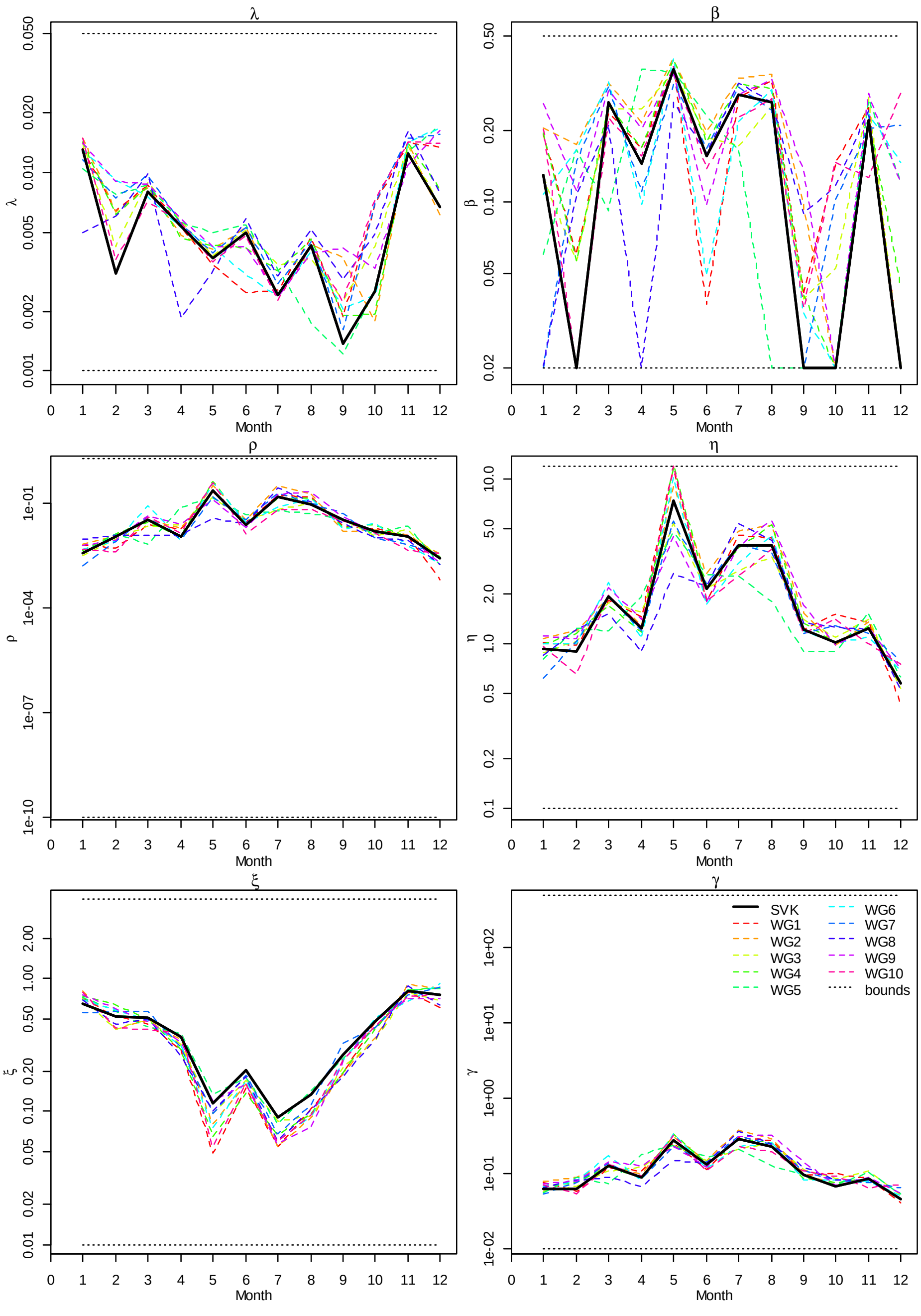


Figure 3 Spatial variation of the mean monthly precipitation calculated from the CGD data set for the model area. Isohyets are 3 mm between.



1

2 Figure 4 Monthly variation of the model parameters estimated from the SVK data set and from
3 the simulated 10 WG data sets. Upper and lower fitting bounds are shown in light grey.

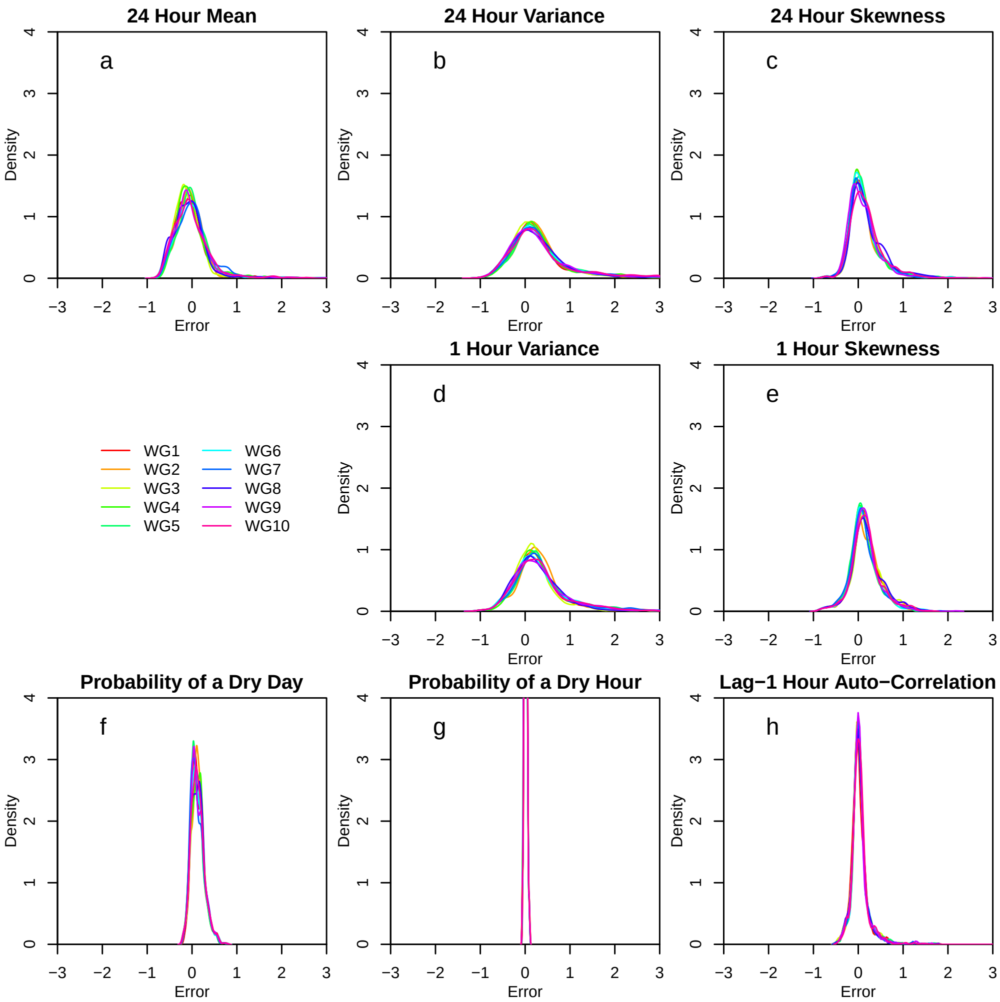


Figure 5 Density plots for the normalized error between the WG and the SVK data sets.

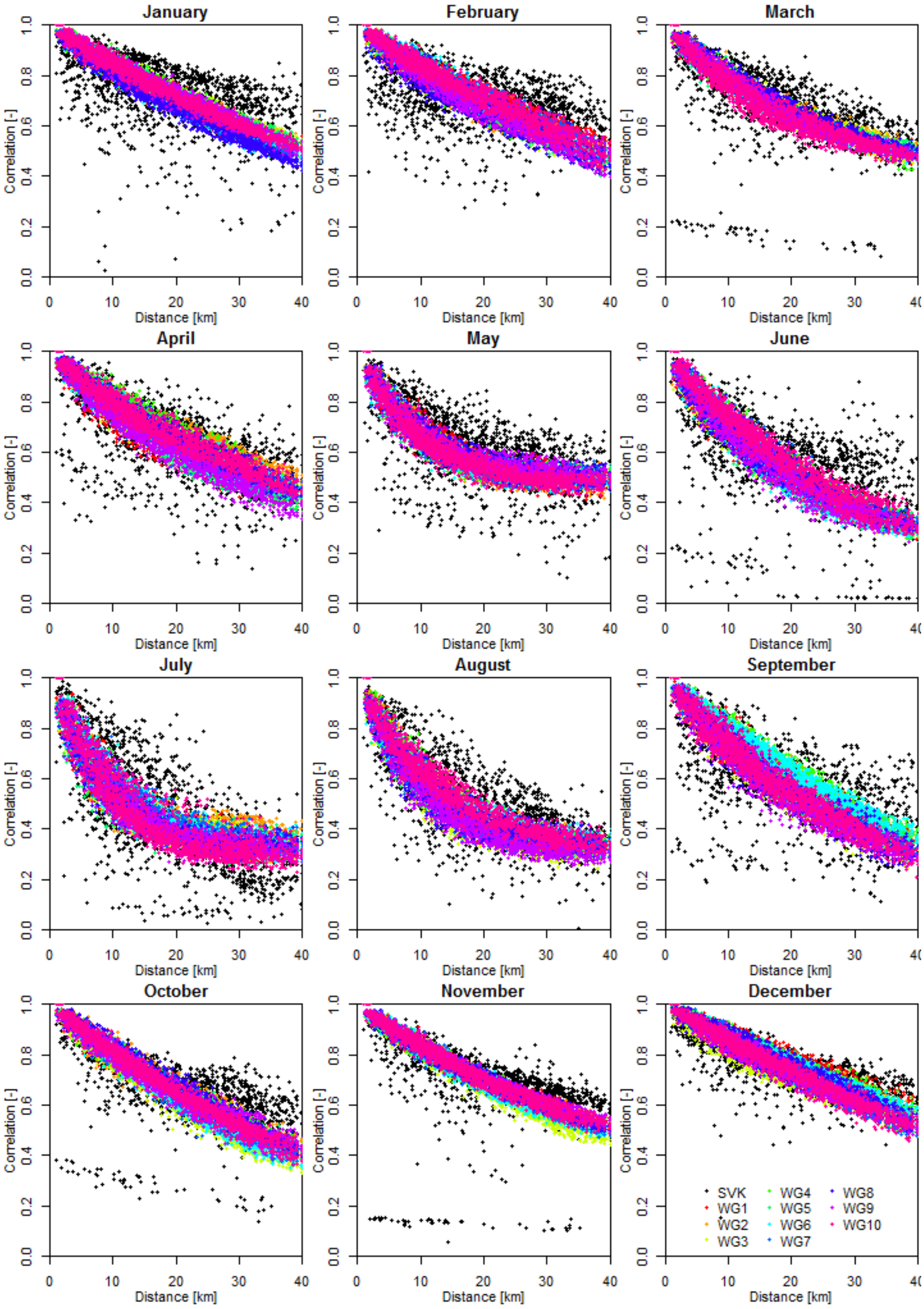


Figure 6 Variation of cross-correlation of the 1 hour intensity with distance between pairs of gauges in the SVK data set (black dots) and grid points in the WG data set (coloured dots).

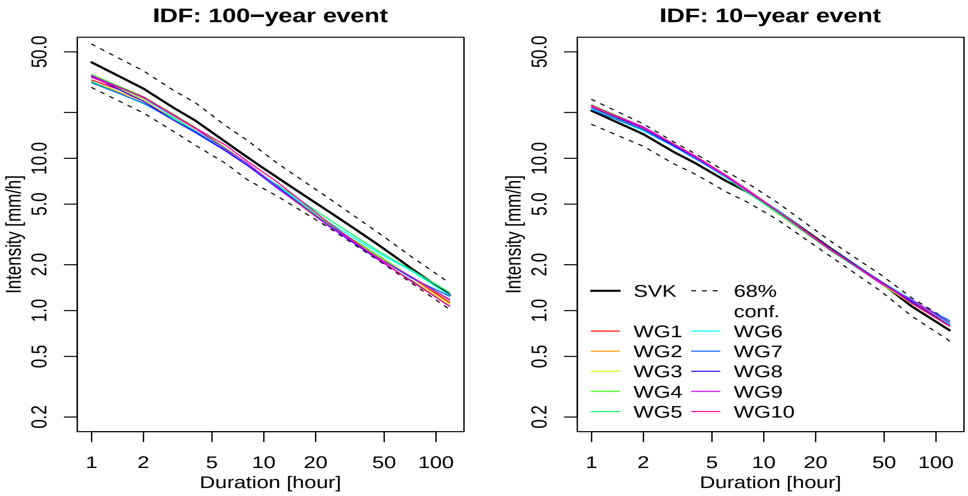


Figure 7 Mean intensity-duration-frequency curves for 100 and 10 year return periods calculated from the SVK data set and for all 10 WG realisations. 68% confidence interval for the SVK data set.

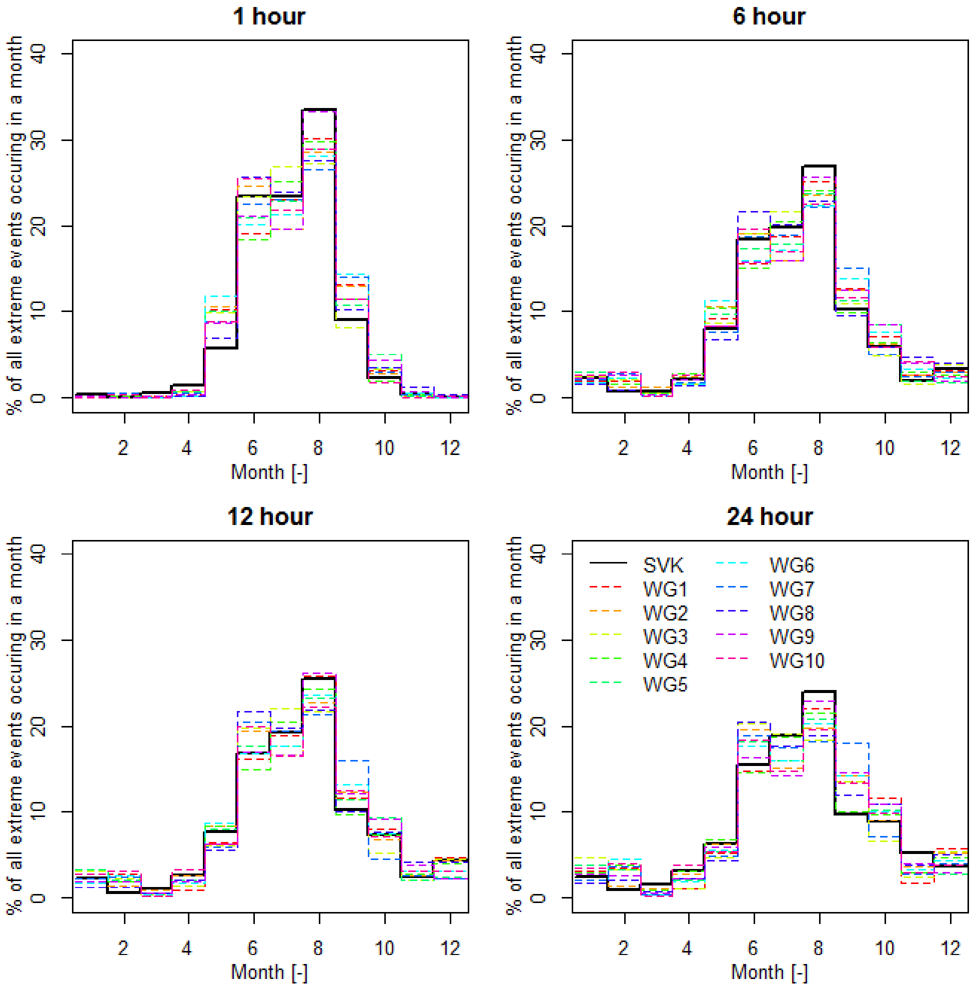


Figure 8 Monthly variation for 1, 6, 12 and 24-hour durations of the frequency of extreme events in the SVK and WG data sets.

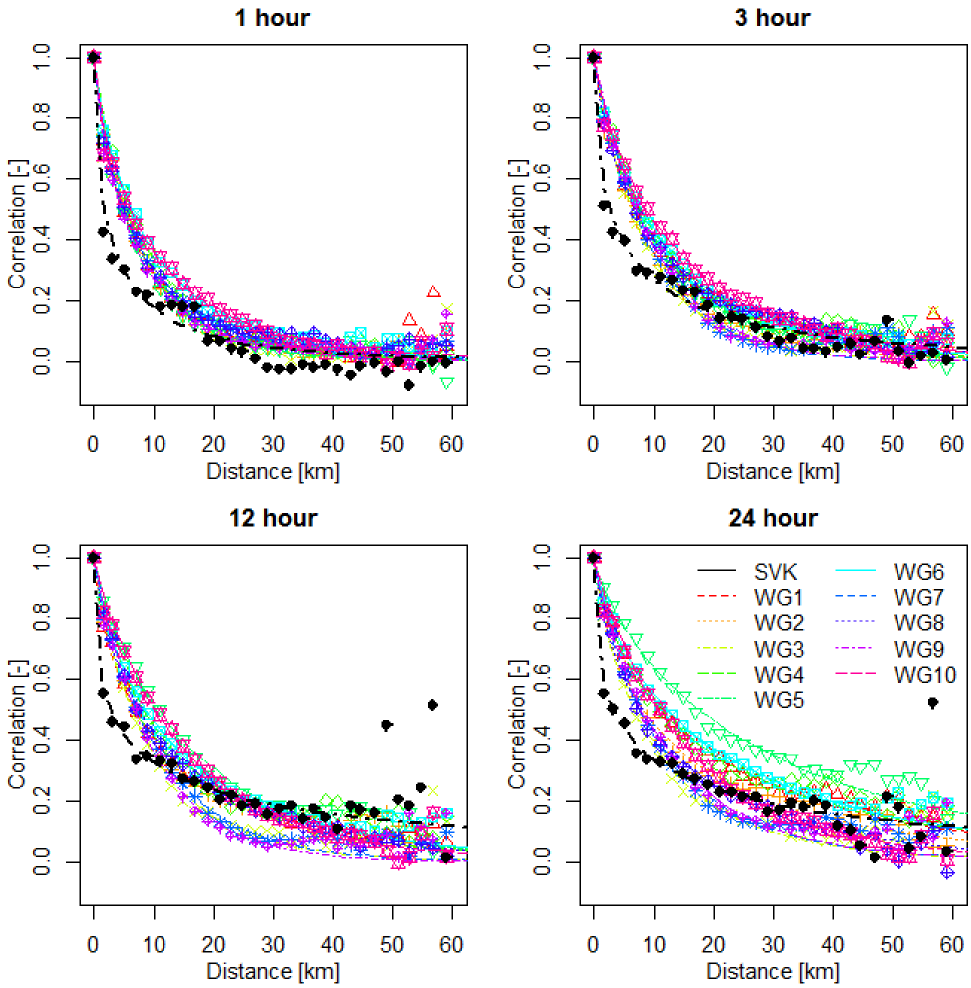


Figure 9 Unconditional spatial correlation for the SVK and WG data sets, calculated from maximum averaged intensities of extreme events for 1, 6, 12 and 24 hours duration. Fitted exponential models that highlight overall tendencies are shown.

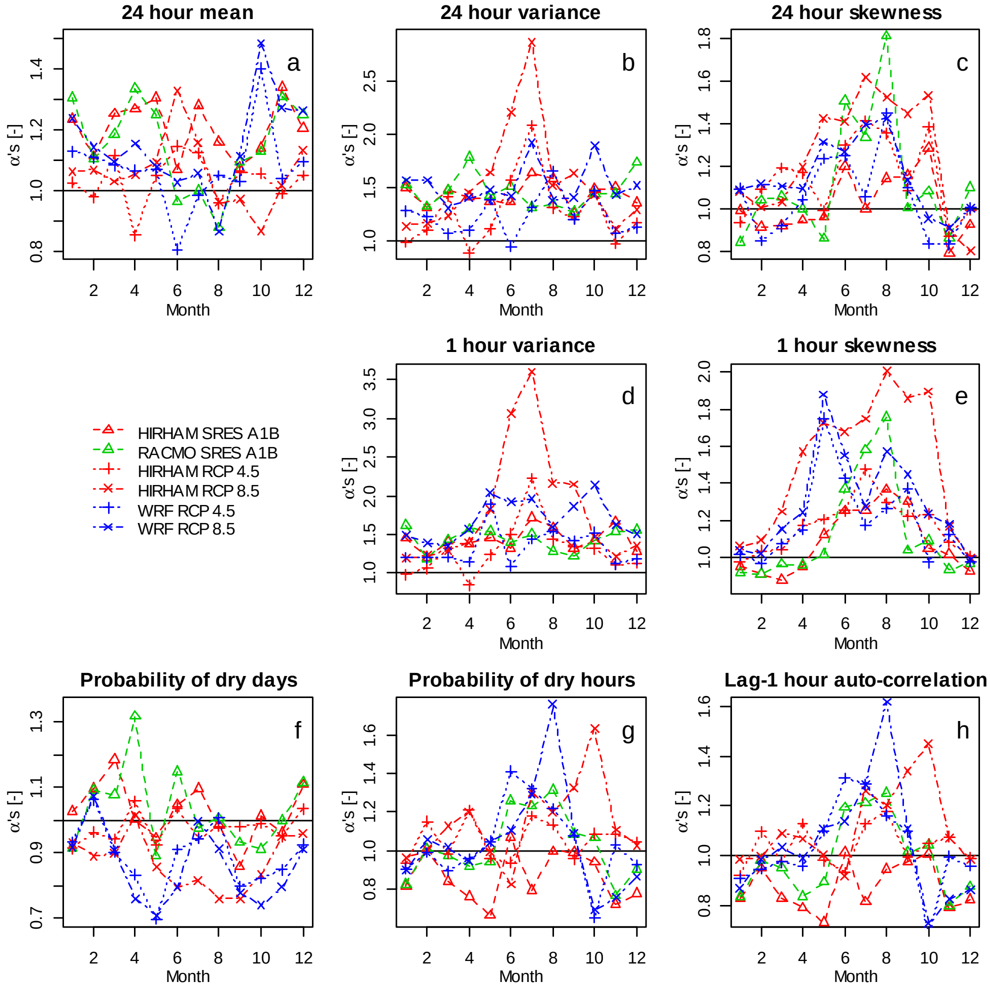


Figure 10 Change factors, α's, calculated on a monthly basis for each statistic and each RCM. Each set of α's from an RCM act as a perturbation scheme for the WG.

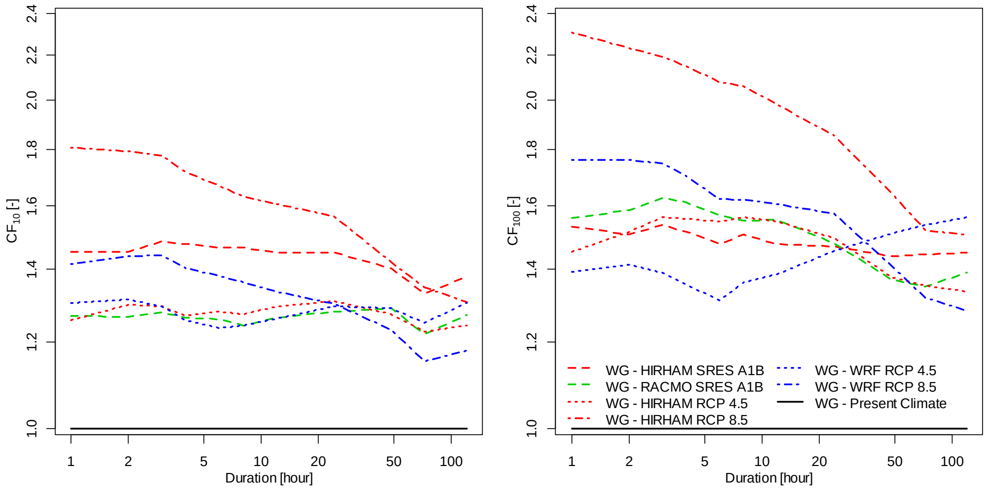
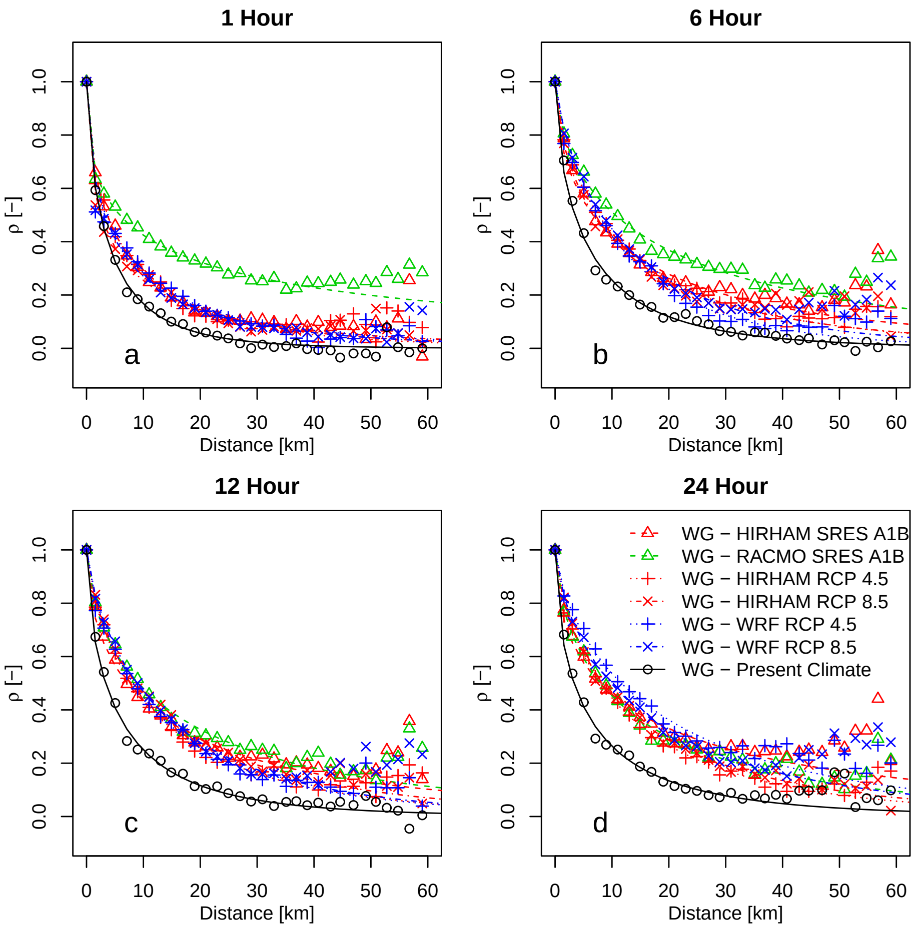


Figure 11 Climate factors for different return periods for the different perturbed WG runs. *T*=10 years (left) and *T*=100 years (right).



1

2 Figure 12 The unconditional spatial correlation of all *T*-year events for perturbed WG output for
3 event durations of 1, 6, 12 and 24 hours.