3^{rd} SWITCH Scientific Meeting
Belo Horizonte, Bresil
29 November – 4 December 2008

Predicting rainfall for the city of the future

Xavier Beuchat†, Marc Soutter

Ecohydrology Laboratory
Swiss Federal Institute of Technology (EPFL)
Lausanne, Switzerland

Abstract
There is now considerable evidence of global climate change. As these changes are likely to have dramatic impacts on hydrology, a great effort is required to provide future scenarios for local hydrological impact assessment studies. However, lack of knowledge and the complexity of the physical processes involved make these predictions highly uncertain. A probabilistic modelling framework is therefore essential.

Among others, urban hydrology will probably be perturbed by global change. In particular, existing classical stormwater drainage infrastructures may no longer work properly in the future. To cope with uncertain rainfall scenarios, structural best management practices (BMPs) or sustainable urban drainage systems (SUDs) may offer a robust and flexible alternative.

Here we propose and illustrate a methodology for producing probabilistic rainfall scenarios for the city of the future by downscaling coarse global climate model prediction to finer temporal and spatial scales. Availability of these scenarios should then allow an investigation of the potential impacts of climate change on urban stormwater managment strategies and the ability to test different alternatives to mitigate these impacts.

Keywords: Climate Change, Downscaling, Precipitations, Urban Drainage, Sustainable Development, Uncertainty

1 Introduction

SWITCH is a project that aims at catalysing change towards more sustainable urban water management in the "City of the Future". It seeks also to be as integrated as possible by comprehending all drivers related to water management.

When thinking in terms of sustainable development, one key step is to identify pressures that are acting on the system under study *now* and how we can control these pressures so as that their impacts on the quality of the system remain limited trough time. Among the spectrum of pressures one may identify, some will be more difficult to control than others and hence their evolution in the futur will be hard to assess with certainty.

† Corresponding Author: `xavier.beuchat@epfl.ch`

Therefore, when planning for the "City of the Future", one must take these uncertain aspects into account. There are essentially two ways to achieve this: either one has to take decisions that are *robust* [1], or one has to be prudent and adopt *flexible* and *adaptative* [2] strategies.

In particular, when planning stormwater-related infrastructures, we may think of urbanization, land use change, economics, energy demand, climatic conditions as being some pressures acting on the system. Among these, climate change may have uncertain impacts that should be considered seriously.

Classical stormwater drainage infrastructures (pipes and sewers) are perfect examples of infrastructures that are neither robust (pipes are deigned to accept a limited volume of water) nor flexible (re-dimensioning the sewer system is not easy and very expensive). More flexible and adaptative solutions may be found trough the concept of Sustainable Urban Drainage systems (SUDs). SUDs replicate natural drainage, they aim to prevent pollution, control flooding, recharge groundwater and enhance the environment. Intense research is currently being done to assess the performance of various SUDs, however there have only been limited studies addressing their usefulness to mitigate climate change impacts in urban areas (e.g. *Denault et al.* (2006); *Semadeni-Davies et al.* (2008)).

2 Climate change: global predictions and downscaling

The last release of the Intergovernmental Panel on Climate Change assessment report (*IPCC Working Group I* (2007)) demonstrates that climate is globally warming. According to this document, most of the observed increase in global average temperatures since the mid-20th century is very likely due to the observed increase in anthropogenic greenhouse gas concentrations.

Along with this global warming there is evidence for intensification of the global water cycle *Huntington* (2006); *Waliser et al.* (2007) which may have serious impacts on hydrology. Climate change impacts should therefore be assessed for long term hydrological studies.

To predict the magnitude of global warming in the future, IPCC developed a set of 40 Green-House Gas (GHG) emission scenarios corresponding respectively to various demographic, socio-economic and technological development alternatives. These can be found in the Special Report on Emissions Scenarios (SRES) (*IPCC Working Group III* (2000)) and are thought to cover the full range of uncertainties related to future evolution.

Global Climate Models (GCMs) are numerical coupled models that allow represention of the atmospheric system of the earth. These models can be forced by selected SRES scenarios and climate perturbations can then be predicted. The main disadvantages of GCMs are the following: i) they are deterministic (can't handle uncertainties) and ii) due to computational cost, their spatial resolution is typically 2.5° latitude by 3.75° longitude.

The coarse resolution of GCMs implies that subgrid processes such as clouds formation, land use and topography are not resolved and therefore prevents direct use of their results for hydrological impacts studies. Depending on the phenomena we want to model, rainfall temporal resolution for urban impact studies may typically be 10 minutes or less (*Hingray and Ben Haha* (2005); *Segond et al.* (2007)). Therefore the rough results provided by GCMs have to be downscaled, either dynamically or statistically, to a finer scale suitable for hydrological impact studies (see *Fowler et al.* (2007) for a review).

However, to date, very few studies have attempted to develop an integrated framework for producing probabilistic scenarios suitable for hydrological impact studies.

[1] Robustness is the quality of being able to withstand stresses, pressures, or changes in procedure or circumstance. A system, organism or design may be said to be "robust" if it is capable of coping well with variations (sometimes unpredictable variations) in its operating environment with minimal damage, alteration or loss of functionality.

[2] Adaptability is to be understood here as the ability of a system to adapt itself efficiently and fast to changed circumstances.

3 Modelling strategy: theoretical considerations

The approach chosen here to produce probabilistic rainfall scenarios for the city of the future is based on a methodology developed by *Chandler et al.* (2007) and is very briefly outlined below.

Generalized Linear Models (GLMs) are used to downscale rainfall at daily time step from GCMs outputs. This is achieved by using the software package GLIMCLIM (*Chandler and Wheater* (2002)) which is based on generalized linear models and aimed primarily at modelling and simulation of daily rainfall time series at a single or multiple sites. Basically, GLMs derive relationships between a variable of interest, $\mathbf{Y} = (Y_1, \ldots, Y_n)^t$, called the *predictand* and a set of p temporally varying predictor variables, or *covariates*, whose values can be arranged in a $n \times p$ matrix $\mathbf{X}$. The relationships between the predictand and the predictors are assumed to be given by

$$g(\mu) = \mathbf{X}\beta \qquad ; \qquad \mu = \mathbf{E}(\mathbf{Y}) \tag{1}$$

where g is some monotonic differentiable *link function* and β is the vector of coefficients that has to be estimated.
Moreover, the distribution of each Y_i is assumed to belong to the exponential family. That is

$$f_Y(y; \theta, \phi) = \exp\left[\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\right] \tag{2}$$

where a, b and c are some specific functions and ϕ is called the *dispersion* parameter.

In our case, the predictand is rainfall series. Among predictors that are used, we may mention terms relating to seasonality, terms relating to the series itself (autocorrelations, intensity of rain events...) and external covariates such as coarse-scale atmospheric variables.

Models are fitted using historical data (rainfall and large scale atmospheric variables derived from NCEP reanalysis) to describe relationships. Then, for future scenarios, one may simulate from the fitted models driven by GCM-generated atmospheric variables.

Rainfall series are highly nonstationnary. Simulating nonstationary rainfall sequences is achieved by allowing some predictors to modulate the effect of other predictors incorporated via interactions (alternative to, e.g., fitting separate models in each month of the year).

Actually, without going into details, the GLIMCLIM modelling framework is composed of two components: first, an *occurrence model* based on logistic regression is used to model the pattern of wet and dry days, and second, an *amount model* based on gamma distributions *with a common dispersion parameter* allows the simulation of rainfall amounts on wet days.

To fully explain the notion of the common dispersion parameter in the amount model, it needs to be mentionned that, traditionally, GLMs assume that the relationship between the means and the variances is given by

$$\mathrm{var}(Y_i) = \phi\nu(\mu_i) \tag{3}$$

where ν is some scalar, non negative, *variance function* determined by the probability distribution of Y_i and ϕ is the unknown, but *fixed*, dispersion parameter.

Yet, it has been noted that this assumption of a common shape parameter in the amount model may be doubtful and lead to an underestimation of rainfall extremes in summer (*Yang et al.* (2005); *Chandler et al.* (2007)).

To cope with this limitation of GLIMCLIM, the software is being modified by implementing ideas found in *Smyth* (1989). Indeed, *Smyth* (1989) show how classical generalized linear models can be extended to allow non-homogeneous dispersion parameter, which can be modelled in terms of covariates. Mean and dispersion sub-models are formulated for this; the dependent variable for the dispersion sub-model being the deviance componants of the mean sub-model. Therefore, we generalize (3) to

$$\sigma_i^2 = \phi_i\nu(\mu_i) \tag{4}$$

and assumes that the dispersions ϕ_i can be modelled by

$$h(\phi_i) = \omega_i^T \gamma \tag{5}$$

where h is another link function, ω_i is a vector of covariates and γ is another vector of unknown parameters.

The choice of climate models (GCMs) represents a significant source of uncertainty in future scenarios. A possible approach to take this into account is to use a mixture of simulations driven by different climate models. Hence in this work, a set of approximately 20 GCMs simulations is considered for each of three SRES scenarios.[3].

Once daily series of rainfall have been downscaled using GLMs, they may then be further disagregated to hourly and even sub-hourly data. Disagregation to hourly data is generally achieved via methods based on Poisson cluster models (see e.g.*Onof et al.* (2000); *Koutsoyiannis and Onof* (2001)), while sub-hourly series are obtained using methods based on cascades (see e.g. *Onof et al.* (2005)).

4 Case study: Geneva, Switzerland

4.1 Data

Daily rainfall data of 4 stations located near the city of Geneva in Switzerland have been used to illustrate the first part of the methodology (i.e. the downscaling at a daily time step).

Figure 1: Orthophoto of the region of Geneva showing the stations used in the study. (1) Genève-Cointrin; (2) Genève-Aïre; (3) Croix-de-Rozon; (4) Jussy. Source: SITG.

[3]This approach is simple, but underestimates the true uncertainty (other climate models could yield more extreme projections). In order to tackle this problem, *Chandler et al.* (2007) proposed a very promising pilot methodology that enables to generate easily a large set of pseudo GCMs covariates that are consistent with available GCMs outputs.

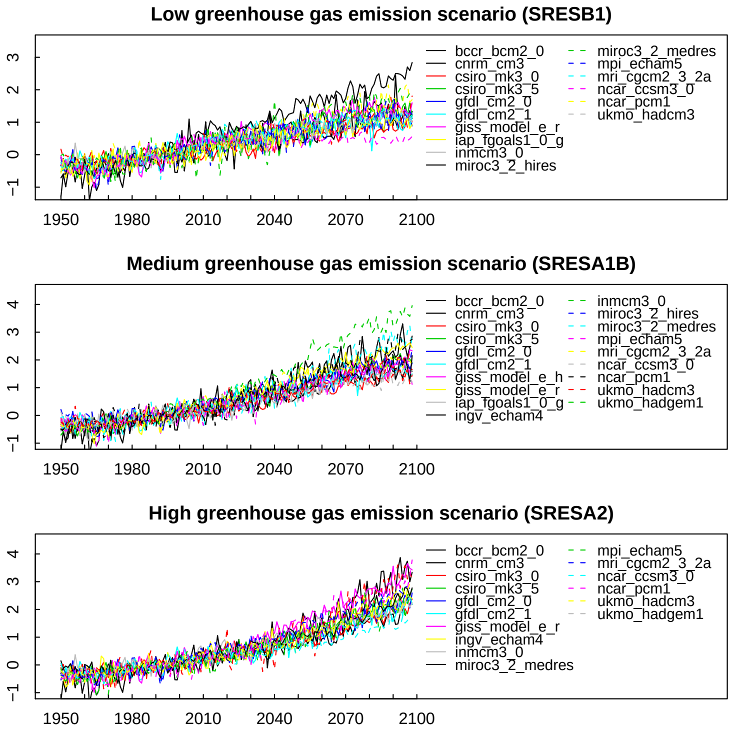

Figure 2: Predictions of annual mean air temperature for the city of Geneva. Values are standarized by their overall mean and variance during the period 1982-2007. Each graphics corresponds to a different greenhouse gas emission scenario and each line correspends to a specific GCM simulation.

The locations of these stations are shown in Figure 1 and some of their specifications are described in Table 1. Data from all stations during period 1982-2007 were used to calibrate the modified version of GLIMCLIM, while data from station *Genève-Cointrin* during the period 1956-1981 were used to validate the fitted model

Station	Genève-Cointrin	Genève-Aïre	Croix-de-Rozon	Jussy
Northings [km]	508 810	498 580	499 480	495 800
Eastings [km]	120 310	122 320	111 080	116 900
Altitude [m]	420	375	478	465
Data availability	1900–...	1968–...	1974–2005	1900–...
Type	Auto. since 1980; TB before	TB	TB	TB

Table 1: Some specifications of the stations used the study (coordinates are given along the swiss grid; TB=Tipping Bucket). Source: Météosuisse.

As described in Section 3, the objective is to include large scale atmospheric predictors as predictors in GLMs. *Chandler et al.* (2007) recommend to use the monthly means of: surface temperatures, relative humidity [4] and sea level pressure. Such observed variables are available through NCEP Reanalysis archives. NCEP Reanalysis datasets are derived by feeding quality controlled observations into a physical model, which then output gridded values of several variables. NCEP Reanalysis for the period 1948-2007 were provided by the NOAA/OAR/ESRL PSD, Boulder,

[4] *Chandler et al.* (2007) use relative humidity at surface, but I found that relative humidity at pressure level of 600 hPa gave better results.

Colorado, USA, from their Web site at `http://www.cdc.noaa.gov/`. GCMs predictions for the same variables and for the period 1950-2100 where collected from the "World Climate Research Programme's (WCRP's) Coupled Model Intercomparison Project phase 3 (CMIP3) multi-model dataset". Three greenhouse gas emission scenarios were considered (SRESB1: low, SRESA1B: medium, SRESA2: high). For each scenario around 15-20 GCM simulations were available. All NCEP and GCMs archives were regridded into a common grid (the one of model HADCM3), and, following the procedure explained in *Leith* (2006), the time series for each of the three variables of interest were extracted as a weighted sum of neighbouring grid cell. Finally, these time series were standardised with respect to the overall mean and standard deviation of the fitting period [5] (1982-2007). Figure 2 illustrates the evolution of standardised temperatures for the city of Geneva as simulated by different GCMs.

4.2 Results of the downscaling at a daily time step

Models were fitted using observed rainfall series and NCEP predictors during the period 1982-2007. The details of the fitting of the three GLMs (one GLM for the occurence model, and two interconnected GLMs for the amount model) are not described in detail here. Instead we move directly to the simultation results.

Fitting and simulations were carried out simultaneously for the all the four sites. Figure 3 depicts the monthly summary statistics of 200 simulations [6] of the fitted models for the station *Genève-Cointrin*: we can see that all monthly statistics are well reproduced. Similar results are obtained at the three other sites and for the overall area. The period 1956-1981 has been used for validation; that is, the fitted models were constrained with atmospheric predictors corresponding to this time gap. Simulation results show that the monthly statistics are still well reproduced (results not shown). This is encouraging and indicates that GLIMCLIM is able to produce sensible results outside of the fitting period.

Figure 4 displays annual (first row) and seasonal (rows 2 and 3) rainfall totals corresponding to various simulations. Plots in the first column show the statistics of 200 simulations of the fitted models conditional on NCEP atmospheric predictors for the station *Genève-Cointrin*. Despite the fact that these statistics are not used to fit the model, they are reasonably well reproduced.

Columns 2 to 4 show the same statistics but for the period 2072-2098 and under, respectively, scenarios SRESA1B (medium forcing), SRESA2 (high forcing) and SRESB1 (low forcing). For each SRES scenario and for each GCM available, a set of 50 simulations has been carried out. The summary statistics displayed in columns 2 to 4 correspond to the mixture of all available simulations for the given SRES scenario.

If we compare future predictions (columns 2 to 4) and simulations during the fitting period (column 1), we may conclude that no great changes are going to happen concerning annual and seasonal total rainfall. However, the range of the simulated total rainfall is wider for the period 2072-2098, reflecting the fact that GCMs do not totally agree. Another conclusion that can be drawn is that the higher the forcing is, the wider the uncertainty is.

Results of monthly summary statistics for the end of 21th century are not shown, but, on average, simulations reflect more intense rain events in summer along with a greater variability of rainfall during the summer months.

An analysis of extreme values has also been carried out using the technique known as *peak over threshold*. Following *Coles* (2001) (cf chapter 4), a generalized Pareto distribution has been

[5] *Chandler et al.* (2007) do this standardisation by month.

[6] There are 200 simulations of 26 years length, that is 5200 simulated years altogether.

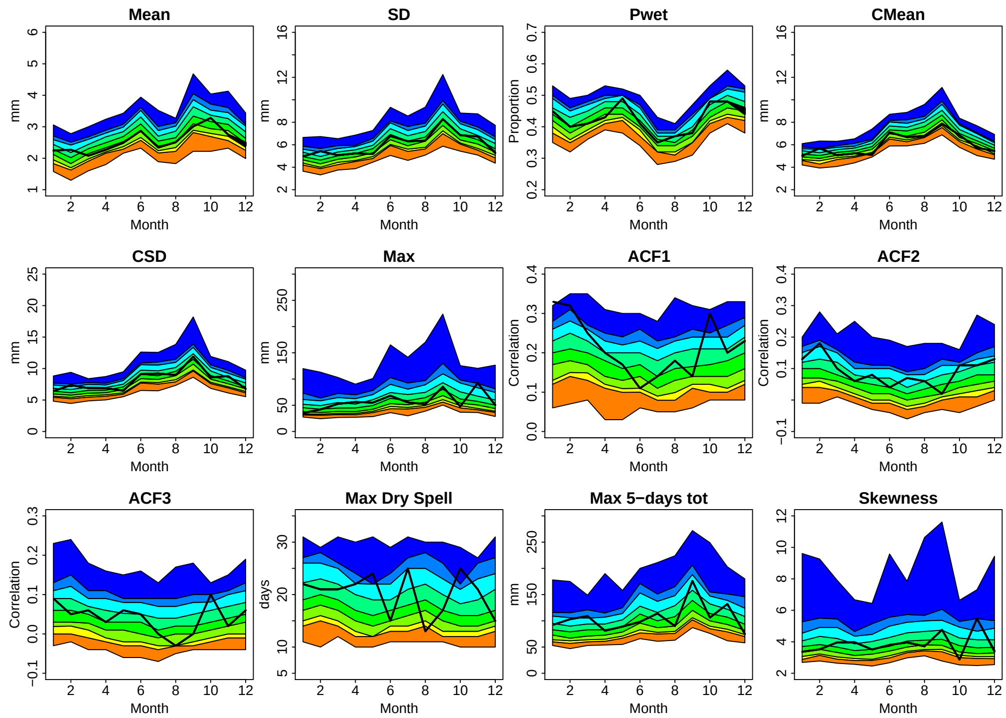

Figure 3: Summary monthly statistics of 200 simulations for the site *Genève-Cointrin* during the fitting period (from top-left to bottom-right: mean rainfall, standard deviation, proportion of wet days, mean rainfall on wet days only, standard deviation on wet days only, maximum rainfall, autocorrelation lag 1 to 3, maximum length of dry spell, maximum cumulative 5-days rainfalls, skewness). Solid black line represents observed values and coloration shows the range of the simulated distributions along with deciles.

fitted to daily rainfall values of the station *Genève-Cointrin* exceeding a threshold of 25 mm during the period 1956-2007. Diagnostic plots of the fitted distribution are displayed on the right of the Figure 5 and indicate a correct fit. The graphics on the left of Figure 5 indicate that the 100-year return level is estimated at 106 mm (with a 95% confidence interval being [84.5mm,152mm]).

In order to assess the performance of the GLMs simulations, in terms of extreme values, a frequency analysis has been carried out on the annual maxima of the 200 simulations made using NCEP predictors during the fitting period (5200 maxima altogether). Given these simulations, the 100-year return level is 108.86 mm. This is fairly close to the value of 106 mm estimated by the generalized Pareto distribution fitted to the observed rainfall series.

The same exercise has been accomplished for each GCM-based simulation during the fitting period and during the forecasting period. As around 20 GCMs are available per SRES scenario, we have, for each SRES scenario, a set of about 20 estimated values of 100-year return levels. Figure 6 displays, for each SRES scenario and for each period of interest, boxplots of the distributions of these sets of estimated return values. The horizontal red line indicates the 100-year return level estimated trough the NCEP-based simulation. The three boxplots on the right concern the fitting period: in terms of 100-year return values, NCEP-based and GCM-based simulations are in accordance regardless of the SRES scenario considered. The three remaining boxplots on the right relate to the forecasting period (2072-2098). Here three conclusions can be drawn. First, independently of the SRES scenario considered, the average forecasted 100-year return level is higher than for the fitting period. Second, the higher the forcing is, the higher the average 100-year return level is (SRESB1: 117.71 mm, SRESA1B: 124.40 mm, SRESA2: 133.97 mm). Finally, once again, the higher the forcing is, the more variable the predictions are.

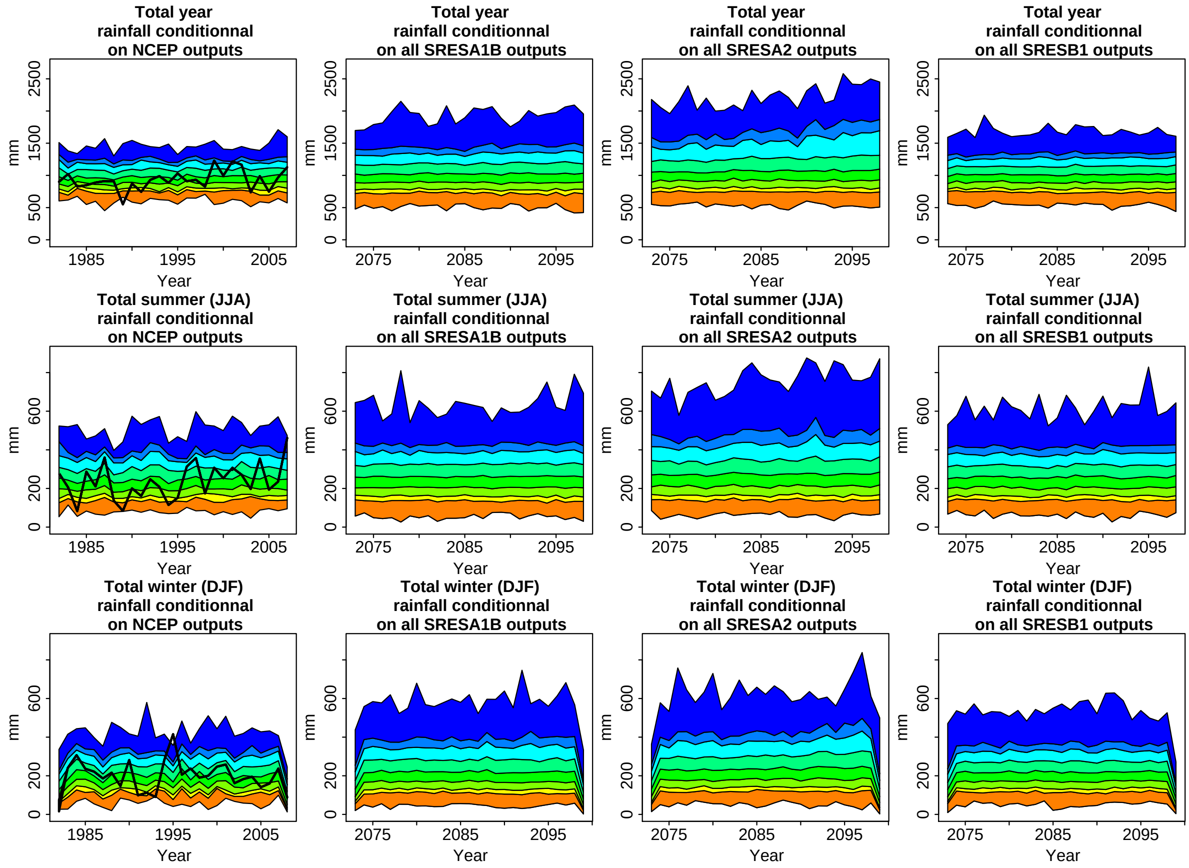

Figure 4: **First row:** annual total rainfall simulated by the models. **Second and third rows:** seasonal totals. Each column corresponds to a different set of simulations. **First column:** 200 NCEP-based simulations. **Remaining columns:** ensemble of GCM-based simulations for the futur period (50 simulations per GCM), each column corresponds to a given SRES scenario. Solid black lines (first column only) represent observed values and coloration shows the range of the simulated distributions along with deciles.

Acknowledgements

We would like to thank Richard Chandler for providing GLIMCLIM software along with several plotting routines and also for providing useful advices.

We acknowledge the modeling groups, the Program for Climate Model Diagnosis and Intercomparison (PCMDI) and the WCRP's Working Group on Coupled Modelling (WGCM) for their roles in making available the WCRP CMIP3 multi-model dataset. Support of this dataset is provided by the Office of Science, U.S. Department of Energy.

Finally, rainfall data have been provided by MeteoSwiss, the Swiss Federal Office of Meteorology and Climatology.

References

Chandler, R., V. Isham, H. Wheater, C. Onof, N. Leith, A. Frost, and M.-L. Segond (2007), Spatial-temporal rainfall modelling with climate change scenarios, *Technical Report FD2113*, Imperial College and University College London, London.

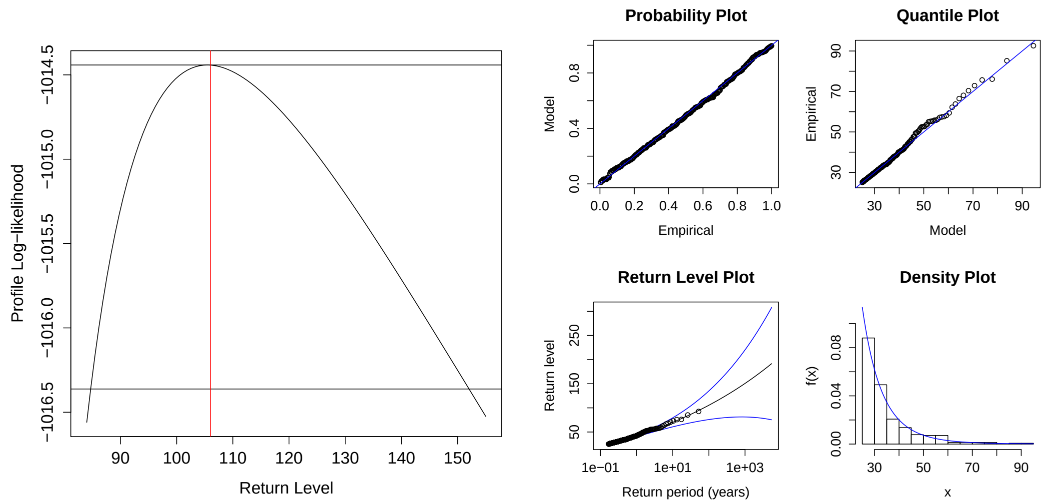

Figure 5: **Left plot:** profile likelihood for the 100-year level in threshold excess model of daily rainfall data from station *Genève-Cointrin*. **Right plots:** Diagnostic plots for the threshold excess model of daily rainfall data from station *Genève-Cointrin*.

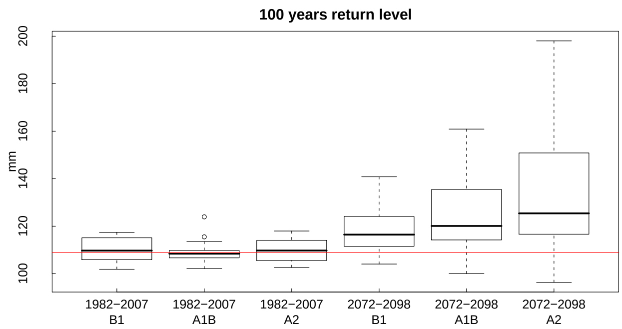

Figure 6: Graphics on the first column shows respectively total Return values.Summary statistics of 100 simulations for the station *Genève-Cointrin* during the fitting period using varying dispersions.

Chandler, R. E., and H. S. Wheater (2002), Analysis of rainfall variability using generalized linear models: A case study from the west of Ireland, *Water Resources Research*, *38*(10).

Coles, S. (2001), *An introduction to statistical modeling of extreme values*, Springer, London.

Denault, C., R. G. Millar, and B. J. Lence (2006), Assessment of possible impacts of climate change in an urban catchment, *Journal of the American Water Resources Association*, *42*(3), 685–697.

Fowler, H., S. Blenkinsop, and C. Tebaldi (2007), Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling, *International Journal of Climatology*, *27*(12), 1547–1578, 10.1002/joc.1556.

Hingray, B., and M. Ben Haha (2005), Statistical performances of various deterministic and stochastic models for rainfall series disaggregation, *Atmospheric Research*, *77*(1-4), 152–175.

Huntington, T. G. (2006), Evidence for intensification of the global water cycle: Review and synthesis, *Journal of Hydrology*, *319*(1-4), 83–95.

IPCC Working Group I (2007), *Climate Change 2007 - The Physical Science Basis: Working Group I Contribution to the Fourth Assessment Report of the IPCC (Climate Change 2007)*, Cambridge University Press.

IPCC Working Group III (2000), *Special report on emissions scenarios*, Cambridge University Press, Cambridge.

Koutsoyiannis, D., and C. Onof (2001), Rainfall disaggregation using adjusting procedures on a poisson cluster model, *Journal of Hydrology*, *246*(1-4), 109–122, times Cited: 15.

Leith, N. A. (2006), Using generalised linear models to simulate daily rainfall under scenarios of climate change, *Tech. rep.*, UCL, available at http://www.ucl.ac.uk/stats/research/Rainfall/reports.html.

Onof, C., R. E. Chandler, A. Kakou, P. Northrop, H. S. Wheater, and V. Isham (2000), Rainfall modelling using poisson-cluster processes: a review of developments, *Stochastic Environmental Research and Risk Assessment*, *14*(6), 384–411, times Cited: 19.

Onof, C., J. Townend, and R. Kee (2005), Comparison of two hourly to 5-min rainfall disaggregators, *Atmospheric Research*, *77*(1-4), 176–187.

Segond, M. L., H. S. Wheater, and C. Onof (2007), The significance of spatial rainfall representation for flood runoff estimation: A numerical evaluation based on the Lee catchment, UK, *Journal Of Hydrology*, *347*, 116–131.

Semadeni-Davies, A., C. Hernebring, G. Svensson, and L. G. Gustafsson (2008), The impacts of climate change and urbanisation on drainage in Helsingborg, Sweden: Combined sewer system, *Journal Of Hydrology*, *350*(1-2), 100–113.

Smyth, G. K. (1989), Generalized linear-models with varying dispersion, *Journal Of The Royal Statistical Society Series B-Methodological*, *51*(1), 47–60.

Waliser, D., K. W. Seo, S. Schubert, and E. Njoku (2007), Global water cycle agreement in the climate models assessed in the IPCC AR4, *Geophysical Research Letters*, *34*(16).

Yang, C., R. E. Chandler, V. S. Isham, and H. S. Wheater (2005), Spatial-temporal rainfall simulation using generalized linear models, *Water Resources Research*, *41*(11).