AWE-GEN

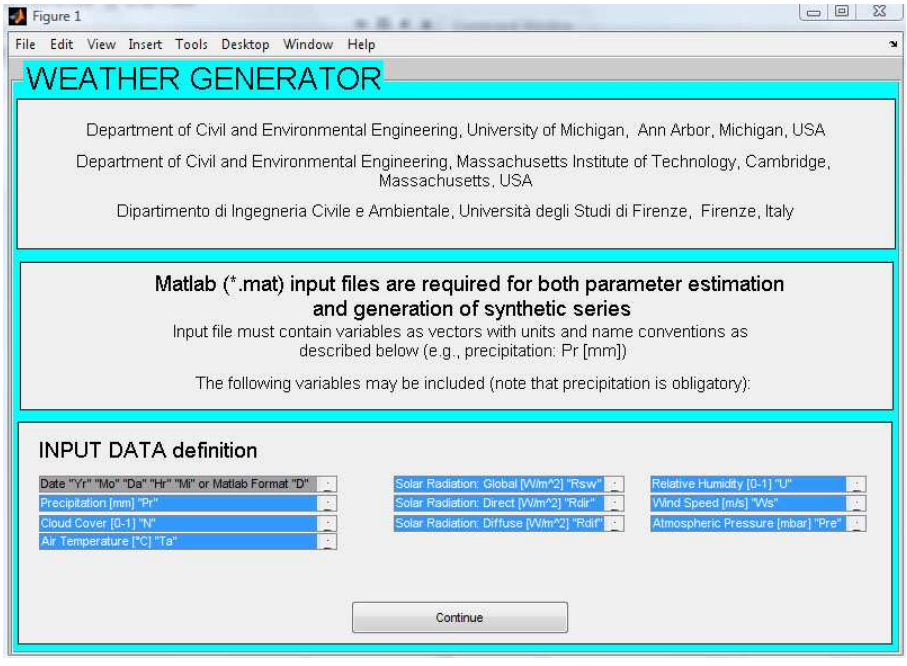
Advanced WEather GENerator

Technical Reference, version 1.0

Release: 7-2-2011

Simone Fatichi[1,2], Valeriy Y. Ivanov[2,3], and Enrica Caporali[1]

[1]Department of Civil and Environmental Engineering, University of Firenze, Firenze, Italy,
[2]Department of Civil and Environmental Engineering, University of Michigan, Ann Arbor, Michigan, USA
[3]Department of Civil and Environmental Engineering, Massachusset Institute of Technology, Cambridge, Massachusset, USA



Contents

List of Figures

List of Tables

List of Symbols

α Shape parameter of the Gamma distribution of rainfall intensity [−], page 7

α^{FUT} Shape parameter of the Gamma distribution of rainfall intensity for future climate conditions [−], page 49

α_Λ Ångström turbidity parameter [−], page 30

β^{-1} Mean waiting time for cell origins after the origin of the storm [h], page 7

β_e Effective turbidity coefficient, page 64

β_T Site slope [rad], page 33

β_Λ Ångström turbidity parameter [−], page 30

$\check{p}$ Rejection threshold for difference in annual precipitation series [%], page 10

Δe Vapor pressure deficit [Pa], page 38

ΔGMT Time difference between the local time zone and Greenwich Mean Time [h], page 57

ΔT Variation of mean monthly temperature from CTS and FUT scenario [$^\circ C$], page 49

ΔT_{day} Daily temperature amplitude [$^\circ C$], page 48

Δt_{SL} Time difference between the standard and local meridian [h], page 57

δ Solar declination [rad], page 22

$\eta(i)$ Random deviate of annual precipitation, page 9

$\eta(t)$ Random deviate of the stochastic component of wind speed, page 43

η^{-1} Mean duration of the cell [h], page 7

γ Coefficient controlling the transition function of the cloud process [h^{-1}], page 18

γ Daily angle, page 29

γ_n Skewness of $\eta(i)$, page 9

γ_n Skewness of $\eta(t)$, page 43

γ_{dW_S} Skewness the stochastic component of wind speed, page 43

$\gamma_{h,l}$ Second moment of Y_h in the NSRP model, page 7

$\gamma_{P_{yr}}$ Skewness of annual precipitation, page 9

$\gamma_{P_{yr}}^{FUT}$ Skewness of annual precipitation for future climate conditions, page 49

$\hat{f}_i$ Observed statistical properties or moments, page 8

$\kappa(h)$ Skewness of Y_h in the NSRP model, page 8

$\Lambda 1$ First band *UV/VIS* of shortwave radiation, page 29

$\Lambda 1_e$ Effective wavelength for the entire band $\Lambda 1$, page 62

$\Lambda 2$ Second band *NIR* of shortwave radiation, page 29

$\Lambda 2_e$ Effective wavelength for the entire band $\Lambda 2$, page 62

$\lambda\,, = 1\,, 2\,, 3\,, 4$ Spectral bands of the *Slingo* (1989) model, page 32

λ^{-1} Mean storm origin arrivals $[h]$, page 7

μ_c Mean number of cell per storm $[-]$, page 7

μ_c^{FUT} Mean number of cell per storm for future climate conditions $[-]$, page 49

μ_h Mean of Y_h in the NSRP model, page 7

ω_i Weights in the objective function, page 8

ω_Λ Single scattering albedo $[-]$, page 29

$\overline{d\Delta e}$ Mean of the stochastic component of vapor pressure deficit $[Pa]$, page 38

$\overline{dT}_h$ Average of the stochastic component of air temperature $[^\circ C]$, page 22

$\overline{dW_s}$ Mean of the stochastic component of wind speed $[m\ s^{-1}]$, page 43

$\overline{P_{atm}}$ Mean of the atmospheric pressure $[mbar]$, page 45

$\overline{P}_{yr}$ Average annual precipitation $[mm]$, page 9

$\overline{P}_{yr}^{FUT}$ Average annual precipitation for future climate conditions $[mm]$, page 49

Φ Local latitude $[rad]$, page 22

Φ' Local longitude $[angular\ degree]$, page 57

$\Phi(h)$ Probability that an arbitrary interval of length h is dry, page 7

$\rho(h)$ Lag-1 autocorrelation of Y_h in the NSRP model, page 8

ρ_g Ground albedo referring to a large area of 5-50 $[km]$ radius surrounding the point of interest, page 29

ρ_m Lag-1 autocorrelation function of correlated deviation in cloud cover simulation, page 18

$\rho_{csB,\Lambda}$ Cloud albedo, which is different for direct beam $[-]$, page 32

$\rho_{csD,\Lambda}$ Cloud albedo, incident diffuse radiation fluxes $[-]$, page 32

$\rho_{d\Delta e}$ Lag-1 autocorrelation of the stochastic component of vapor pressure deficit [−], page 38

ρ_{dT} Lag-1 autocorrelation of the stochastic component of air temperature, page 22

ρ_{dW_s} Lag-1 autocorrelation of the stochastic component of wind speed [−], page 43

$\rho_{P_{atm}}$ Lag-1 autocorrelation of the atmospheric pressure [−], page 45

$\rho_{P_{yr}}$ Lag-1 autocorrelation of annual precipitation, page 9

$\rho_{s,\Lambda}$ Sky albedo [−], page 29

σ Stefan-Boltzmann constant [$W\ m^{-2}\ K^{-4}$], page 22

σ_M^2 Variance of cloud cover in the "fair weather region, page 18

σ_m^2 Variance of correlated deviation in cloud cover simulation, page 18

$\sigma_{d\Delta e}$ Standard deviation of the stochastic component of vapor pressure deficit [Pa], page 38

$\sigma_{dT,h}$ Standard deviation of the stochastic component of air temperature [$^\circ C$], page 22

σ_{dW_s} Standard deviation of the stochastic component of wind speed [$m\ s^{-1}$], page 43

$\sigma_{P_{atm}}$ Standard deviation of the atmospheric pressure [$mbar$], page 45

$\sigma_{P_{yr}}$ Standard deviation of annual precipitation [mm], page 9

$\sigma_{P_{yr}}^{FUT}$ Standard deviation of annual precipitation for future climate conditions [mm], page 49

τ_N Cloud optical thickness [−], page 33

$\tau_S(t)$ Hour angle of the Sun [rad], page 57

$\tau_{a\Lambda}$ Spectral aerosol optical depth [−], page 29

θ Scale parameter of the Gamma distribution of rainfall intensity [$mm\ h^{-1}$], page 7

θ^{FUT} Scale parameter of the Gamma distribution of rainfall intensity for future climate conditions [$mm\ h^{-1}$], page 49

$\tilde{\omega_\lambda}$ Single scatter albedo of cloud optical properties, page 64

$\varepsilon(i)$ Standard normal deviate, page 9

$\varepsilon(t)$ Normal random deviates, Beta distributed random deviates in the cloud component, page 18

$\varphi_{S,T}$ Local solar illumination angle [rad], page 71

ς Coefficient controlling the transition function of the cloud process $[h^{-1}]$, page 18

$\vec{x}$ Position, page 33

$\widehat{\Delta e}$ Deterministic component of vapor pressure deficit $[Pa]$, page 38

$\widehat{W_s}$ Deterministic component of wind speed $[m\ s^{-1}]$, page 43

$\widetilde{E\{N(t)\}}$ Smoothed mean cloud cover $[-]$, page 52

$\widetilde{PAR}_{Bn}$ direct normal PAR $[W\ m^{-2}]$, page 33

$\widetilde{PAR}_{D}$ diffuse PAR $[W\ m^{-2}]$, page 33

$\widetilde{R}_{Bn,\Lambda}$ Direct beam radiation at normal incidence for cloudy sky conditions $[W\ m^{-2}]$, page 32

$\widetilde{R}_{D,\Lambda}$ Total diffuse radiation for cloudy sky conditions $[W\ m^{-2}]$, page 32

$\widetilde{R}_{Dd,\Lambda}$ Backscattered radiation for cloudy sky conditions $[W\ m^{-2}]$, page 32

$\widetilde{R}_{Dp,\Lambda}$ Incident diffuse radiation for cloudy sky conditions $[W\ m^{-2}]$, page 32

$\widetilde{R}_{Gn,\Lambda}$ Global radiation at normal incidence for cloudy sky conditions $[W\ m^{-2}]$ $[W\ m^{-2}]$, page 32

$\widetilde{T}(t)$ Deterministic component of air temperature $[^{\circ}C]$, page 22

ξ_h Third moment of Y_h in the NSRP model, page 7

ζ_S Solar azimuth $[rad]$, page 29

ζ_T Site aspect $[rad]$, page 33

a Beta probability distribution parameter, page 18

a_i Regression coefficient for the deterministic component of vapor pressure deficit, page 38

$A_{B,\lambda}$ Cloud diffuse reflectivity for direct beam incident radiation $[-]$, page 32

$A_{D,\lambda}$ Cloud diffuse reflectivity for diffuse incident radiation $[-]$, page 32

b Beta probability distribution parameter, page 18

B_a Aerosol forward scattering factor $[-]$, page 29

b_i Regression coefficient in the deterministic component of air temperature, page 22

$B_{R,\Lambda}$ Forward scattering fractions for Rayleigh extinction $[-]$, page 29

C Random number of cells in NSRP model $[-]$, page 7

c_i Regression coefficient for the deterministic component of wind speed, page 43

C_t Terrain configuration factor [−], page 71

$C_v(h)$ Coefficient of variation of Y_h in the NSRP model, page 8

$d\Delta e(t)$ Stochastic component of vapor pressure deficit $[Pa]$, page 38

D_{LH} Daily length $[h]$, page 29

$dT(t)$ Stochastic component of air temperature $[°C]$, page 22

dW_s Stochastic component of wind speed $[m\ s^{-1}]$, page 43

E_0 Ratio between the actual Earth-Sun distance and the mean Earth-Sun distance [−], page 29

e_a Ambient vapor pressure $[Pa]$, page 38

e_{sat} Vapor pressure at saturation $[Pa]$, page 38

f_i Theoretical statistical properties or moments, page 8

F_Λ Correction factor to compensate for multiple scattering effects [−], page 29

F_{obj} Objective function, page 8

g_λ Asymmetry parameter of cloud optical properties, page 64

h Rainfall aggregation period $[h]$, page 6

h_S Solar height $[rad]$, page 22

H_ζ Horizon angle $[rad]$, page 71

$J(t)$ Transition function of the cloud process between the boundary of a storm period [−], page 18

J_1 Cloud cover of the first hour after a storm and of the last hour of an inter-storm $[0-1]$, page 18

J_{Day} Julian Day, page 29

$K(\Lambda)$ Fractions of solar irradiance at the top of atmosphere in each band (*Gueymard*, 2008), page 32

$k(\lambda)$ Fractions of solar irradiance at the top of atmosphere in each band (*Slingo*, 1989), page 32

$K(t)$ Cloud attenuation factor [−], page 22

K_N Correction for the cloudiness in the atmospheric long-wave radiation calculation [−], page 22

L_{atm} Incoming atmospheric long-wave radiation $[W\ m^{-2}]$, page 22

$m(t)$ Stationary sequence of correlated deviation in cloud cover simulation $[0-1]$, page 18

m' Reference air mass $[-]$, page 29

M_0 Mean cloud cover in the "fair weather region $[0-1]$, page 18

m_A Aerosol extinction air mass $[-]$, page 62

M_B Reduction factor for direct PAR $[-]$, page 33

M_G Reduction factor for global PAR $[-]$, page 33

m_O Ozone absorption air mass $[-]$, page 62

m_R Rayleigh (molecular) scattering and uniformly mixed gases absorption air mass $[-]$, page 62

m_W Water vapor absorption air mass $[-]$, page 62

N Cloud cover, i.e., fraction of the celestial dome occupied by clouds $[0-1]$, page 18

P_{atm} Atmospheric pressure $[mbar]$ or $[Pa]$, page 45

$q(t)$ Scaled incoming long-wave radiation $[W\ m^{-2}]$, page 22

$r(t)$ Function of the solar height h_S, page 22

R_0 Solar constant $[W\ m^{-2}]$, page 29

R_0 "Fair weather region, page 18

R'_0 Extraterrestrial radiation $[W\ m^{-2}]$, page 29

r_e Effective radius of cloud-droplet size distribution $[\mu m]$, page 64

$R^T_{B,\Lambda}$ Direct beam radiation at the ground surface with topographic effects $[W/;m^{-2}]$, page 71

$R_{Bn,\Lambda}$ Direct beam radiation at normal incidence for clear sky $[W\ m^{-2}]$, page 29

$R_{D,\Lambda}$ Total diffuse radiation for clear sky conditions $[W\ m^{-2}]$, page 29

$R^T_{D,\Lambda}$ Diffuse radiation at the ground surface with topographic effects $[W/;m^{-2}]$, page 71

$R_{Dd,\Lambda}$ Backscattered radiation for clear sky conditions $[W\ m^{-2}]$, page 29

$R_{Dp,\Lambda}$ Incedent diffuse radiation for clear sky $[W\ m^{-2}]$, page 29

$R_{Gn,\Lambda}$ Global radiation at normal incidence for clear sky conditions $[W\ \ m^{-2}]$, page 29

$R^T_{R,\Lambda}$ Reflected radiation at the ground surface with topographic effects $[W/;m^{-2}]$, page 71

$R_{sw,n}$ Global shortwave radiation at normal incidence $[W\ m^{-2}]$, page 32

R_{sw} Global shortwave radiation $[W\ m^{-2}]$, page 32

$s(t)$ Function of the solar height h_S, page 22

$S_h(\vec{x},t)$ Shadow effect $[-]$, page 33

t Time, page 6

$T(t)$ Air temperature $[^\circ C]$, page 22

t_0 Time at which inter-storm period begins $[h]$, page 18

t_b Length of the inter-storm period $[h]$, page 18

T_m Average value of air temperature between sea level and Z_{ref} $[K]$, page 62

T_R Length of the post-storm transition period after which the cloud cover process can be considered stationary $[h]$, page 18

$T_{a,\Lambda}$ Aerosol extinction transmittance $[-]$, page 29

t_{aft} Forward difference between the standard time and the limits of integration of the sun variables $[h]$, page 57

$T_{as,\Lambda}$ Aerosol scattering transmittance $[-]$, page 29

T_a Air temperature at the reference height z_{atm}, $[^\circ C]$ or $[K]$, page 22

$T_{B,\lambda}$ Cloud transmissivity for direct beam flux $[-]$, page 32

t_{bef} Backward difference between the standard time and the limits of integration of the sun variables $[h]$, page 57

$T_{DB,\lambda}$ Cloud diffuse transmissivity for direct radiation $[-]$, page 32

$T_{DD,\lambda}$ Cloud diffuse transmissivity for incident diffuse radiation $[-]$, page 32

T_{dew} Dew point temperature $[^\circ C]$, page 38

$T_{g,\Lambda}$ Uniformly mixed gas absorption transmittance $[-]$, page 29

$T_{H\,rise}$ Local time of sunrise $[local\ hour]$, page 22

$T_{H\,set}$ Local time of sunset $[local\ hour]$, page 22

$T_{is\,\max}$ Maximum length of the interstorm period $[h]$, page 52

T_{is} Length of the interstorm period $[h]$, page 52

T_{mon} Monthly temperature [$°C$], page 49

$T_{n,\Lambda}$ Nitrogen dioxide absorption transmittance [$-$], page 29

$T_{o,\Lambda}$ Ozone absorption transmittance [$-$], page 29

$T_{R,\Lambda}$ Rayleigh scattering transmittance [$-$], page 29

$T_{w,\Lambda}$ Water vapor absorption transmittance [$-$], page 29

U Relative humidity [$-$], page 38

u_n Nitrogen dioxide amounts in the atmospheric column [cm], page 30

u_o Ozone amounts in the atmospheric column [cm], page 30

w Precipitable water in atmospheric column [cm], page 62

W_s Wind speed [$m\ s^{-1}$], page 43

X Rainfall intensity [$mm\ h^{-1}$], page 7

Y Intensity of precipitation, page 6

$Y(t)$ Hourly temperature change [$°C$], page 59

Y_h Aggregated rainfall depth, page 6

Z_{ref} Reference elevation [m], page 62

$\beta^{FUT^{-1}}$ Mean waiting time for cell origins after the origin of the storm for future climate conditions [h], page 49

$\eta^{FUT^{-1}}$ Mean duration of the cell for future climate conditions [h], page 49

$\lambda^{FUT^{-1}}$ Mean storm origin arrivals for future climate conditions [h], page 49

LWP Liquid water path [$g\ m^{-2}$], page 32

LWP_R Reference value of LWP for overcast condition [$g\ m^{-2}$], page 33

$S_{vf}(\vec{x})$ Sky view factor [$-$], page 33

1 Introduction

Records of meteorological variables around the world are often very short, with substantial gaps and low spatial coverage. This creates a problem of data inadequacy in numerous applications. In order to overcome such a problem, weather generators as the tools capable of generating consistent time-series of climatic variables have been proposed and used in the past (*Wilks and Wilby*, 1999). Specific motivations for using a weather generator can be found in several fields of science. These models have been significantly used in agricultural applications, e.g., studies of crop sensitivity and productivity to climate realizations (*Semenov and Porter*, 1995; *Mavromatis and Hansen*, 2001; *Dubrovský et al.*, 2004). In water resource engineering, climate simulators were used to generate long time series of precipitation that are required for flood risk analysis or water resource evaluations (*Fowler et al.*, 2000; *Wheater et al.*, 2005). Other possible applications are related to the generation of inputs to hydrological models (e.g., *Rigon et al.*, 2006; *Ivanov et al.*, 2008a), ecosystem models, or in long-term land management and erosion studies (e.g., *Collins et al.*, 2004; *Istanbulluoglu et al.*, 2004). Sometimes the weather generators are also employed to replace missing data from recorded time-series.

The generation of meteorological variables in a weather generator is frequently based on empirical statistical models. In these cases, statistical properties and correlations among variables are inferred from observed data. Precipitation is the most important variable in weather generators. It is frequently modeled by using an approach of separating the process of precipitation occurrence from the problem of determining the precipitation amount (*Wilks*, 1999; *Wilks and Wilby*, 1999; *Srikanthan and McMahon*, 2001). Other climate variables, or their residuals, since the mean and variance are typically removed, are simulated by means of regression equations. The regression parameters are usually estimated differently for wet, dry, and transitional states. The time scales at which these variables are simulated can range from daily to annual. A number of well known models can be listed in the category of empirical statistical approaches, such as the WGEN (*Richardson*, 1981; *Richardson and Wright*, 1984), the WXGEN model (*Sharpley and Williams*, 1990) used in the hydro-sedimentological SWAT model, the CLIGEN model (*Nicks et al.*, 1995) used in the hydro-sedimentological WEPP model, LARS-WG (*Semenov and Barrow*, 2002), ClimGen (*McKague et al.*, 2003), and Met&Roll (*Dubrovský et al.*, 2004).

A physically consistent method to generate meteorological variables is to directly use dynamic meteorological models, that solve the non-linear partial differential equations governing the dynamics of the atmosphere (*Cox et al.*, 1998). While being attractive, this approach has been mainly used for weather forecasting but not for long-term weather realizations, given the computational feasibility constraints of such simulations.

A third intermediate approach can be defined as the joint use of empirical statistical relations and physically-based methods. In essence, the approach adopts

stochastic models using some description of the underlying physical process, such as a simulation of rain cells and clustering, the cloudiness dependence on precipitation, the dependence of temperature on long- and short-wave radiation, etc. This third approach has received particular attention in rainfall modeling but it has been almost neglected in the implementation of complete climate simulators. A recent effort to fill this gap was undertaken by *Ivanov et al.* (2007), based on the earlier developed methodology of *Curtis and Eagleson* (1982), who proposed a weather generator at the hourly scale. While a number of variables are simulated stochastically, the approach attempts to preserve the underlying physical relations among them. The use of causal physical relationship within a weather generator allows one to simulate finer temporal scales (hourly or minute). Using an empirical statistical weather generator for the same purpose is more difficult since statistical correlations become more complex to model at shorter time scales.

Wilks and Wilby (1999) underlined that testing a weather generator in attempt to only reproduce the mean climate, for the way itself in which weather generators are realized, is somewhat naive. The real challenge is in reproducing higher order moments, correlations among the variables, and low and high frequency properties such, as extreme events and inter-annual variability. The latest efforts to test and improve weather generators are directed towards these directions (*Wilks*, 1999; *Hansen and Mavromatis*, 2001; *Kyselý and Dubrovský*, 2005; *Fowler et al.*, 2005; *Kilsby et al.*, 2007). A comparison between inter-annual variability of observed and simulated data is a crucial test of reliability of a weather generator. For instance, it has been noticed previously that a common characteristic of weather generators is the underestimation of the inter-annual variability (*Wilks*, 1989, 1999; *Wilks and Wilby*, 1999; *Srikanthan and McMahon*, 2001; *Kyselý and Dubrovský*, 2005). Several studies attempted to quantify this underestimation sometimes referred to as "overdispersion" (*Katz and Parlange*, 1998; *Wilks*, 1999). The random generation of numbers produce in the realizations a smaller variance than that of the corresponding observed data. *Wilks and Wilby* (1999) suggest that a possible explanation for the missing variance is that climate statistics change somewhat in the real world from year to year. A simple weather generator, with the underlining assumption of stationarity, cannot capture such a variation. In order to capture such a behavior, the model internal parameters should change in time, thus violating the stationarity assumption.

The issue of under-predicting inter-annual variability and extremes is the most difficult challenge for weather generators. Especially when a weather generator is used for simulation of future scenarios, both low and high frequency statistics should be tested (*Kyselý and Dubrovský*, 2005; *Semenov*, 2008). The capability of such models to reproduce extremes is related to the internal structure of the model. On the other hand, the capability to generate inter-annual variability is introduced by conditioning the model with external information. A common approach is to link parameters of the weather generator to some properties of large-scale atmospheric circulation. For instance, different sets of weather generator parameters can be used for different ranges of values assumed by large-scale atmospheric properties.

Several climate characteristics have been used for this scope: the mean monthly sea level pressure (SLP) (*Katz and Parlange*, 1993); the geostrophic wind direction (GWD) (*Kiely et al.*, 1998); the air masses provenance (*Wallis and Griffiths*, 1997); the objective Lamb weather type of atmospheric circulation (*Fowler et al.*, 2000); and the low-frequency realizations of monthly variable such as precipitation (*Wilks*, 1989) or temperature (*Hansen and Mavromatis*, 2001; *Kyselý and Dubrovský*, 2005).

The possibility to condition models externally allows one to use a weather generator when constructing climate change scenarios. The dependence of weather generator parameters on properties of large-scale atmospheric circulation links weather generators and General Circulation Models (GCMs). GCM realizations, in fact, can provide information about climate properties suitable to condition a weather generator. The GCM-predicted changes of large atmospheric patterns can therefore affect the weather generator parametrization and thus their realizations. *Wilks* (1992) pioneered the use of weather generator for climate change studies, and there is a recent evidence of a growing interest in such studies (*Semenov and Porter*, 1995; *Katz*, 1996; *Semenov and Barrow*, 1997; *Fowler et al.*, 2000, 2005; *Elshamy et al.*, 2006; *Kilsby et al.*, 2007; *Manning et al.*, 2009).

An hourly weather generator, AWE-GEN (Advanced WEather GENerator), is presented in this technical note. The generator is capable of reproducing low and high-frequency characteristics of hydro-climatic variables and essential statistical properties of these variables. The weather generator employs both the physically-based and stochastic approaches and is a substantial evolution of the model presented by *Ivanov et al.* (2007). Enhancements of the original formulation are the following: a new formulation of the precipitation module based on the Poisson-Cluster process; a new formulation of the module simulating vapor pressure instead of dew point temperature; simulation of the daily cycle of wind speed; significant modifications of the shortwave radiation module, in particular the inclusion of explicit simulation of the photosynthetically active radiation; minor modifications of the cloudiness and air temperature components; and a new model to reproduce time-variability of the atmospheric pressure.

An important capability for simulating the inter-annual variability of the precipitation process has been added. The capability to reproduce a wide set of statistics including extremes is also tested. Furthermore, a procedure to take into account non-stationary change of climate has been incorporated in the AWE-GEN framework. The procedure is based on a stochastic downscaling of GCM predictions (*Fatichi et al.*, 2011). The variables simulated by the weather generator at hourly scale are precipitation, cloud cover, shortwave radiation with partition into various type and spectral bands, air temperature, vapor pressure, wind speed, and atmospheric pressure. These variables are typically necessary as inputs for ecological, hydrological, geomorphological, and crop-dynamic models.

Although we are aware of the possible inaccuracy related to the random number-generation (*Meyer et al.*, 2007), this problem has not been addressed herein. The random generator tools available in the *Matlab*© software are used.

2 Data and model validation

The performance of AWE-GEN has been tested to reproduce observations at several locations with different climates. The weather generator has been validated for 10 airport meteorological stations located in the USA: Tucson (Arizona), Muskegon (Michigan), Albuquerque (New Mexico), Boston (Massachusetts), Nashville (Tennessee), San Francisco (California), Chicago (Illinois), Miami (Florida), Philadelphia (Pennsylvania), Atlanta (Georgia), and one meteorological station in Italy (Firenze University). Time series of hourly meteorological variables ranging in duration from 8 to 40 year period were available for these stations. The data for the USA location have been downloaded by *Webmet* meteorological resources center (http://www.webmet.com/). The data for Firenze have been provided by the Tuscany Functional Center. It should be noted that given gaps or absence of some meteorological variable, the test has not been realized for all variables in each station. The results of weather generator performance are shown only for Boston (MA), where time series of 80 years are simulated starting with 18 years of observations. The accuracy of the results is very similar among all the stations. Information about the data are available in the Webmet web-site.

The precipitation component has been further tested for four stations in the Tuscany region (Italy): Arezzo, Camaldoli, Vallombrosa and Firenze Ximeniano (data from: Tuscany Functional Center). Twenty-five years of precipitation at 20 minute resolution were available for these locations. This dataset for its particularly accuracy has been the object of previous studies where precipitation properties were analyzed and rainfall models developed (*Becchi et al.*, 1994; *Veneziano and Iacobellis*, 2002; *Cowpertwait et al.*, 2002).

3 Precipitation

Existing weather generators emphasize precipitation as the primary variable of interest (*Wilks and Wilby*, 1999; *Srikanthan and McMahon*, 2001). The underlying reason is that, due to the nature of coupling physical mechanisms, other hydroclimatic variables are affected directly or indirectly by the precipitation occurrence. Consequently, a correct reproduction of the precipitation regime influences all of the other weather variables. The use of models of stochastic precipitation has been addressed by the scientific community for many years, given their possible use in flood mitigation design, and in agricultural and ecological applications, etc. There is, indeed, a need of precipitation data across a range of temporal scales and for different purposes such as design of storm-water sewerage systems, flood hydrographs, and reservoir size. Among the first contributions in this field are worth to be mentioned the studies of *LeCam* (1961); *Todorovic and Yevjevich* (1969); *Todorovic and Woolhiser* (1975); *Waymire and Gupta* (1981a,b,c); *Foufoula-Georgiou and Lettenmaier* (1987). The first models of rainfall were developed treating separately the process of rainfall occurrence from the modeling of rainfall intensity. Frequently,

the occurrences of wet and dry states were simulated using Markov chains and the precipitation intensity with statistical distributions such as the Exponential or the Gamma distribution. Although more complex physically-based methods have been proposed, it is still common for weather generators to use this approach to generate precipitation.

Currently, generation of stochastic precipitation is mainly achieved with two methods, i.e., using models based on multifractality and Poisson-cluster models. Other approaches exist, but they have received less attention. The multifractal approach is based on observed scale invariance of the precipitation process, called multifractality. Multifractality implies that the rainfall process looks statistically the same at small and large scales, except for simple transformations (*Veneziano and Iacobellis*, 2002). The number of models that use multifractal scaling of rainfall has grown in the past decade (*Koutsoyiannis et al.*, 1998; *Veneziano and Furcolo*, 2002; *Veneziano et al.*, 2002, 2006).

In AWE-GEN, the method based on the Poisson-cluster model is used to simulate rainfall (*Onof et al.*, 2000). The development of the Poisson-cluster models has began with *Rodriguez-Iturbe et al.* (1987) and *Rodriguez-Iturbe and Eagleson* (1987) and was further developed by *Rodriguez-Iturbe et al.* (1988), *Entekhabi et al.* (1989), and *Cowpertwait* (1991); *Cowpertwait et al.* (1996). Both Neyman-Scott and Bartlett-Lewis types were used as stochastic point process. These two are different types of Poisson processes of storm origins. In Poisson processes each storm has associated a random number of rectangular pulse (cells) with random intensity and duration. Different cells and storms may overlap to produce the total hyetograph. At any time the rainfall is the sum of the cells active at that time, eventually belonging to different storms. The difference between the Neyman-Scott and the Bartlett-Lewis types, is concentrated in the method of cell origin displacement within a storm. In the Neyman-Scott model the time between storm origin and origin of each cell is considered a random variable. Conversely, in the Bartlett-Lewis model, the time between cell origins is considered to be a random variable. A schematic representation of the two models is shown in Figure: 1.

The development of Poisson-cluster models has continued, over the years, extending the models into two-dimensional space, in order to provide a framework for modeling multi-site and spatio-temporal rainfall data (*Cowpertwait*, 1995; *Northrop*, 1998; *Cowpertwait et al.*, 2002; *Wheater et al.*, 2005; *Cowpertwait*, 2006; *Leonard et al.*, 2008). Another improvement has been introduced explicitly calculating the theoretical function of the third moment of the rainfall process, in order to better fit the extreme values (*Cowpertwait*, 1998; *Cowpertwait et al.*, 2002). The process has been also generalized allowing convective and stratiform rain cells to occur (*Cowpertwait*, 1994). This purpose was reached overlapping two rectangular pulse models, thus enhancing the capability of the model in reproducing finer structure of the rainfall process (*Cowpertwait*, 2004; *Cowpertwait et al.*, 2007).

Reliability of the Poisson-cluster model has been confirmed by a comparative analysis of its performance with numerous observed time series of precipitation. The

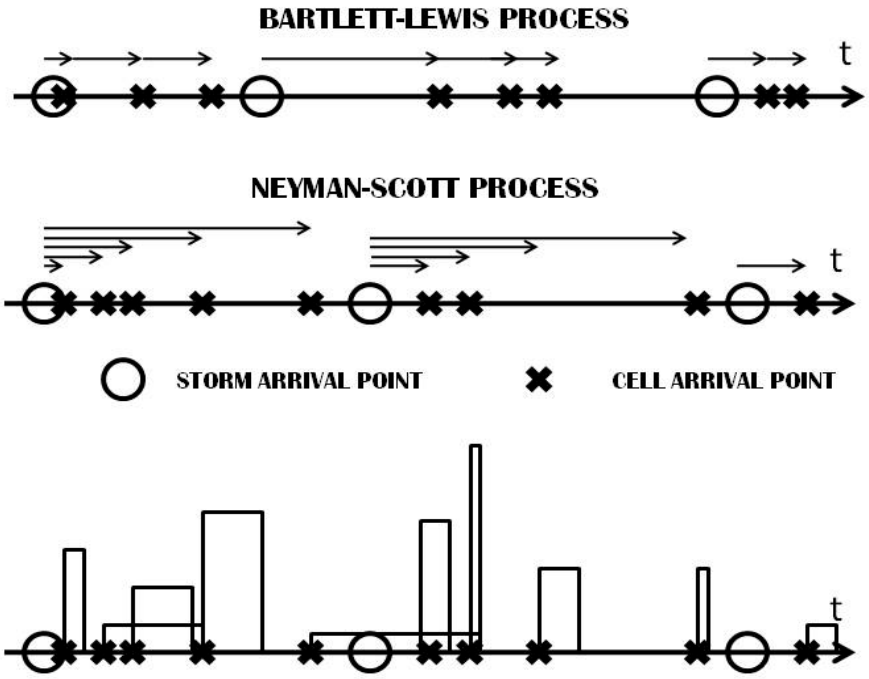


Figure 1: Schematic representation of Neyman-Scott and Bartlett-Lewis models with rectangular pulses.

model has demonstrated the capability to fit the essential characteristics of the precipitation process at a large number of time scales, including extreme events (*Cowpertwait*, 1991; *Cowpertwait et al.*, 1996; *Cowpertwait*, 1998; *Onof et al.*, 2000; *Cowpertwait et al.*, 2002; *Burton et al.*, 2008).

3.1 Neyman-Scott Rectangular Pulse model

The total intensity of precipitation $Y(t)$ of the Neyman-Scott Rectangulr Pulse (*NSRP*) model is the sum of the overlapping cells at any time t. This statement could be expressed formally with equation (1):

$$Y(t) = \int_{u=0}^{\infty} X_{t-u}(u) dN(t-u)\,, \tag{1}$$

where $dN(t-u)$ is 1 if there is a cell at the time $t-u$ and 0 otherwise, and $X_{t-u}(u)$ is the intensity at time t owing to a cell with origin at $t-u$. Given the characteristic of the rainfall measurements, rainfall data are available in aggregated form. Therefore, theoretical derivations of the statistical properties of the aggregated process $Y_h^{(i)}$ are needed to estimate the parameters of the model. The aggregated rainfall depth in the i^{th} interval of arbitrary length h is:

$$Y_h^{(i)} = \int_{(i-1)h}^{ih} Y(t) dt\,. \tag{2}$$

Under the hypothesis of stationarity the n^{th} moment of the process is $E\{(Y_h^{(i)})^n\} = E\{(Y_h^{(j)})^n\}$ and the indexes i and j could be omitted (*Cowpertwait*, 1998). The theoretical properties of a Neyman-Scott rectangular pulse model for a single site were derived up to the second order moments of Y_h by *Rodriguez-Iturbe et al.* (1987). The probability that an arbitrary interval of length h is dry was derived by *Cowpertwait* (1991) and the third moment was successively derived by *Cowpertwait* (1998). The theoretical expressions for the statistical properties of the *NSRP* model are a function of the statistical distribution chosen to model the random processes of storms and cells occurrence and intensity.

The Neyman-Scott Rectangular Pulse (NSRP) model used to generate the internal structure of precipitation process in AWE-GEN is primarily based on the approach of *Cowpertwait* (1998); *Cowpertwait et al.* (2002); *Cowpertwait* (2004). The storm time origin occurs as a Poisson process with the rate λ $[h^{-1}]$, a random number of cells C is generated for each storm according to the geometrical distribution with the mean μ_c $[-]$. Cell displacement from the storm origin is assumed to be exponentially distributed with the mean β^{-1} $[h]$. A rectangular pulse associated with each precipitation cell has an exponentially distributed life time with the mean η^{-1} $[h]$ and intensity X $[mm\ h^{-1}]$. The latter is distributed according to the Gamma distribution with the parameters α and, θ. X must be positive and its probability density function is:

$$P(X) = \frac{X^{\alpha-1}e^{-X/\theta}}{\Gamma(\alpha)\,\theta^{\alpha}}\,. \tag{3}$$

An overview of the *NSRP* parameters is provided in table 1. The distributions adopted for the random process within the *NSRP* model fully define the statistical properties of the aggregated process $E\{Y_h\}$ over an arbitrary time-scale h (*Cowpertwait*, 1998). The mean is:

$$\mu_h = E\{Y_h\} = \lambda\mu_c E\{X\}h/\eta\,, \tag{4}$$

and the second moment is:

$$\begin{aligned}\gamma_{h,l} \;=\; & COV\{Y_h^i, Y_h^{i+l}\} = \lambda\eta^{-3}A(h,l)[2\mu_c E\{X^2\} + [E\{X\}]^2\beta^2 E\{C^2-C\}/(\beta^2-\eta^2)] \\ & -\lambda[E\{X\}]^2 B(h,l)E\{C^2-C\}/[\beta(\beta^2-\eta^2)]\,,\end{aligned} \tag{5}$$

where $A(h,l)$ and $B(h,l)$ are defined in *Cowpertwait* (1998) (see Appendix A). The moments of the rainfall intensity for the Gamma distribution are $E\{X^n\} = \theta^n\Gamma(\alpha+n)/\Gamma(\alpha)$; and for the geometric distribution with mean equal to μ_c are $E\{C^2 - C\} = 2\mu_c(\mu_c-1)$ and $E\{(C^2-C)(C-2)\} = 6\mu_c(\mu_c-1)^2$. The third moments $\xi_h = E\{[Y_h - E\{Y_h\}]^3\}$ is also defined in *Cowpertwait* (1998) (see Appendix A). The probability that an arbitrary interval of length h is dry $\Phi(h) = P(Y_h = 0)$ is taken from *Cowpertwait* (1991) and *Cowpertwait et al.* (1996) with some modifications to make use of the geometrical distribution rather than the Poisson distribution in the generation of the random number of cells within a storm (see Appendix: A).

Table 1: The parameters of the point Neyman-Scott Rectangular Pulse model.

Parameter	**Explanation**
λ^{-1}	Mean storm origin arrivals $[h]$
β^{-1}	Mean waiting time for cell origins after the origin of the storm $[h]$
η^{-1}	Mean duration of the cell $[h]$
μ_c	Mean number of cell per storm $[-]$
α	Shape parameter of the Gamma distribution of rainfall intensity $[-]$
θ	Scale parameter of the Gamma distribution of rainfall intensity $[mm\ h^{-1}]$

3.2 Parameter fitting procedure

The utilized model has six unknowns and, thus, at least six equations are required in order to estimate these parameters. An exact estimation of the six parameters would need six statistical properties or moments, $\hat{f}_i$, inferred from the observed data. The $\hat{f}_i$ should be successively compared with the statistical properties obtained from the theoretical equations of the *NSRP* model, f_i. Theoretically the following equation should be verified:

$$f_i(\lambda, \beta, \eta, \mu_c, \alpha, \theta) = \hat{f}_i \,. \tag{6}$$

From equation (4) and the equation for the first moment, $E\{X\}$, one of the six parameters can be made explicit in terms of the mean, $E\{Y_h\}$, and the remaining parameters. Usually, θ is derived as a function of the other parameters (*Cowpertwait*, 1998; *Cowpertwait et al.*, 2002, 2007):

$$\theta = \frac{\eta\, E\{Y_h\}}{\alpha\, \lambda\, \mu_c\, h} \,. \tag{7}$$

Including equation (7) the problem reduces to the estimation of only five parameters. Rather than fitting exactly the parameters of the model, it is more desirable to use a wider set of statistical properties, e.g., $m > 5$, and to find the best approximate solution. This solution allows one to better reproduce a larger set of statistical properties instead of exactly reproducing few of them. In order to achieve this purpose, an objective function F_{obj} is defined following the procedure proposed by *Cowpertwait* (2006); *Cowpertwait et al.* (2007):

$$F_{obj} = \sum_{i=1}^{m} \omega_i \left[\left(1 - \frac{\hat{f}_i}{f_i}\right)^2 + \left(1 - \frac{f_i}{\hat{f}_i}\right)^2 \right] , \tag{8}$$

where ω_i are the weights in the objective function to emphasize the importance of certain statistical properties over the others. The choice of the m statistical properties, of $\hat{f}_i$, and of the weights ω_i into the objective function depends on the primary scope of the rainfall model. In the weather generator context, $\hat{f}_i$ are selected such that allow the model to fit a wide set of statistical properties without emphasiz-

ing any one in particular. After having carried out a large number of tests using available data, the four following properties were selected: the coefficient of variation, $C_v(h) = \sqrt{\gamma_{h,0}}/\mu_h$, the lag-1 autocorrelation, $\rho(h) = \gamma_{h,1}/\gamma_{h,0}$, the skewness $\kappa(h) = \xi_h/\gamma_{h,0}^{3/2}$, and the probability that an arbitrary interval of length h is dry, $\Phi(h)$. The utilized fitting procedure assumes that rainfall time series are available as the coarsest temporal resolution of 1 hour. It specifically uses the statistical properties of the rainfall process at four different aggregation periods, h, 1, 6, 24, and 72 hours. The weights, ω_i, are taken equal to "1" for all statistical properties and for the four different aggregation times. Totally, $m = 16$ statistical properties of rainfall observations are used to fit the five parameters $(\lambda, \beta, \eta, \mu_c, \alpha)$ and $E\{Y_1\}$ is finally used to estimate θ. Given the high non-linearity in the parameter functions, the automated procedure should be well constrained to avoid unrealistic values of the parameters (e.g., *Cowpertwait* (1998)). The adopted feasible regions are taken from *Cowpertwait* (1998): $0.0001 < \lambda < 0.05$; $0.01 < \beta < 0.99$; $1 < \mu_c < 80$; $0.5 < \eta < 30$; $0.1 < \alpha < 20$; in comparison to the original formulation, the regions of validity for η, μ_c, and α are restricted to reduce the tendency of the optimization procedure toward the boundaries. Overall, the simplex method (*Nelder and Mead*, 1965) is used as a minimization method for the imposed objective function. The method has been previously employed with good performance also in terms of its convergence characteristics (*Cowpertwait*, 1998; *Cowpertwait et al.*, 2007). In order to take into account the seasonality of site climatology, the parameters are estimated on a monthly basis, i.e., six parameters for each months need to be inferred to completely define the *NSRP* model.

3.3 Low-frequency properties of the rainfall process

Previous efforts of validation of the *NSRP* model at larger time intervals, for instance at the yearly time scales, have indicated that the variance of the simulated process was smaller than the one inferred from observed data. As already stated in the introduction, due to their nature the conventional weather generator techniques often fail to capture entirely inter-annual variability (*Wilks and Wilby*, 1999). This observation is related to the underlying assumption of stationarity of precipitation process. This underestimation of inter-annual variability can be problematic for numerous applications in hydrology or when climate change scenario needs to be explicitly introduced. *Kilsby et al.* (2007) highlight that this problem is present not only in the framework of weather generators but in the physically-based climate models as well.

Attempts to resolve this issue have typically used an external conditioning of the parameters of the rainfall models. The conditioning was derived by climate characteristics, such as monthly statistic (*Wilks*, 1989) or indices of large-scale circulation (*Kiely et al.*, 1998). These approaches have been used with Markov chain or renewal process of precipitation. However, examples to link *NSRP* models to patterns of large-scale circulation also exist (*Fowler et al.*, 2000, 2005). The external condition-

ing allows one to produce realizations for a non-stationary climate. For instance, changes in large-scale circulation patterns can be inferred from climate models and transferred into "future" meteorological realizations using a rainfall generator.

In this study, the capability of reproducing low-frequency properties of the precipitation process is introduced selecting externally the total annual precipitation generated with the *NSRP* model, on the basis of an annual precipitation model, as explained later in this section. Following this approach, the variance and autocorrelation properties of precipitation process at the annual scale are preserved. Note that this does not automatically assure the preservation of the monthly variance of precipitation. The preservation of the annual variance, without preserving the monthly one might induce a theoretical error. For instance, extremely drought years may be obtained with a uniformly lower amount of precipitation in all of the months rather than due to a drastic reduction of rainfall in a few months and vice-versa for extremely wet years. This artifact becomes larger as the difference between the simulated and the observed variance of monthly precipitation increases. Fortunately, the "overdispersion" at monthly scale is generally limited and the artifacts imposed by the methodology are typically negligible, as shown in the result analysis. Furthermore, it can be argued that inter-annual variability of precipitation is the the preferred property to be preserved for most agricultural, ecological, and hydrological applications.

Markov-type models have been commonly used to reproduce annual time series of precipitation (*Srikanthan and McMahon*, 1982, 2001), although they neglect the long term persistency of the process (*Koutsoyiannis*, 2003). In this study the inter-annual variability of precipitation is simulated using an autoregressive order-one model, AR(1), with the skewness modified through the Wilson-Hilferty transformation (*Wilson and Hilferty*, 1931; *Fiering and Jackson*, 1971):

$$P_{yr}(i) = \overline{P}_{yr} + \rho_{P_{yr}}(P_{yr}(i-1) - \overline{P}_{yr}) + \eta(i)\sigma_{P_{yr}}\sqrt{1-\rho_{P_{yr}}^2}\,, \tag{9}$$

where $\overline{P}_{yr}$ $[mm]$ is the average annual precipitation, $\sigma_{P_{yr}}$ $[mm]$ is the standard deviation, and $\rho_{P_{yr}}$ is the lag-1 autocorrelation of the process. The term $\eta(i)$ represents the random deviates of the process and is skewed according the Wilson-Hilferty transformation:

$$\eta(i) = \frac{2}{\gamma_n}\Big(1 + \frac{\gamma_n\,\varepsilon(i)}{6} - \frac{\gamma_n^2}{36}\Big)^3 - \frac{2}{\gamma_n}\,, \tag{10}$$

where the skewness of $\eta(i)$ is $\gamma_\eta = (1-\rho_{P_{yr}}^3)\gamma_{P_{yr}}/(1-\rho_{P_{yr}}^2)^{1.5}$; $\gamma_{P_{yr}}$ is the skewness inferred from observations and $\varepsilon(i)$ are the standard normal deviates. The Wilson-Hilferty transformation is not exact. However, the lag-one autocorrelation and the coefficient of skewness of annual rainfall data are usually within the limits of validity of the transformation and thus no significant errors are introduced (*McMahon and Miller*, 1971).

The *NSRP* model that captures intra-annual precipitation regime (the high-frequency

properties) is coupled with the AR(1) model (equation 9) that reproduces precipitation inter-annual variability (the low-frequency properties) in the following manner. First, the *NSRP* model is used to simulate precipitation series at the hourly time scales for the period of one year. The obtained total precipitation is then compared with the annual value estimated with the autoregressive model (9). If the difference between the two values is larger than a certain percentage $\check{p}$ of the measured long-term mean annual precipitation, the simulated one-year long hourly series are rejected, a new series is generated and the above comparison is repeated. Once the difference between the two values is below the $\check{p}$ threshold, the simulated with the *NSRP* model time series of precipitation are accepted. The entire procedure is repeated until all annual values generated with the model AR(1) model have matching hourly series generated with the *NSRP* model. The rejection threshold $\check{p}$ can be chosen according to the information about observational errors of annual precipitation. An illustrative example of the coupling between the two methods is shown in Figure 2 for the location of Tucson airport (AZ), with $\check{p} = 2.5\%$. As seen, the inter-annual variability of precipitation process is reproduced correctly.

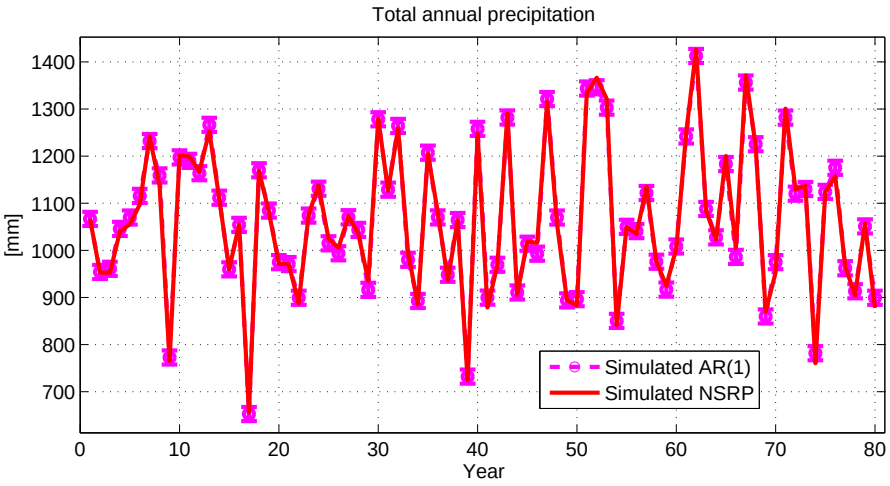


Figure 2: The annual precipitation simulated with the *NSRP* model (red line) after the external selection based on the AR(1) precipitation series (magenta dots) has been carried out. The vertical bars denote the $\check{p} = 2.5\%$ of the long-term average annual precipitation.

Given the stationary nature of the *NSRP* model, the search of "suitable" years can be computationally exhaustive for locations characterized by a high variance of annual precipitation. For instance, it can be the case when observed time series have limited duration. The short duration leads to a larger variance and does not permit a correct evaluation of the internal parameters of the *NSRP* model. The computationally exhaustive search of "suitable" years is only related to time constraints, because although it is not easily demonstrable, it should be expected that some rare combination of random numbers in the *NSRP* model would reproduce total annual precipitation equal to the one simulated by the annual model, AR(1).

In order to reach the convergence in a reasonable computational time, an adjustment procedure similar to that proposed by *Kyselý and Dubrovský* (2005) is

introduced after a pre-defined number of iterations without a satisfactory match. Specifically, discarded one-year long hourly NSRP precipitation series are first selected that have the closest match to the precipitation simulated with the AR(1) model. These series are subsequently multiplied by a correction factor to match the annual precipitation simulated with the AR(1) model. This correction alters the internal structure of the data generated with the *NSRP* model. However as justified by *Kyselý and Dubrovský* (2005), the magnitude of such an adjustment is several times lower than the inter-diurnal or diurnal variability, and consequently the effect of the adjustment procedure is insignificant. The correction are minimized by choosing the discarded precipitation series with the total annual precipitation as closest as possible to the unmatched annual precipitation simulated by the AR(1) model. As concluded from our numerous experiments, such adjustments were found to be necessary only for a few years in a millennium and only for stations with limited records.

Overall, it is theoretically possible that the proposed procedure might somewhat alter the intra-annual structure of the rainfall process because of the applied correction factors and since the output of the *NSRP* model is sampled in a non-random fashion. However, it is argued that these drawbacks are minor with respect to the overall capability of reproducing the inter-annual variability of precipitation process. The negligible effect of the procedure is confirmed by the results obtained for precipitation statistics at shorter aggregation periods (shown in the result analysis). Furthermore, using long-term simulations we have successfully verified that the results of simulating high-frequency precipitation are not distinguishable between the cases when the model feature of inter-annual variability is enabled or not (results not shown).

When inter-annual variability is the most important statistical property to reproduce, the proposed procedure could be enhanced using an ARFIMA(p, d, q) model instead of the AR(1). Without describing the details of using an ARFIMA model, such a type of model allows one to take into account the long-memory eventually present in the time series (*Montanari et al.*, 1997; *Koutsoyiannis*, 2000, 2003).

3.4 Results and validation

The capability of the model to reproduce the main statistics of the precipitation process at different aggregation periods is tested. The simulated mean, variance, lag-1 autocorrelation, skewness, frequency of non-precipitation, i.e. the probability that an arbitrary interval of length h is dry, and the transition probability from a wet-spell to another wet-spell are compared with observations at the monthly scale. The comparison is shown at the periods of aggregation of 1, 24, and 48 hours (Figure 3, 4 and 5).

Reproducing statistical properties different from the ones used in the calibration of precipitation parameters such as transition probability from wet-spells or all the statistics different from mean at aggregation of 48 $[h]$ is challenging. The results shown in Figure 5 for the 48 $[h]$ aggregation period confirm that the statistical

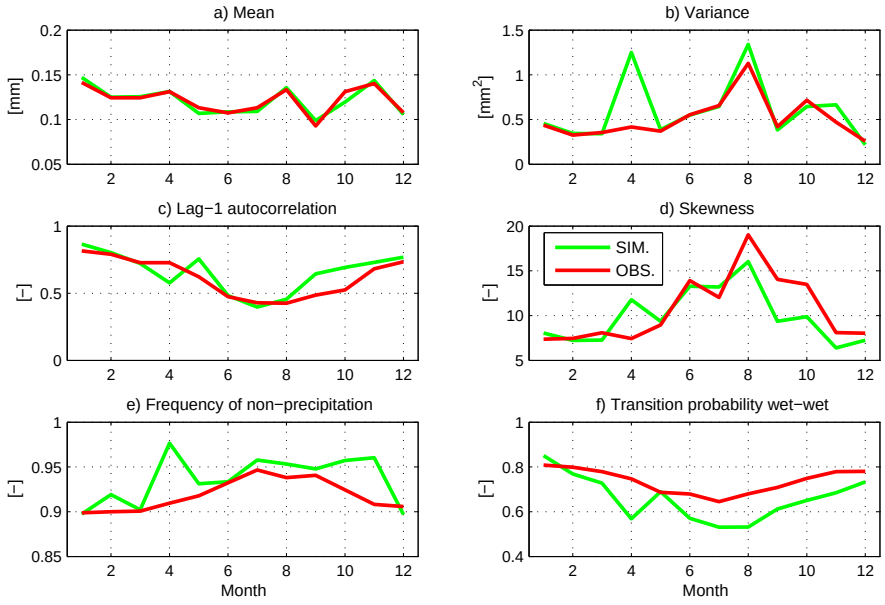


Figure 3: A comparison between observed (red) and simulated (green) monthly statistics of precipitation (mean, variance, lag-1 autocorrelation, skewness, frequency of non-precipitation, transition probability wet-wet), for the aggregation period of 1 hour.

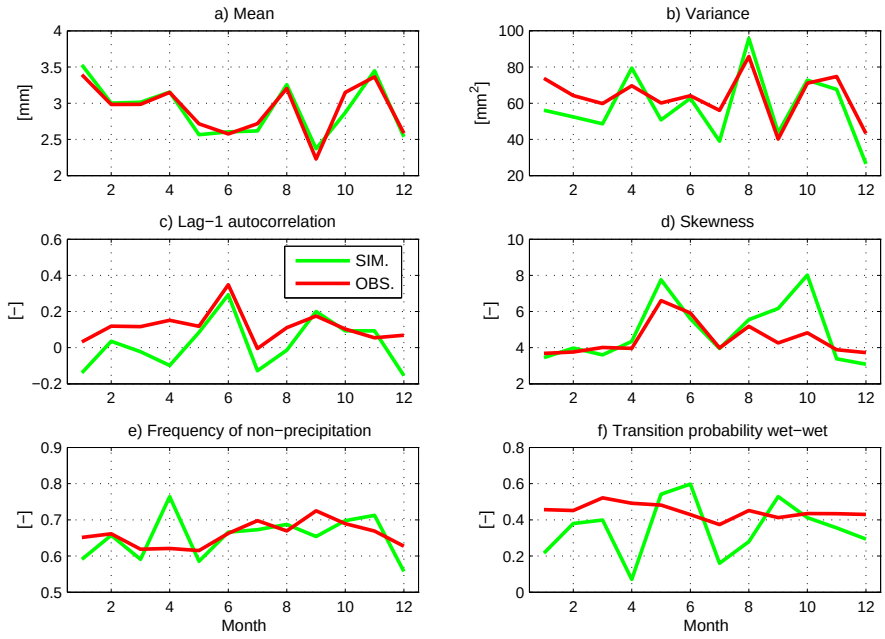


Figure 4: A comparison between observed (red) and simulated (green) monthly statistics of precipitation (mean, variance, lag-1 autocorrelation, skewness, frequency of non-precipitation, transition probability wet-wet), for the aggregation period of 24 hours.

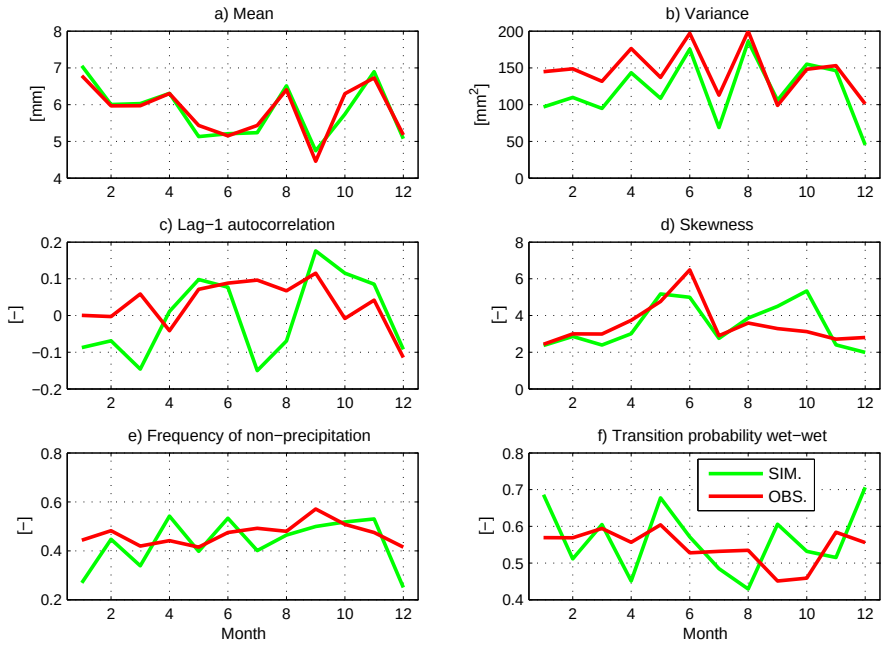


Figure 5: A comparison between observed (red) and simulated (green) monthly statistics of precipitation (mean, variance, lag-1 autocorrelation, skewness, frequency of non-precipitation, transition probability wet-wet), for the aggregation period of 48 hours.

properties are also preserved at this aggregation time. After the verification of statistics at short aggregation periods, the entire annual cycle of the rainfall process is checked in Figure 6. The simulated process perfectly preserves the mean but underestimates the monthly variance of observations, almost in every month. These differences are related to the poor skill of the *NSRP* model in reproducing, the low frequency variances as discussed in Section 3.3.

The precipitation component must be checked to properly reproduce also extreme values including rainfall maxima and occurrences of dry and wet periods. The performance of the *NSRP* model with regards to the reproduction of the extreme values is influenced by its internal structure (probability distributions of random variables). Specifically, the distribution used to simulate the random intensity of the rainfall cell, X, directly affects the realizations of extremes. A Gamma probability distribution as previously tested by *Cowpertwait* (1998) is employed. Weibull and Mixed-Exponential probability distributions have been also tested and compared with the Gamma. Nonetheless, no appreciable differences have been observed. All of these probability distributions provide consistent results in terms of fitting of extreme values of precipitation intensity. The simulated and observed extreme precipitations for time aggregation periods of 1 hour and 24 hours are illustrated in Figures 7a and 7b. For all test locations, there is a good match between the simulated and observed values, especially for the return periods at up to 20-30 years. This is not appreciable for the location at Boston (Figure 7a,b), where only 18 years of observed values were available. For larger return periods, multiple simulations would be necessary

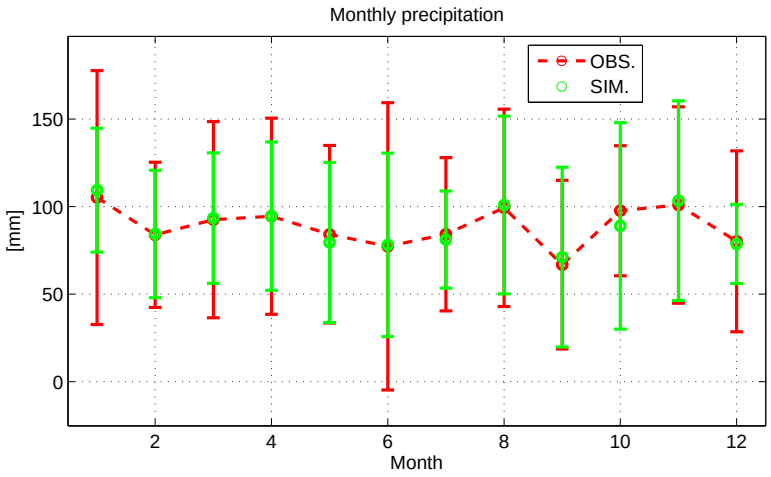


Figure 6: A comparison between observed (red) and simulated (green) monthly precipitation. The vertical bars denote the standard deviations of the monthly values.

to define the mean and confidence intervals of extreme precipitation and effectively corroborate the weather generator (*Semenov*, 2008). Cumulative probabilities associated with the data are estimated with the method of plotting position (*Cunnane*, 1978).

Extremes of dry spell and wet spell durations are generally poorly captured by the model, especially for dry climates. Simulations and observations sometimes differ also for return periods of less than one year. The results for Boston are illustrated in Figure 7c, where extreme dry spells are well simulated, while extreme wet spells, are slightly overestimated for return periods larger than 10 years (Figure 7d).

The fractions of total time that precipitation exceeds a certain depth are shown for different aggregation periods in Figure 8a. As seen, precipitation events with depth larger than 1 $[mm]$ are somewhat overestimated for aggregations periods longer than 48 $[h]$. Conversely, the fractions of time with precipitation depth larger than 20 $[mm]$ is slightly underestimated for the same aggregation periods. This is consistent given the preservation of precipitation average at each aggregation period. Errors of such type are almost unavoidable in the *NSRP* model, as parameterized in AWE-GEN. They might be related to the use of a single set of parameters to describe rainfall cells and clusters, that in the natural process are the result of different mechanisms, such as stratiform and convective rainfall. The use of *NSRP* models that overlap two different kind of cells (*Cowpertwait*, 2004; *Cowpertwait et al.*, 2007) might be used to reduce this error. The distribution of dry spell duration (Figure 8b) is usually represented reasonably well, although for temperate climates its mean is slightly underestimated. For Boston it is underestimated by 0.7 days, as confirmed from Figure 8b. In drier climates, the mean is usually preserved but the shape of the distribution can deviate from the observed one for intermediate dry spell durations. The distribution of the wet spell durations is generally captured by the weather generator with respect to the mean and the shape of the probability distribution (Figure 8c). This performance is realized whether the climate is dry or wet.

The errors seen in Figure 8a, are generally negligible. For example, the difference in the fraction of time precipitation depth exceed 1 $[mm]$ at aggregation time of 96 $[h]$ is typically around 0.1 but usually less than 0.05. These errors are acceptable and imply that only a small amount of rainfall is transferred from intense rainfall to drizzle. Errors in the representation of dry spell duration mean between 0.5 and 1.5 days are generally not desirable. It should be noted that this is usually the most difficult precipitation property to be simulated by the weather generator. This feature is important since the distribution of dry spell duration significantly affects the simulation of all other variables. Fortunately, it has been checked that although some inaccuracy can be appreciated, it does not influence the simulation performance of the other variables, as testified from the results in the following.

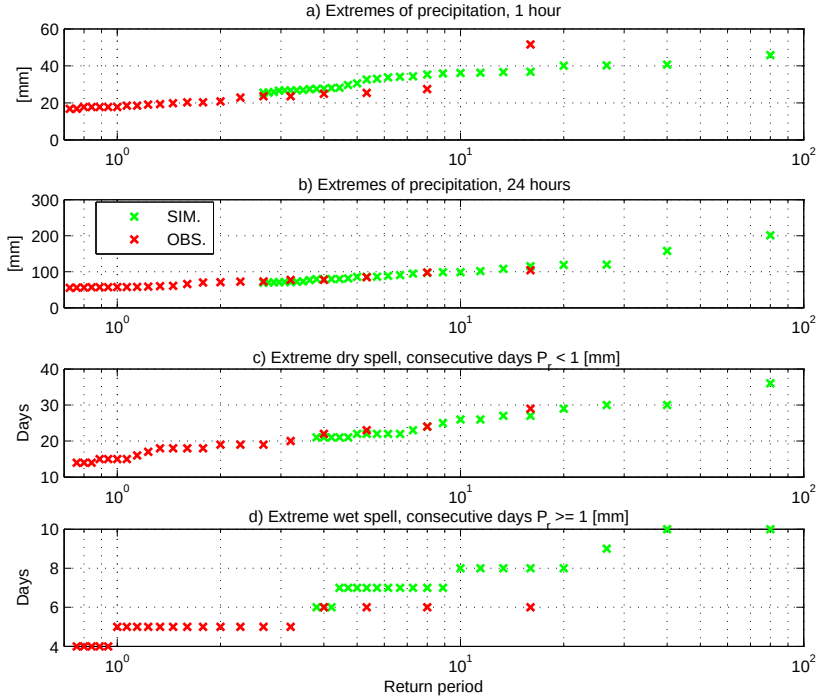


Figure 7: A comparison between the observed (red crosses) and simulated values of extreme precipitation (green crosses) at (a) 1-hour and (b) 24-hour aggregation periods; (c) extremes of dry and (d) wet spell durations. Dry/wet spell duration is the number of consecutive days with precipitation depth lower/larger than 1 $[mm]$.

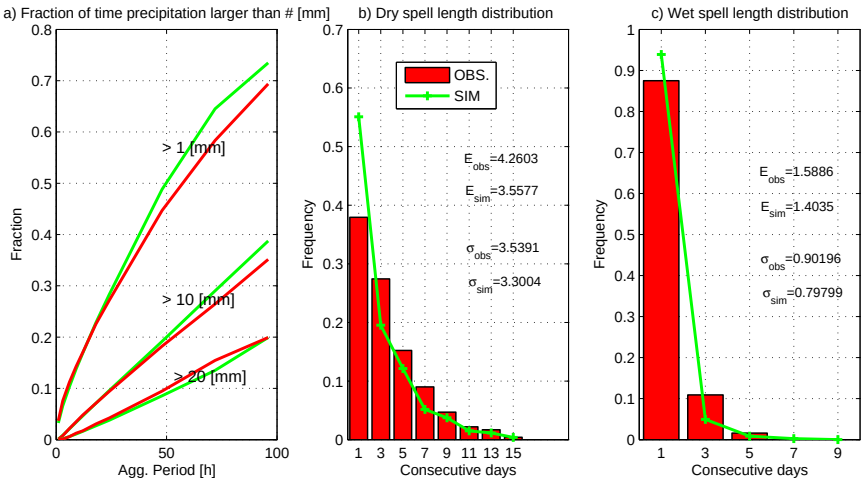


Figure 8: A comparison between observed (red) and simulated (green) fractions of time with precipitation larger than a given threshold $[1 - 10 - 20mm]$ at different aggregation periods (a). The same comparison for dry spell length distribution (b), i.e. consecutive days with precipitation depth lower than 1 $[mm]$ and for wet spell length distribution (c), i.e. consecutive days with precipitation depth larger than 1 $[mm]$. E_{obs} and σ_{obs} are the observed mean and standard deviation and E_{sim} and σ_{sim} are the simulated ones.

4 Cloud cover

Cloud cover is an important climatic variable directly affecting radiation fluxes and indirectly influencing air temperature and relative humidity. This variable is often neglected in empirical statistical weather generators (*Richardson*, 1981; *Semenov et al.*, 1998; *Parlange and Katz*, 2000). Commonly, weather generators simulate variables dependent on cloud cover, such as air temperature, on the basis of precipitation occurrence, e.g., dry and wet states. The latter are only weak implicit proxies for the process of cloud cover, that is not explicitly included. In simulations of the hydrological cycle and energy budget, the cloud cover is often assumed constant or its definition is oversimplified. This is a theoretically incorrect assumption incompatible with the high-frequency variation of cloud cover and its effect on shortwave radiation fluxes. In some applications, such as modeling of snowpack or vegetation dynamics, this assumption can lead to unrealistic results. The cloud cover simulated in AWE-GEN is based on the framework first developed by *Curtis and Eagleson* (1982) and further modified by *Ivanov et al.* (2007).

4.1 Model

Cloud cover $N(t)$ is the fraction of the celestial dome occupied by clouds. It can be measured in oktas $[0-8]$ or in cloud fraction $[0-1]$, where 0 signifies clear sky conditions and 1 is used to describe complete overcast conditions (*Muneer et al.*, 2000). The fraction notation will be used in the following. In the model of *Ivanov et al.* (2007), $N(t)$ $[-]$ is considered to be a random variable that has different dynamics during intra-storm and inter-storm periods. During an intra-storm period, i.e. the hours with precipitation different from zero, the value of cloudiness is assumed to be equal to 1. During an inter-storm period, the existence of the "fair weather" region, R_0, is assumed. The region is sufficiently distant from storms, thus the cloud cover can be assumed stationary and fully characterized by the first two statistical moments: the mean $E\{N(t)\}_{t\in R_0} = M_0$ and the variance $VAR\{N(t)\}_{t\in R_0} = \sigma_M^2$ of the process. The length of the post-storm transition period after which the cloud cover process can be considered stationary is indicated with T_R $[h]$. The second assumption is that the transition of the cloud process between the boundary of a storm period and the fair-weather takes place through an exponential function $J(t)$. The latter is characterized by two coefficients controlling the transition rates, ς and γ $[h^{-1}]$, and by the average cloud cover of the first hour after a storm and of the last hour of an inter-storm: J_1. The expression for the cloudiness becomes:

$$N(t) = M_0 + \big(J_1 - M_0\big)\big(1 - J(t)\big) + m(t)J(t)\,, \tag{11}$$

where $m(t)$ is the stationary sequence of correlated deviation with $E\{m(t)\} = 0$; $VAR\{m(t)\} = \sigma_m^2$ and autocorrelation function $\rho_m(l)$ (where l is the lag). The time varying conditional expectation and variance of cloud cover under this assumption

have been estimated from *Curtis and Eagleson* (1982):

$$E\{N(t)\}_{t\in t_b} = M_0 + \big(J_1 - M_0\big)\big(1 - J(t)\big) \ , \tag{12}$$

$$VAR\{N(t)\}_{t\in t_b} = \sigma_m^2 J(t)^2 \, . \tag{13}$$

The stationary sequence of correlated deviation $m(t)$ is modeled through an AR(1) model where the random deviates $\varepsilon(t)$ are distributed following a Beta probability distribution with parameter a and b and evaluated with the same procedure of lower and upper bound constraining proposed by *Ivanov et al.* (2007). The a and b parameters are estimated on the basis of the of the discretized cloudiness [0:0.1:1] at the step before $N(t-1)$ and therefore are in total eleven parameters. The procedure proposed by *Ivanov et al.* (2007) allows to use a different distribution of the correlated deviation $m(t)$ function of the cloudiness $N(t-1)$. This has been shown to improve significantly the results of the model in comparison to a fixed distribution for $\varepsilon(t)$, as used in *Curtis and Eagleson* (1982). The AR(1) model for $m(t)$ can be expressed as follows:

$$m(t) = \rho_m m(t-1) + \varepsilon(t)\sigma_m \sqrt{1-\rho_m^2} \, . \tag{14}$$

The transition function is calculated with the same expression defined by *Ivanov et al.* (2007):

$$J(t) = \big(1 - e^{-\varsigma(t-t_0)}\big)\big(1 - e^{-\gamma(t_0+t_b-t)}\big) \, , \tag{15}$$

where t_0 is the time at which inter-storm period begins and t_b is the length of the inter-storm period. The decay coefficients ς and γ are taken equal and are calculated with the procedure proposed by *Curtis and Eagleson* (1982). Assuming symmetry for the transition period, the second term of equation (15) can be neglected. Further, equation (15) can be reduced to $0.99 = \big(1 - e^{-\varsigma(T_R)}\big)$, when $t_0 = 0$ and $J(t) = 0.99$ for $t = T_R$. It follows that $\gamma = \varsigma = 4.61/T_R$.

The differences with the formulation of *Ivanov et al.* (2007) include the relaxation of the requirements of the minimum length of inter-storm period between two successive precipitation events; and the explicit computation of cloudiness in the first hours following and preceding rainfall spells J_1, instead of using a theoretical value equal to 1.

The parameters required for the model are estimate monthly and are: M_0, σ_m^2, $\rho_m(1)$, $\gamma = \varsigma$, J_1, and eleven values of a and b. The procedure for the parameter estimation follows that of *Curtis and Eagleson* (1982) and *Ivanov et al.* (2007), with some modifications (see Appendix B). First, the threshold value T_R of the transition period is determined to identify the fair-weather region, i.e., the region where $N(t)$ is stationary. Once the fair-weather region is identified the parameters M_0, σ_m^2, $\rho_m(1)$, $\gamma = \varsigma$, J_1 are easily evaluated with conventional techniques. The value of the first hour of the transition period J_1 is obtained as the average of all the first and last hours of the inter-storm periods. The empirical random deviate $\varepsilon(t)$ are estimated

for fair weather region considering that when $t \in R_0$ the transition function $J(t) = 1$ and equation 11 reduces to $N(t) = M_0 + m(t)$. a and b are finally evaluated from $m(t)$ as a function of the discretized cloudiness [0:0.1:1] at the step before $N(t-1)$. This procedure might be regarded as over-parameterized. However, the complexity of the approach is required by the difficulty of simulating a stochastic process such as cloud cover. When time series of hourly cloudiness are available, the identification of all parameters is computationally efficient.

4.2 Results and validation

A comparison between the observed and simulated monthly distributions of cloud cover is shown in Figure 9 for the fair-weather period. The weather generator performs generally well, in reproducing both the shape of the probability density function and the seasonality of the process. A less than perfect agreement was noticed for few locations for summer months. This shortcoming was first noted by *Ivanov et al.* (2007) and is related to the non-stationarity in the cloudiness process when passing of atmospheric precipitation systems do not necessarily result in rainfall at a given location. The non stationarity of cloud cover occurrence in such periods cannot be identified from the weather generator and consequently biases in the mean value are produced. The shape of probability density functions of total cloud cover are also well reproduced by AWE-GEN (Figure: 10). The differences between the simulated and observed mean cloud cover are generally less than 0.05, although sometimes discrepancies around 0.1 are appreciable. This holds true also for other tested locations. The performance of simulating the total cloudiness is not as satisfying as the one obtained for the fair-weather period. This is due to a higher difficulty of reproducing the transition regions.

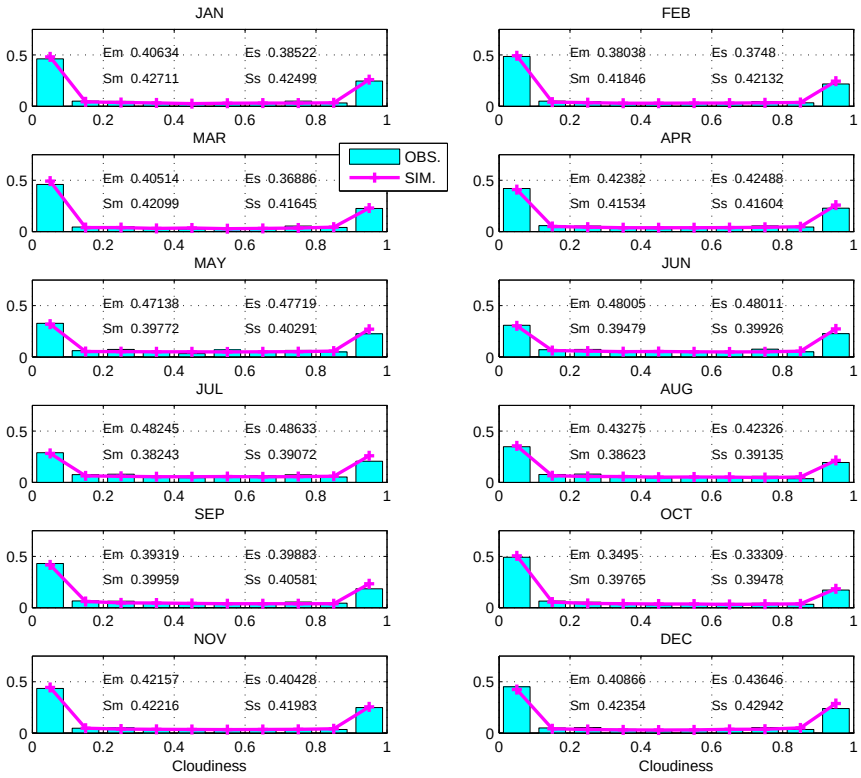


Figure 9: A comparison between the observed (cyan) and simulated (magenta) fair weather cloud cover distribution for every month. E_{obs} and σ_{obs} are the observed mean and standard deviation and E_{sim} and σ_{sim} are the simulated ones.

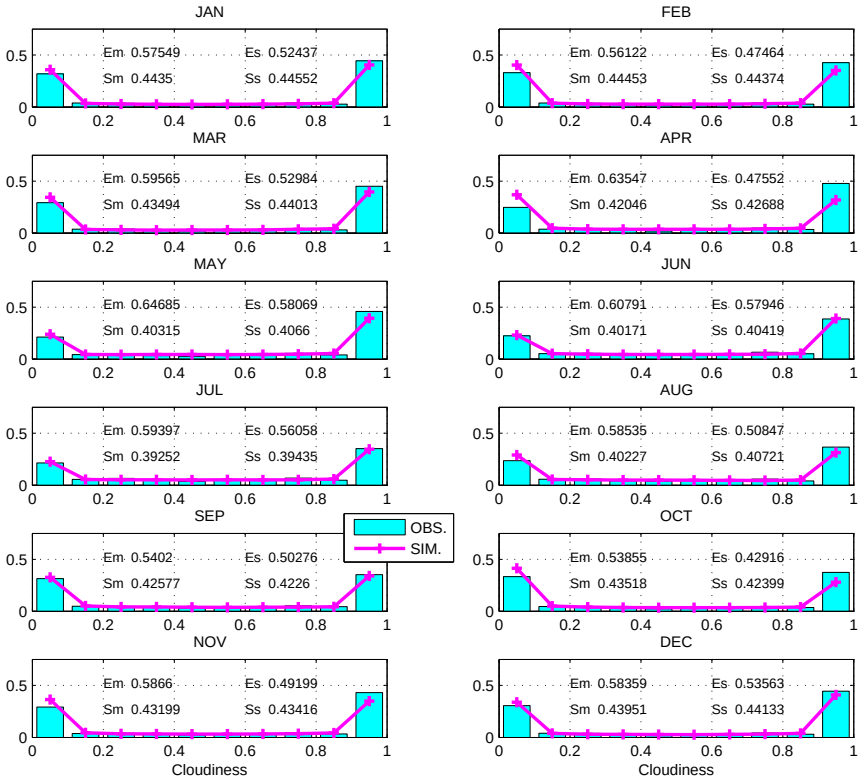


Figure 10: A comparison between the observed (cyan) and simulated (magenta) total cloud cover distribution, for every month. E_{obs} and σ_{obs} are the observed mean and standard deviation and E_{sim} and σ_{sim} are the simulated ones.

5 Air temperature

In weather generators, air temperature is commonly simulated at the daily scale, implying the generation of only maximum and minimum temperature or, alternatively, the temperature mean and its daily range (*Richardson*, 1981; *Semenov et al.*, 1998; *Wilks and Wilby*, 1999; *Parlange and Katz*, 2000). Typically, air temperature or its residuals are simulated through multi-regression equations between air temperature and other variables. Lagged air temperature can be included to take into account the autocorrelation process. Consequently the effects of climate variables such as cloudiness, are considered only indirectly, i.e., in the use of different parameterizations or equations for wet or dry states. Although such an approach can reproduce the mean and the variance accurately, it is not suitable for applications that require information on intra-daily air temperature variation. For these reasons, a mixed physics-based stochastic approach was developed by *Curtis and Eagleson* (1982) and later enhanced by *Ivanov et al.* (2007). This work utilizes the same approach with some further improvements.

5.1 Model

The generation of air temperature, $T(t)$ $[^{\circ}C]$, is simulated as the sum of a stochastic component, $dT(t)$ $[^{\circ}C]$, and a deterministic component, $\widetilde{T}(t)$ $[^{\circ}C]$:

$$T(t) = \widetilde{T}(t) + dT(t)\,. \tag{16}$$

The deterministic component of air temperature, $\widetilde{T}(t)$, is assumed to be directly related to the underlying physical processes such as the divergence of radiative and eddy heat fluxes. More specifically the deterministic time-gradient of temperature, $d\widetilde{T}(t)/dt$, is a function of the air temperature itself and of the incoming long-wave radiation. It is further related through two functions to the Sun's hourly position and site geographic location (*Curtis and Eagleson*, 1982; *Ivanov et al.*, 2007). Thus the deterministic component, $\widetilde{T}(t)$, is expressed as follows:

$$\frac{d\widetilde{T}(t)}{dt} \;=\; b_0 - b_1\widetilde{T}(t) + b_2K(t)s(t) + b_3K(t)r(t) + b_4q(t)\,, \tag{17}$$

where b_i $(i = 0,\ 1,\ \ldots,\ 4)$ are the five regression coefficients of the model, $q(t) = L_{atm}/1000$ $[W\ m^{-2}]$ is a scaled incoming long-wave radiation, L_{atm} $[W\ m^{-2}]$, and $K(t) = 1 - 0.75N^{3.4}$ $[-]$ is the cloud attenuation factor defined by *Kasten and Czeplak* (1980). The longwave radiation L_{atm} $[W\ m^{-2}]$ is modeled using the air temperature:

$$L_{atm} \;=\; K_N(N)\sigma T_a^4\,, \tag{18}$$

where T_a $[K]$ is the air temperature at the reference height, $\sigma = 5.670410^{-8}$ $[W\ m^{-2}\ K^{-4}]$ is the Stefan-Boltzmann constant; and $K_N(N) = 1 + 0.17N^2$ is the correction for the cloudiness N $[-]$ (*TVA*, 1972). The terms $r(t)$ $[-]$ and $s(t)$ $[-]$ are functions of

the solar height, h_S $[rad]$, defined by *Curtis and Eagleson* (1982):

$$\begin{aligned} s(t) &= sin(\delta)sin(\Phi) - cos(\delta)cos(\phi)cos(\frac{\pi t}{12})\,, \quad T_{H\,rise} \leq t \leq T_{H\,set}\,, \\ s(t) &= 0\,, \qquad\qquad\qquad\qquad\qquad\qquad \text{otherwise}\,, \\ r(t) &= \frac{ds(t)}{dt} = \frac{\pi}{12}cos(\delta)cos(\Phi)sin(\frac{\pi t}{12})\,, \quad T_{H\,rise} \leq t \leq 12\,, \\ r(t) &= 0\,, \qquad\qquad\qquad\qquad\qquad\qquad \text{otherwise}\,, \end{aligned} \tag{19}$$

where t is the local hour, δ $[rad]$ is the solar declination, Φ $[rad]$ is the local latitude, $T_{H\,rise}$ $[local\ hour]$ is the local time of sunrise and $T_{H\,set}$ $[local\ hour]$ is the local time of sunset. For details on the calculation of these quantities see the Appendix C. The factors $r(t)$, $s(t)$, and $q(t)$ are subjected to modification daily and seasonally and they explain the deterministic variation of air temperature. The differential equation, $d\widetilde{T}(t)/dt = f\big(\widetilde{T}(t), s(t), r(t), q(t), K(t)\big)$, is solved each day to compute the deterministic cycle of air temperature $\widetilde{T}(t)$ once the initial value of deterministic temperature $\widetilde{T}(t-1)$ is provided. *Curtis and Eagleson* (1982) provide a solution method of equation (17) summarized in Appendix D.

The stochastic temperature component, $dT(t) = T(t) - \widetilde{T}(t)$, is estimated through an autoregressive model AR(1). At the hourly scale, the random deviate of temperature exhibits a significant dependence in the hour of the day. Differences are noticeable in the statistics of $dT(t)$ for different phases of the day: morning, midday, afternoon, evening, and night. The stochastic component is particularly important for the determination of extreme of air temperature, such as minimum and maximum temperatures. Consequently, the average of the stochastic component $\overline{dT}_h$, and its standard deviation $\sigma_{dT,h}$ are estimated differently for each hour of the day $h \in [0, \ldots, 23]$. Note that this is an improvement in comparison to the original models of *Curtis and Eagleson* (1982) and *Ivanov et al.* (2007).

$$dT(t) = \overline{dT}_h + \rho_{dT}\big(dT(t-1) - \overline{dT}_h\big) + \varepsilon(t)\sigma_{dT,h}\sqrt{(1-\rho_{dT}^2)}\,, \tag{20}$$

where $\varepsilon(t)$ are the standard normal deviate, ρ_{dT} is the lag-1 autocorrelation of the process. The average, $\overline{dT}_h$, and the standard deviation, $\sigma_{dT,h}$, of $dT(t)$ depend on the hour of the day.

The coefficients and the parameters used to estimate the deterministic and stochastic components are evaluated at the monthly scale. *Ivanov et al.* (2007) describe the procedure for the estimation of the coefficients (see Appendix E). Once the regression coefficients are determined, the parameter: $\overline{dT}_h$, $\sigma_{dT,h}$, and ρ_{dT} are estimated from $dT(t)$ using conventional techniques. A constrain on $\rho_{dT} < 0.96$ is required to avoid numerical instability. Otherwise, it has been observed that combinations of random numbers might lead to unrealistic values of temperature.

5.2 Results and validation

The assessment of the performance of an hourly weather generator should not be limited to the daily means, especially for the air temperature process. The reproduction of the daily cycle and minimum and maximum temperatures is indeed fundamental for evaluating its capability.

Figure 11 shows the seasonal variation of mean air temperature and its standard deviation at the two aggregation periods of 1 hour and 24 hours. The observed values are reproduced almost perfectly. Note that the mean does not change with aggregation period. The seasonal variability of daily maximum and minimum air temperatures including standard deviations are illustrated in Figure 12. These quantities are well captured by the weather generator, although the variances can be slightly overestimated or underestimated. The daily cycle and the probability density function of air temperature are also well reproduced, as shown in Figure 13.

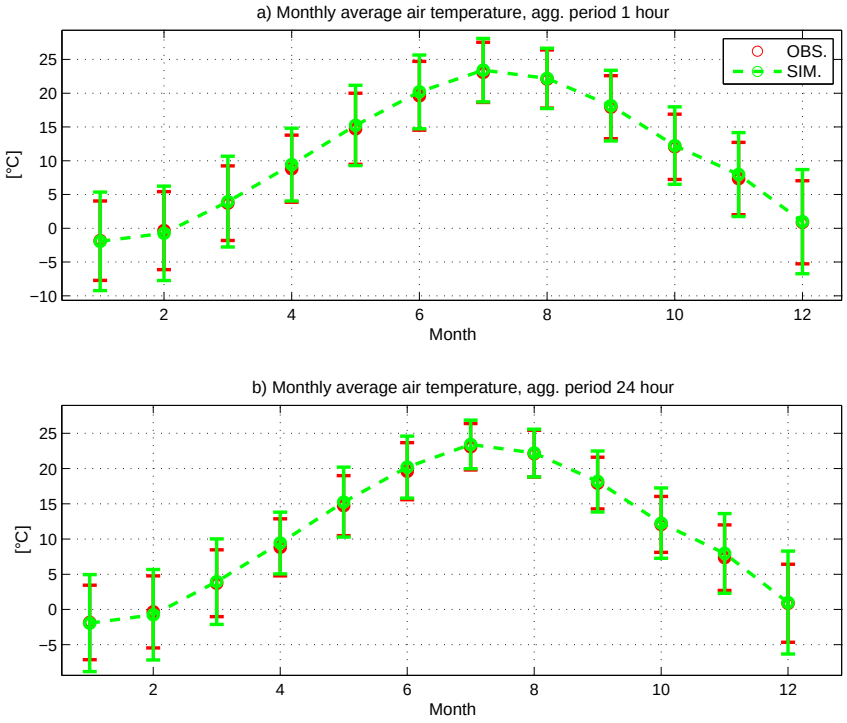


Figure 11: A comparison between the observed (red) and simulated (green) average air temperature for every month, aggregation periods of 1 $[h]$ (a) and 24 $[h]$ (b). The vertical bars denote the standard deviations.

Air temperature extremes at different return periods are reproduced satisfactorily, though overestimation or underestimation are often present both for minimum and maximum temperature. As seen in Figure 14 extremes of daily temperature (24 hour aggregation period) are generally reproduced better than hourly values. Nonetheless errors around 2-4 $[°C]$ for return periods of more than 10 years are not unusual. This shortcoming can be considered fairly insignificant for most hydrological applications.

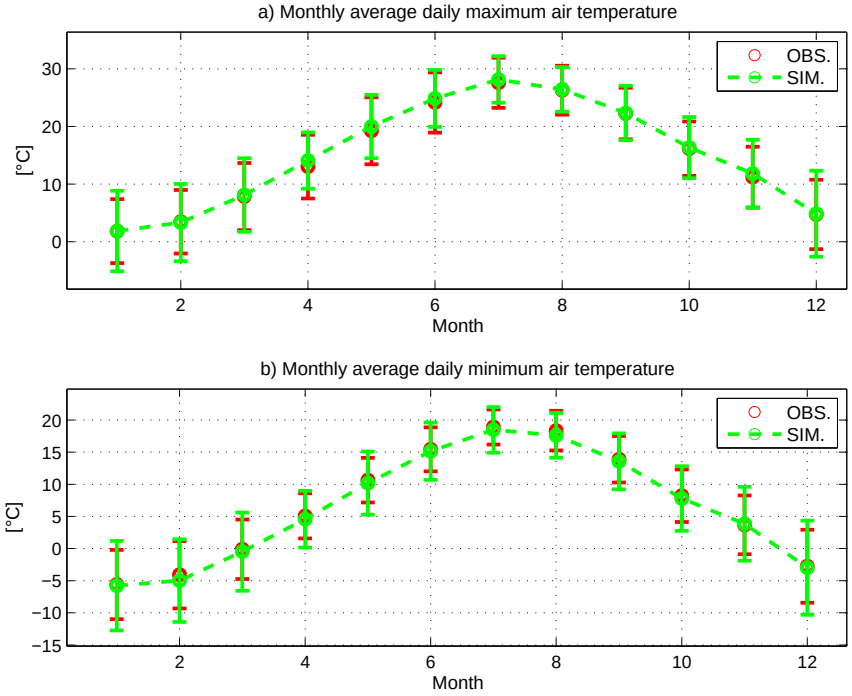


Figure 12: A comparison between the observed (red) and simulated (green) daily maximum (a) and minimum (b) air temperature for every month. The vertical bars denote the standard deviations.

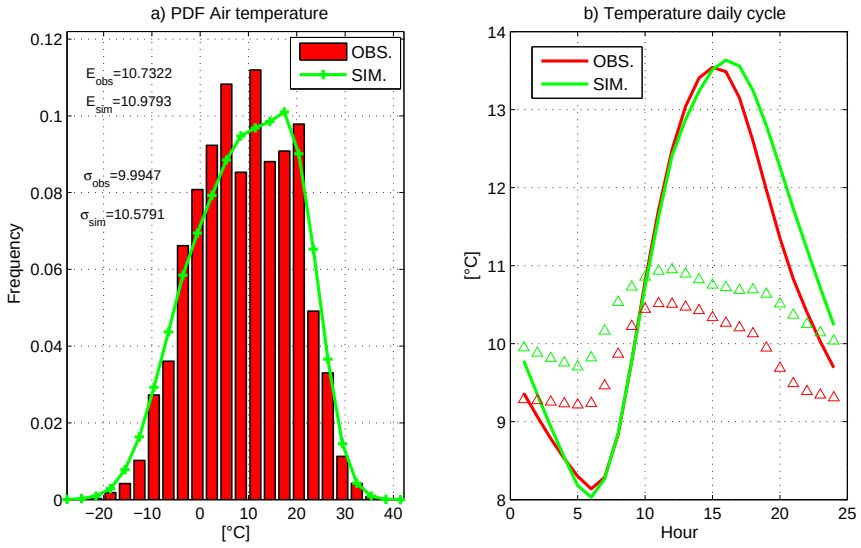


Figure 13: A comparison between the observed (red) and simulated (green) air temperature distribution (a) and average daily cycle (b). The triangles are the standard deviations for every day hour, E_{obs} and σ_{obs} are the observed mean and standard deviation and E_{sim} and σ_{sim} are the simulated ones.

Some problems might arise only when the introduction of temperature thresholds for plant mortality or other natural processes is required.

The occurrence of heat and cold waves, i.e. the number of consecutive days with air temperature higher than the 90th percentile (heat wave) or lower than the 10th percentile (cold wave) are poorly represented (Figure 15). There are differences in the accuracy of the results among the tested stations but generally, the temperature wave occurrence is underestimated. In order to simulate these climatic characteristics correctly, information about larger patterns of the atmospheric circulation is required. Obviously, a point scale weather generator cannot capture such features. Fortunately, the simulation of extreme heat and cold waves can be assumed to have a minor influence for numerous hydrological applications. Therefore, this shortcoming of AWE-GEN is acceptable in such applications. The estimation of heat and cold waves might be significant when climate change impacts on human health are required (*Rebetez et al.*, 2006), in this case other approaches are required.

An analysis has been also performed for four indices of air temperature as defined in the "Expert Team on Climate Change Detection, Monitoring and Indices" (*ETCCDMI*) (http://cccma.seos.uvic.ca/ETCCDMI/list_27_indices.shtml). These four indices allow an evaluation of the weather generator from a climatologist perspective. Specifically, results are compared between observation and simulations for icing days (days with $T_{max} < 0$ $[^{\circ}C]$), summer days (days with $T_{max} > 25$ $[^{\circ}C]$), frost days (days with $T_{min} < 0$ $[^{\circ}C]$) and tropical nights (days with $T_{min} > 20$ $[^{\circ}C]$). As shown in Table 2, AWE-GEN is able to reproduce such indices confirming its overall skill.

Table 2: Comparison between observed and simulated climatological indices of air temperature

	Observed fraction	**Simulated fraction**
Icing days	0.080	0.091
Summer days	0.198	0.207
Frost days	0.254	0.272
Tropical nights	0.078	0.070

The inter-annual variability of air temperature is neglected in this version of AWE-GEN. This is related to the difficulties in finding a proper external conditioning to reproduce low-frequency of the air temperature process. Nonetheless, the simulated annual temperature presents a certain variance, larger than 0, due to random number generation and to the influence of precipitation in the deterministic component of air temperature. The simulated variance is somewhat lower than the observed one because of the "overdispersion" issue already discussed in the description of annual precipitation (Section 1). Nevertheless, for several tested locations, the annual variance is only slightly underestimated and for typical applications the mismatch between the simulated and natural process can be neglected.

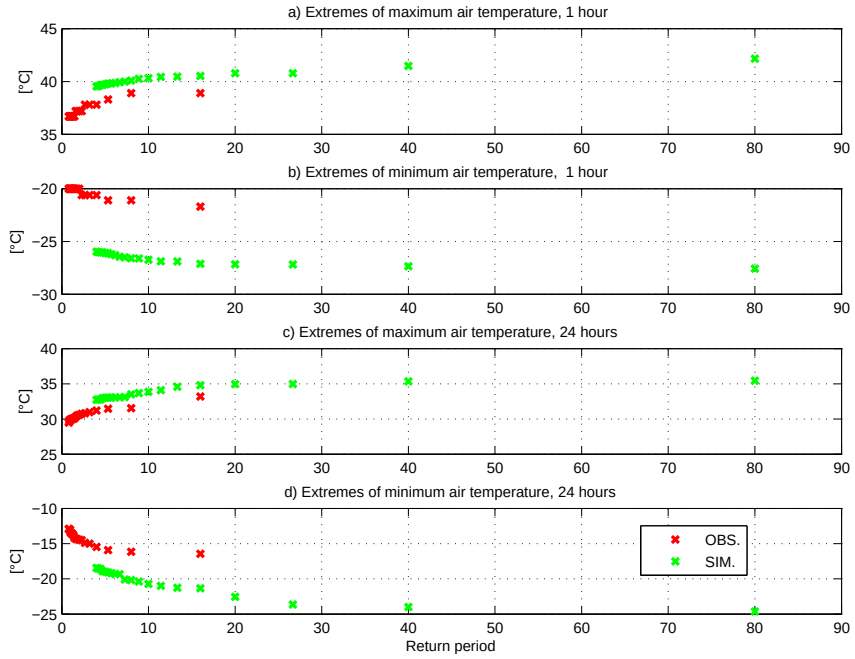


Figure 14: A comparison between the observed (red) and simulated (green) extremes of air temperature. a) Maxima of hourly temperature. b) Minima of hourly temperature. c) Maxima of daily temperature. d) Minima of daily temperature.

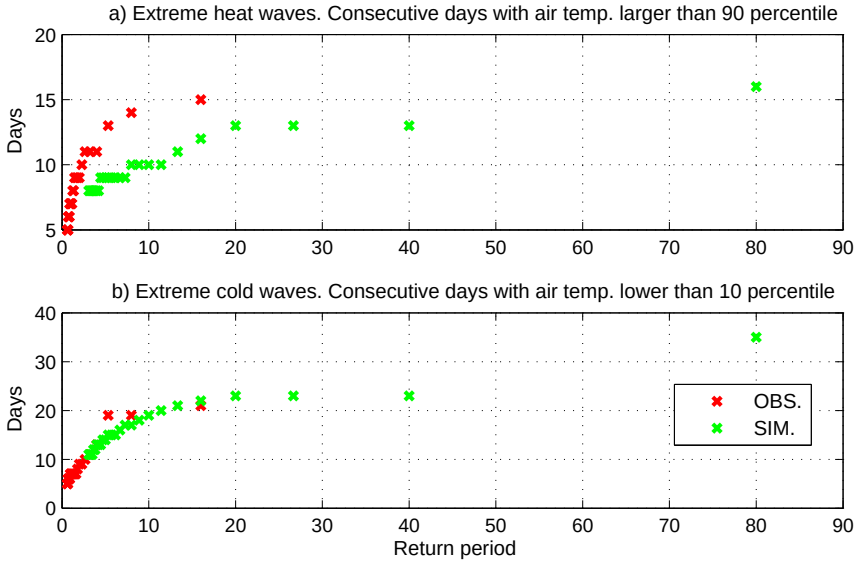


Figure 15: A comparison between the observed (red) and simulated (green) occurrence of heat (a) and cold (b) waves, i.e. consecutive days with temperature higher than the 90 percentile or lower than 10 percentile.

6 Shortwave incoming radiation

A correct estimation of the shortwave radiation is important because it represents the main source of incoming energy in the land-surface systems, directly affecting several ecological and hydrological processes. In weather generators, radiation is commonly estimated through regression with other variables (*Richardson and Wright*, 1984; *Parlange and Katz*, 2000). The likely reason for such an approach is a conventional lack of methodology for direct estimation of cloudiness. However, once the site geographic location and cloudiness are known, several deterministic models with different degrees of complexity can be used to calculate the incoming shortwave radiation for clear-sky and overcast conditions (*Gueymard*, 1989; *Freidenreich and Ramaswamy*, 1999; *Muneer et al.*, 2000; *Gueymard*, 2001, 2008; *Ineichen*, 2006). These methods recur to a large use of empirical coefficients to determine the atmospheric transmittances and the scattering fractions for direct and diffuse shortwave radiation. In this study, the incoming shortwave radiation is estimated with the model *REST2* developed by *Gueymard* (2008) for clear sky conditions. The parameterizations of *Stephens* (1978) and *Slingo* (1989) are used to compute transmittances for arbitrary cloudy conditions. This approach follows mainly the one proposed by *Ivanov et al.* (2007), improving the clear sky component where the model of *Gueymard* (1989) is substituted with the more recent model of *Gueymard* (2008). In hydrological applications only the global shortwave radiation component is typically considered, yet recent solar radiation models offer the convenience of shortwave flux computation in multiple-bands (*Freidenreich and Ramaswamy*, 1999; *Gueymard*, 2001). The partition of the incoming energy into different spectral bands could be useful for several purposes such as ecological or eco-hydrological simulations that require the photosynthetically active radiation, *PAR*, as input. Moreover using a multi-band approach allows one to minimize the overlapping effect between water vapor and gas, making the transmittance modeling more reliable (*Freidenreich and Ramaswamy*, 1999). The clear sky radiation component in AWE-GEN, considers two bands Λ: the ultraviolet/visible, *UV/VIS*, band with wavelengths $[0.29 - 0.70\ \mu m]$ and the near infrared, *NIR*, band with wavelengths $[0.70 - 4.0\ \mu m]$ (*Gueymard*, 2008). In the first band, ozone, nitrogen dioxide absorption, and Rayleigh scattering are concentrated; the absorption by water vapor and uniformly mixed gases is mainly concentrated in the second band.

The two-band model is a compromise between more complex formulations adopted in General Circulation Models and simple broadband approaches. This compromise allows one to compute explicitly the *PAR* without requiring a computationally intensive multiple-bands parametrization rather unadapt for applications of weather generators.

According to *Gueymard* (2008) the extraterrestrial radiation, R'_0, is partitioned in the fractions of 0.4651 in the *UV/VIS* band, and 0.5195 in the *NIR* band. The extraterrestrial radiation, R'_0, can be obtained starting with the value of the solar constant $R_0 = 1366.1\ [W\ m^{-2}]$, as suggested by *Darula et al.* (2005). This value is

corrected to take into account the ratio between the actual Earth-Sun distance and the mean Earth-Sun distance $R'_0 = E_0\ R_0$ $[W\ m^{-2}]$. The correction factor E_0 $[-]$ was derived by *Iqbal* (1983) as a function of the daily angle $\gamma = 2\pi(J_{Day} - 1)/365$:

$$\begin{aligned} E_0 &= 1.00011 + 0.034221\cos\gamma + 0.00128\sin\gamma \\ &\quad +0.000719\cos 2\gamma + 0.000077\sin 2\gamma\,, \end{aligned} \tag{21}$$

where J_{Day} is the Julian Day. The equations to calculate the instantaneous values of other variables used in the radiation computation, such as the solar altitude, h_S $[rad]$, solar azimuth, ζ_S $[rad]$, solar declination, δ $[rad]$, sunrise local time, $T_{H\,rise}$ $[local\ hour]$, sunset local time, $T_{H\,set}$ $[local\ hour]$, and daily length, D_{LH} $[h]$, are defined in the Appendix C. The equations are mainly drawn from *Iqbal* (1983) and *Eagleson* (2002).

6.1 Direct and diffuse radiation for clear sky conditions

When extraterrestrial radiation enters the atmosphere, it is attenuated by Rayleigh scattering $T_{R,\Lambda}$ $[-]$, uniformly mixed gas absorption $T_{g,\Lambda}$ $[-]$, ozone absorption $T_{o,\Lambda}$ $[-]$, nitrogen dioxide absorption $T_{n,\Lambda}$ $[-]$, water vapor absorption $T_{w,\Lambda}$ $[-]$, and aerosol extinction $T_{a,\Lambda}$ $[-]$ (*Gueymard*, 1989, 2008). The equations to compute the transmittance terms $T_{X,\Lambda}$ for both bands are given in *Gueymard* (2003, 2008) and in Appendix F.

The direct beam radiation at normal incidence, $R_{Bn,\Lambda}$ $[W\ m^{-2}]$, is computed for the first band *UV/VIS* $\Lambda 1$, and for the second band *NIR* $\Lambda 2$:

$$R_{Bn,\Lambda 1} = 0.4651 R'_0 \prod_X T_{X,\Lambda 1}\ , \tag{22}$$

$$R_{Bn,\Lambda 2} = 0.5195 R'_0 \prod_X T_{X,\Lambda 2}\,. \tag{23}$$

Following the model of *Gueymard* (2008) the incedent diffuse irradiance, $R_{Dp,\Lambda}$ $[W\ m^{-2}]$, on a perfectly absorbing ground (zero albedo) is defined in equation (24) where the prime indicate that the transmittances are calculated with a reference air mass $m' = 1.66$ $[-]$.

$$\begin{aligned} R_{Dp,\Lambda} &= T_{o,\Lambda} T_{g,\Lambda} T'_{n,\Lambda} T'_{w,\Lambda} \Big[B_{R,\Lambda}\big(1 - T_{R,\Lambda}\big) T_{a,\Lambda}^{0.25} \\ &\quad + B_a F_\Lambda T_{R,\Lambda}\big(1 - T_{as,\Lambda}^{0.25}\big) \Big] R'_{0,\Lambda} \sin(h_S)\,, \end{aligned} \tag{24}$$

where h_S $[rad]$ is the solar altitude, $B_{R,\Lambda}$ $[-]$ are the forward scattering fractions for Rayleigh extinction, B_a $[-]$ is the aerosol forward scattering factor, and F_Λ $[-]$ is a correction factor to compensate for multiple scattering effects and shortcomings of the simplified approach (*Gueymard*, 2008). The term $T_{as,\Lambda}$ $[-]$ is the aerosol scattering transmittance function of the single scattering albedos, $\omega_{\Lambda 1}$ $[-]$ and $\omega_{\Lambda 2}$ $[-]$, and of the spectral aerosol optical depth, $\tau_{a\Lambda}$ $[-]$ (*Gueymard*, 1989, 2008). For the parameterizations of the above quantities, see *Gueymard* (2008) and Appendix

F.

Backscattered radiation, $R_{Dd,\Lambda}$ $[W\ m^{-2}]$, must be added to the diffuse fluxes because of the interaction between the reflecting Earth surface and the scattering layer of the atmosphere. This component is computed as follows (*Gueymard*, 2008):

$$R_{Dd,\Lambda} = \rho_g \rho_{s,\Lambda} \big(R_{Bn,\Lambda} \sin(h_S) + R_{Dp,\Lambda} \big) / \big(1 - \rho_{g,\Lambda} \rho_{s,\Lambda} \big) \,, \tag{25}$$

where ρ_g $[-]$ is the ground albedo referring to a large area of 5-50 $[km]$ radius surrounding the point of interest and $\rho_{s,\Lambda}$ $[-]$ is the sky albedo, which is described in *Gueymard* (2008) (Appendix: F). Finally, the total diffuse radiation for clear sky conditions in each band is $R_{D,\Lambda} = R_{Dp,\Lambda} + R_{Dd,\Lambda}$ and the normal global radiation is $R_{Gn,\Lambda} = R_{Bn,\Lambda} + R_{D,\Lambda}$.

The parameters required for the clear sky radiation model of *Gueymard* (2008) are the ozone, u_o $[cm]$, and nitrogen dioxide, u_n $[cm]$, amounts in the atmospheric column, the single scattering albedos, $\omega_{\Lambda 1}$ $[-]$ and $\omega_{\Lambda 2}$ $[-]$, the surrounding ground albedo, ρ_g $[-]$, and the Ångström turbidity parameters, α_Λ $[-]$ and β_Λ $[-]$, from which the spectral aerosol optical depth, $\tau_{a\Lambda}$, can be obtained trough the Ångström equation:

$$\tau_{a\Lambda} = \beta_\Lambda \Lambda^{-\alpha_\Lambda} \,. \tag{26}$$

In the two band model, the wavelength Λ is substituted by an effective wavelength Λ_e for each of the two bands. Further α_Λ and β_Λ are taken equal for the two band (*Gueymard*, 1989, 2008). These parameters are not commonly available for typical applications of the weather generator. Nonetheless, the ranges of variation of several of these parameters are limited. In most cases typical values can be assumed. The value of the single scattering albedo ω_Λ is typically constrained between 0.75-0.98 for most applications (*Russell et al.*, 2002); *Gueymard* (2008) suggests to adopt a value of $\omega_{\Lambda 1} = 0.92$ and a value $\omega_{\Lambda 2} = 0.84$ when no-information is available. The ozone, u_o, and the nitrogen dioxide, u_n, amounts have a minimal influence in the overall process and constant values of 0.35 $[cm]$ and 0.0002 $[cm]$, respectively, are assumed in the weather generator. The surrounding ground albedo, ρ_g $[-]$, depends on the location but for snow-free region its value is typically between 0.1 and 0.25. The contribution of backscattered radiation is very small (*Gueymard*, 2008) and may become important only in snow-covered region, where ρ_g can reach the values of 0.7-0.85. The Ångström turbidity parameters α and β require a more detailed discussion. These parameters, especially β, have a strong effect in determining the clear sky irradiance. Suitable values of α and β can be derived from the spectral irradiance measurement, typically Aerosol Optical Depth (AOD) from n discrete bands using a linearization of Ångström equation (26) (*Gueymard*, 2008). The development of various sun-photometric ground networks, especially AERONET (http://aeronet.gsfc.nasa.gov/) (*Holben et al.*, 1998) has provided a large data archive of measured AOD and other atmospheric states with a

world-wide coverage. The possible values that the Ångström turbidity, α, can assume are 1.3±0.5. The parameter β, on the other hand, can vary several orders of magnitude reflecting sky conditions, from nearly zero (0.001 or less) for clear sky to 0.5 for very hazy conditions (*Chaiwiwatworakul and Chirarattananon*, 2004). The globally and annually averaged value of AOD (at [0.55 μm]) is about 0.12, that implies average values of β around 0.05 (*Ramanathan et al.*, 2001; *Robertson et al.*, 2001). When site-specific values of α and β are not available, the weather generator uses reference values: 1.3 for α (*Gueymard*, 1989) and β is calibrated to fit the average monthly clear sky radiation.

An example of the performance obtained with the model of *Gueymard* (2008) for clear sky condition is shown in Figure 6.1, where global, direct and diffuse broadband shortwave radiation are compared with the observations. The daily cycles of the clear sky shortwave radiation are reproduced satisfactorily for the different components, although midday differences typically in the order of [10−20] [$W\ m^{-2}$] are detectable. Generally, the results tend to underestimate the peaks of direct and diffuse radiation. Differences in the average monthly values are somewhat smaller.

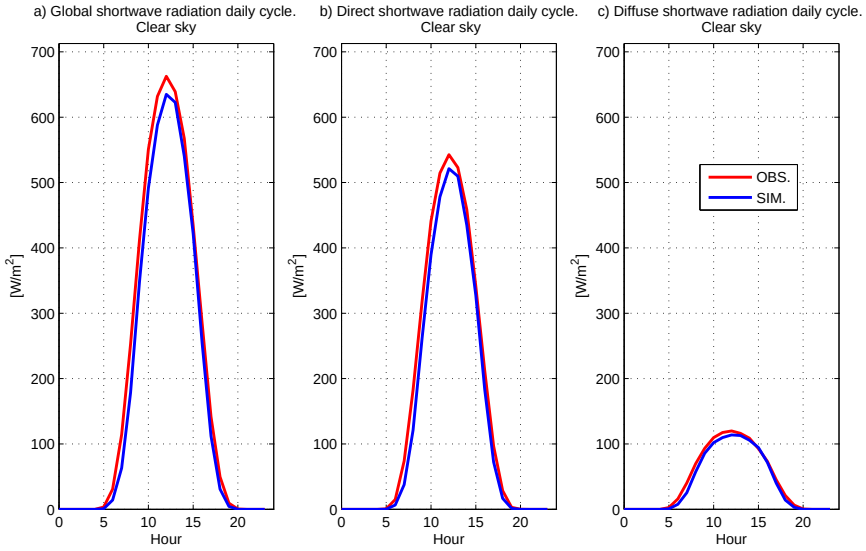


Figure 16: A comparison between the observed (red) and simulated (blue) daily cycle of global (a), direct (b) and diffuse (c) shortwave radiation for clear sky condition.

6.2 Direct and diffuse radiation for overcast conditions

Relative fluxes for cloudy conditions need to be addressed, after clear sky radiation fluxes are reproduced satisfactorily. Typically, in hydrological applications empirical equations relating the cloud cover, N [−], to the ratio between clear sky and total or partially overcast radiations were used to account for clouds effects (*Kasten and Czeplak*, 1980; *Becker*, 2001). Radiative properties of clouds are related to their type and structure. The latter should be taken into account through some parametrization. The approach described in *Ivanov et al.* (2007) is also employed in AWE-GEN

and uses the models developed by *Stephens* (1978) and *Slingo* (1989). These approaches argue that radiative properties of clouds are mainly related to the total vertical liquid water path, LWP $[g\ m^{-2}]$, which remains almost constant for clouds with the same broadband optical thickness (*Stephens*, 1978). Specifically *Slingo* (1989) simplified a multi-band cloud transmittance model to include only four wavelength bands, making the application suitable for weather generator purposes. This model parameterizes cloud transmittances for diffuse, $R_{D,\Lambda}$, and normal direct beam, $R_{Bn,\Lambda}$, clear sky fluxes, considering the latter normally incident on top of the clouds. *Slingo* (1989) accounted for four spectral bands λ, one in *UV/VIS* and three in *NIR* wavelength intervals: $[0.25-0.69\mu m], [0.69-1.19\mu m], [1.19-2.38\mu m], [2.38-4.0\mu m]$. The four band approach of *Slingo* (1989) can be transferred into the two band of *Gueymard* (2008) considering that the first bands of the two model *UV/VIS* almost coincide $\Lambda 1 \simeq \lambda 1$ and the second band $\Lambda 2$ is the sum of the other three bands $\Lambda 2 = \lambda 2 + \lambda 3 + \lambda 4$. The direct normal irradiance in each band $\lambda\,, = 1\,, 2\,, 3\,, 4$ for cloudy conditions, $\widetilde{R}_{Bn,\Lambda}$ $[W\ m^{-2}]$, is estimated as a linear combination of the fluxes from clear and cloudy fractions of the sky (*Slingo*, 1989):

$$\widetilde{R}_{Bn,\Lambda} = R_{Bn,\Lambda}\Big[(1-N) + T_{B,\lambda}N\Big]\frac{k(\lambda)}{K(\Lambda)}\,, \tag{27}$$

where $T_{B,\lambda}$ $[-]$ is the cloud transmissivity for direct beam flux in band λ, $k(\lambda)$ are the respective fractions of solar irradiance at the top of atmosphere in each band for *Slingo* (1989), [0.460 0.326 0.181 0.033]; and $K(\Lambda)$ are the fractions of solar radiation in the model of *Gueymard* (2008) [0.4651 0.5195]. Further details of the parametrization can be found in the auxiliary material of *Ivanov et al.* (2007) and in Appendix G of this work.

The diffuse radiative fluxes for cloudy conditions can result from the diffuse clear sky fraction and from the direct radiation incident on the clouds. The incident component of diffuse radiation in each band $\lambda\,, = 1\,, 2\,, 3\,, 4$ for cloudy conditions, $\widetilde{R}_{Dp,\Lambda}$ $[W\ m^{-2}]$, is estimated as a linear combination of the fluxes from clear and cloudy fractions of the sky (*Slingo*, 1989):

$$\widetilde{R}_{Dp,\Lambda} = (1-N)R_{Dp,\Lambda} + N\Big[T_{DB,\lambda}R_{Bn,\Lambda} + T_{DD,\lambda}R_{Dp,\Lambda}\Big]\frac{k(\lambda)}{K(\Lambda)}\,, \tag{28}$$

where $T_{DB,\lambda}$ $[-]$ and $T_{DD,\lambda}$ $[-]$ are the diffuse transmissivity for direct and incident diffuse radiation, respectively.

The backscattered contribution under a cloudy sky $\widetilde{R}_{Dd,\Lambda}$ $[W\ m^{-2}]$ is computed accounting for the effects of cloud transmittance:

$$\begin{aligned}\widetilde{R}_{Dd,\Lambda} \;=\; & \big[\rho_g\rho_{csB,\Lambda}/\big(1-\rho_{g,\Lambda}\rho_{csB,\Lambda}\big)\big]\,\widetilde{R}_{Bn,\Lambda}\sin(h_S) \\ & +\big[\rho_g\rho_{csD,\Lambda}/\big(1-\rho_{g,\Lambda}\rho_{csD,\Lambda}\big)\big]\,\widetilde{R}_{Dp,\Lambda}\,.\end{aligned} \tag{29}$$

The equation 29 has the same expression as 25, with the difference that the sky albedo for overcast or partially overcast conditions depends on the cloud albedo,

which is different for direct beam, $\rho_{csB,\Lambda}$ [−], and diffuse radiation, $\rho_{csD,\Lambda}$ [−]. The albedos, $\rho_{csB,\Lambda}$ and $\rho_{csD,\Lambda}$, are estimated as a linear combination of clear sky albedo, ρ_s [−], and diffuse reflectivity for direct and diffuse incident radiation, $A_{B,\lambda}$ [−], $A_{D,\lambda}$ [−]:

$$\rho_{csB,\Lambda} = (1-N)\rho_{s,\Lambda} + NA_{B,\lambda}\frac{k(\lambda)}{K(\Lambda)} \ , \tag{30}$$

$$\rho_{csD,\Lambda} = (1-N)\rho_{s,\Lambda} + NA_{D,\lambda}\frac{k(\lambda)}{K(\Lambda)} \, , \tag{31}$$

where the diffuse reflectivity for direct beam and diffuse incident radiation, $A_{B,\lambda}$, $A_{D,\lambda}$, are defined in *Slingo* (1989) (Appendix G) and are considered to be an approximation of cloud albedo.

The total diffuse radiation for cloudy sky is therefore: $\widetilde{R}_{D,\Lambda} = \widetilde{R}_{Dp,\Lambda} + \widetilde{R}_{Dd,\Lambda}$. The final value of the global radiation in each band Λ is $\widetilde{R}_{Gn,\Lambda} = \widetilde{R}_{Bn,\Lambda} + \widetilde{R}_{D,\Lambda}$. For arbitrary sky conditions global normal shortwave radiation is indicated as $R_{sw,n}$ [$W\ m^{-2}$], where $R_{sw,n} = \sum_\Lambda \widetilde{R}_{Gn,\Lambda}$, if $N > 0$, and $R_{sw,n} = \sum_\Lambda R_{Gn,\Lambda}$, if $N = 0$.

The described model requires cloud cover fractions and the cloud optical thickness, τ_N [−], which is essential for the description of the radiative properties of clouds (*Stephens*, 1978). The thickness τ_N can be approximately parameterized in terms of the liquid water path, LWP (*Stephens*, 1978). The cloudy sky condition is assumed to be characterized by a certain amount of $LWP(N)$ [$g\ m^{-2}$], which is estimated from a reference value of LWP for overcast conditions LWP_R [$g\ m^{-2}$] (*Ivanov et al.*, 2007):

$$LWP(N) = LWP_R\, N \, . \tag{32}$$

From equation (32), it follows that LWP varies from 0, when $N = 0$, to LWP_R ($N = 1$). Note that the exponential dependence of LWP on cloudiness N in *Ivanov et al.* (2007) has been replaced with a linear one, which leads to a ratio $R_{sw,n}(N)/R_{sw,n}(0)$ that better matches observations (*Kasten and Czeplak*, 1980). By evaluating different LWP_R for different months, this parametrization allows one to take into account the seasonal differences in cloud properties. In some circumstances LWP measurements or estimations could be also available and can be used directly.

The output of the radiation component of the weather generator contains the direct and diffuse radiation fluxes for the ultraviolet/visible *UV/VIS* band [$0.29 - 0.70\mu m$] and the near infrared *NIR* band [$0.70 - 4.0\mu m$]. As stated previously, the photosynthetically active radiation (*PAR*) can be important in several applications. The *PAR* radiation is the spectral range of solar light between 0.40 [μm] and 0.70 [μm]. This range does not coincide perfectly with the first radiation band *UV/VIS*. Reduction factors between the first radiation band and PAR are adopted (equations 33 and 34), as proposed by *Gueymard* (2008). The reduction factors, M_B [−] and M_G [−], are considered valid also for cloudy conditions, although the original formulation

of *Gueymard* (2008) was proposed only for clear sky conditions. This assumption should not introduce significant errors since the reduction factors depend only on Ångström turbidities, and on the air mass of aerosol extinction and of Rayleigh scattering. These factors are not expected to be modified under cloudy conditions.

$$\widetilde{PAR}_{Bn} = \widetilde{R}_{Bn,\Lambda 1}\, M_B \ , \tag{33}$$

$$\widetilde{PAR}_{D} = \widetilde{R}_{Gn,\Lambda 1}\, M_G - \widetilde{PAR}_{Bn} \, , \tag{34}$$

where $\widetilde{PAR}_{Bn}$ $[W\ m^{-2}]$ and $\widetilde{PAR}_{D}$ $[W\ m^{-2}]$ are the direct beam PAR at normal irradiance and the diffuse PAR, respectively. The parametrization for the two reduction factors, M_B and M_G, can be found in *Gueymard* (2008) and in Appendix F.

The spatial distribution of solar radiation over a surface is function of the surface geometry, i.e. of the local topography. Site slope, β_T $[rad]$, and aspect, ζ_T $[rad]$, can alter the daily distribution of incoming energy at the ground. Furthermore, the reflection and shadow effects of the surrounding terrain can strongly influence radiation fluxes. Obviously, the terrain effects are site-specific and are not accounted for in the weather generator. For a flat unobscured surface the only applicable adjustment is to multiply the $\widetilde{R}_{Bn,\Lambda}$ by the sine of the solar altitude h_S: $\widetilde{R}_{B,\Lambda} = \widetilde{R}_{Bn,\Lambda} \sin(h_S)$ in order to obtain the flux density for unit surface area. The same holds for clear sky conditions. When local topographic effects are non-negligible, $\sin(h_S)$ is substituted by a function of β_T and ζ_T (Appendix H). When the remote shading effect becomes important, the sky view factor, $S_{vf}(\vec{x})$, and the shadow effect, $S_h(\vec{x}, t)$, function at the position, $\vec{x}$, and the local time, t, should be introduced (*Olseth et al.*, 1995; *Kumar et al.*, 1997; *Dubayah and Loechel*, 1997; *Rigon et al.*, 2006; *Ivanov et al.*, 2007). Although topography effects cannot be accounted for directly by the weather generator, insights on the topographic effects on solar radiation are provided in Appendix H.

6.3 Results and validation

In all simulations, the stations are considered to be located on a flat surface, without topography-induced shadow or obstruction effects. The β Ångström turbidity parameter is calibrated monthly to fit the average value of global, direct, and diffuse clear sky radiation. The reference value of the liquid water path, LWP_R, is successively calibrated to fit the global, direct, and diffuse shortwave radiation for all sky conditions. The results obtained are shown in Figures 17, 18, and 19. The monthly average of shortwave radiation (Figure 17) is simulated properly, with occasional differences of 5-15 $[W\ m^{-2}]$. Such differences could be related to the higher frequency variability of parameters such as β Ångström turbidity or LWP_R. Weekly or daily variabilities of these parameters are not captured in AWE-GEN, as the relevant parameters are calibrated at the monthly scale. There is also an effect of error propagation from the simulation of the cloud process, that can worsen the results. Daily cycles of shortwave radiation are reproduced satisfactorily for different com-

ponents (Figure 18), although biases are present during mid-day hours for several stations. The weather generator tends to overestimate direct radiation and underestimate diffuse radiation, as shown in Figure 18 for Boston. However, it should be noted that the opposite holds true sometimes. Annual cycles of global radiation for different hours of local time are simulated very well, except for small deviations at the sunrise and sunset hours (Figure 19). This mismatch can be related to the reflection of beam radiation inducing radiative fluxes before sunrise and after sunset or to the measurement errors that are very likely to occur at low radiation density.

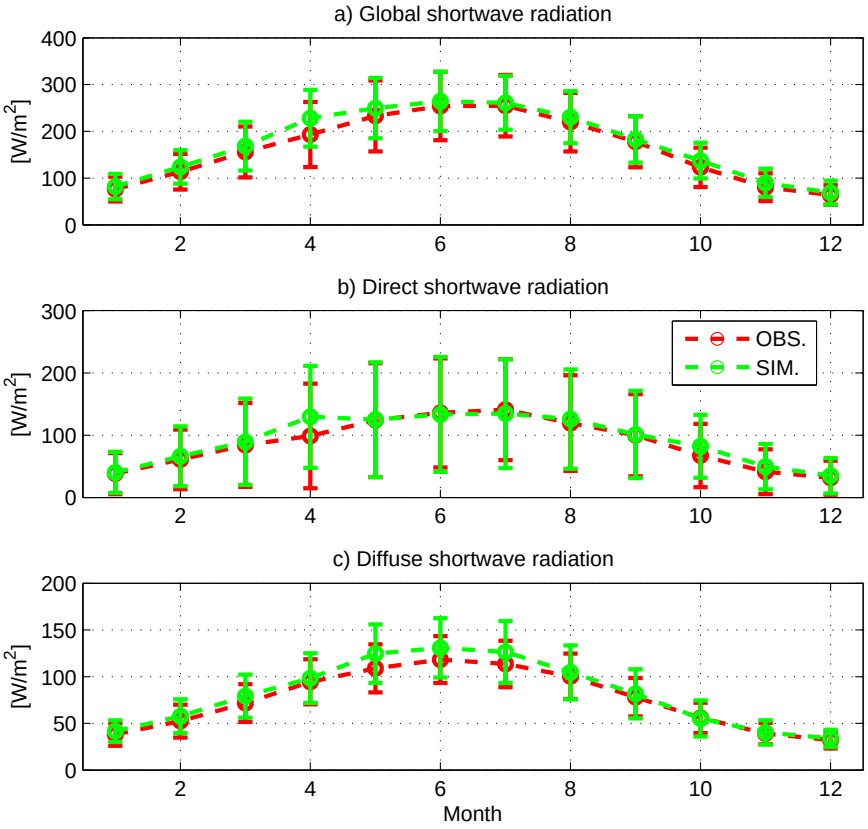


Figure 17: A comparison between the observed (red) and simulated (green) mean monthly shortwave radiation. a) Global radiation. b) Direct beam radiation. c) Diffuse radiation. The vertical bars denote the standard deviations of the monthly values.

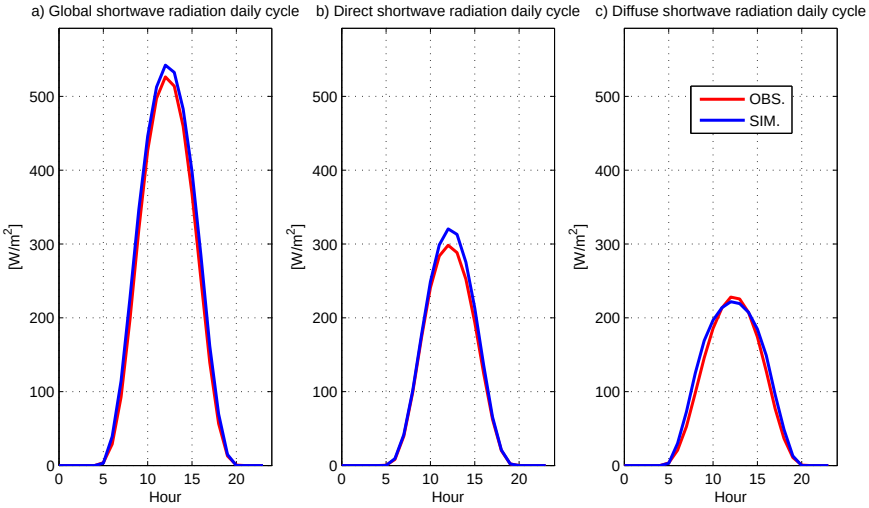


Figure 18: A comparison between the observed (red) and simulated (blue) daily cycle of global (a), direct (b) and diffuse (c) shortwave radiation for all sky conditions.

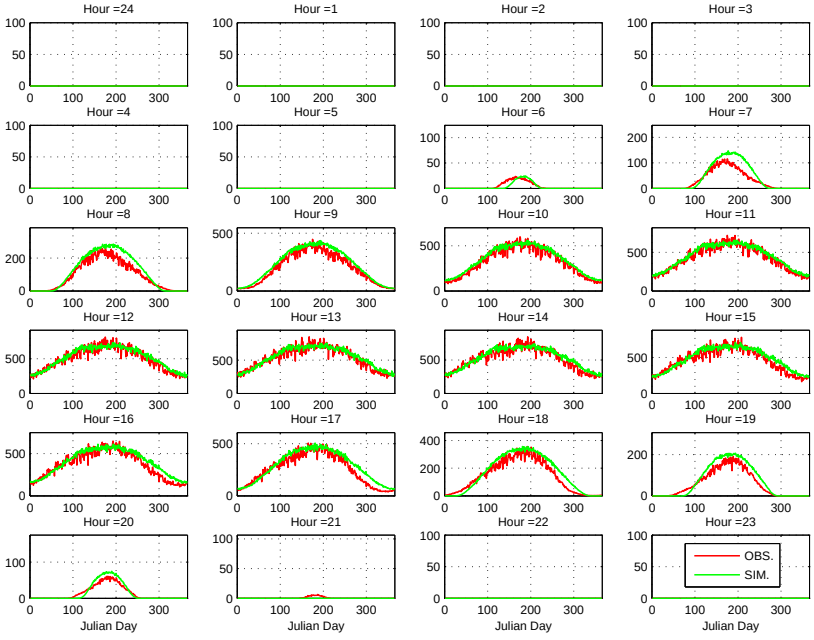


Figure 19: A comparison between the observed (red) and simulated (green) annual cycle of global shortwave radiation for different local time hours. The global shortwave fluxes are expressed in $[W\ m^{-2}]$.

7 Vapor pressure

Given the importance of vapor pressure for several hydrological and ecological applications, it needs to be included as one of the simulated variables. Vapor pressure is not commonly simulated by weather generators (*Semenov et al.*, 1998). Some weather generators include relative humidity (*Sharpley and Williams*, 1990; *McKague et al.*, 2003) or dew point temperature (*Parlange and Katz*, 2000; *Ivanov et al.*, 2007). Relative humidity or dew point temperature are generally estimated with a multi-regressive analysis (*Parlange and Katz*, 2000). An attempt to introduce a more physically-based approach was done by considering that dew point temperature is almost constant during the day time and has the tendency to come into equilibrium with nightly minimum temperatures (*Kimball et al.*, 1997). *Kimball et al.* (1997) pointed out that in arid and semiarid climate the dew point temperature could differ from nightly minimum temperature and proposed an empirical model to take into account these adjustments. A modified version of the same model was used to simulate dew point temperature by *Ivanov et al.* (2007). Typically, a daily or longer time step is used to simulate dew point temperature. *Curtis and Eagleson* (1982) proposed a multi-regressive model to simulate hourly dew point temperature for cases when its cross-correlation with other variables is non-negligible. While the conversion of relative humidity or dew point temperature into vapor pressure is mathematically straightforward, it involves non-linearity. Because of that, accurate simulations of dew point temperature or relative humidity do not necessary imply a good fit for vapor pressure. Dew point or relative humidity outputs of weather generators should be checked before asserting their suitability for applications that require vapor pressure.

7.1 Model

This study approaches the simulation of air humidity via the simulation of vapor pressure deficit, Δe $[Pa]$, i.e., the difference between the vapor pressure at saturation, e_{sat} $[Pa]$, and the air ambient vapor pressure, e_a $[Pa]$, where $e_{sat} = 611 \exp[17.27\, T_a/(237.3 + T_a)]$ $[Pa]$ (with T_a $[°C]$) is a well known expression (*Dingman*, 1994). Following *Bovard et al.* (2005), who noted a correlation between vapor pressure deficit, Δe, and PAR during daylight time, correlations of vapor pressure deficit, Δe, with shortwave radiation and temperature have been analyzed in this study. The vapor pressure deficit, Δe, shows a strong correlation with air temperature and a weaker correlation with global shortwave radiation lagged by several hours. Specific humidity and vapor pressure, e_a, remain almost constant throughout the day, especially in dry climates. Therefore variations of Δe and relative humidity, U $[-]$, are well explained by the diurnal cycle of air temperature. Specifically, there is a positive relation between the daily cycle of air temperature and the daily cycle of vapor pressure deficit. The assumption is primarily valid when the atmosphere is stable and exchanges between air masses with different characteristics are limited.

In order to simulate vapor pressure, a similar model framework as for the case of

air temperature is used: Δe is simulated as the sum of the deterministic component, $\widehat{\Delta e}$ $[Pa]$, and the stochastic component, $d\Delta e$ $[Pa]$:

$$\Delta e(t) = \widehat{\Delta e}(t) + d\Delta e(t)\,. \tag{35}$$

The deterministic component of vapor pressure deficit is related to air temperature through a cubic function, which is essentially an approximation of the commonly used exponential relation between T_a $[^\circ C]$ and e_{sat}. From observational data, a non-negligible correlation was also detected with global shortwave radiation, R_{sw} $[W\ m^{-2}]$, at lag one and two hours. The influence of solar radiation is generally minor, but it becomes important when air temperature effects are secondary. The deterministic component Δe is calculated with the equation:

$$\widehat{\Delta e}(t) = a_0 + a_1 T_a^3(t) + a_2 R_{sw}(t-1) + a_3 R_{sw}(t-2)\,, \tag{36}$$

where a_i $(i = 0,\, 1,\, \ldots,\, 3)$ are the regression coefficients. The deterministic component, $\widehat{\Delta e}$, usually shows a minor hourly variance, when compared with $\Delta e(t)$. The residuals $d\Delta e(t)$, that constitute the stochastic component of vapor pressure deficit, are modeled with the AR(1) approach in a similar fashion as for other variables:

$$d\Delta e(t) = \overline{d\Delta e} + \rho_{d\Delta e}\big(d\Delta e(t-1) - \overline{d\Delta e}\big) + \varepsilon(t)\sigma_{d\Delta e}\sqrt{(1-\rho_{d\Delta e}^2)}\,, \tag{37}$$

where $\overline{d\Delta e}$ is the average of vapor pressure deficit deviations, $\sigma_{d\Delta e}$ is the standard deviation and $\rho_{d\Delta e}$ is the lag-1 autocorrelation of the process. The terms $\varepsilon(t)$ are the standard normal deviates. Finally, the atmospheric vapor pressure, e_a, is calculated as the difference between e_{sat} and $\Delta e(t)$. It is possible that e_a calculated with the proposed procedure will assume values larger than e_{sat} and smaller than 0. Because of that, such values are simply corrected and assigned to the boundary values 0 and e_{sat}. This approximation introduces a bias in the e_a values near the limits. Nonetheless, this shortcoming is expected to slightly affect hydrological or ecological applications and furthermore could be corrected in future versions of the model.

The parameters of the deterministic component, a_i $(i = 0,\, 1,\, \ldots,\, 3)$, are estimated on a monthly basis using conventional regression techniques, for example, the least square approach. The parameters of the stochastic component: $\overline{d\Delta e}$, $\sigma_{d\Delta e}$, and $\rho_{d\Delta e}$ are evaluated using the time series of $d\Delta e(t)$ after removing the deterministic component from the observed series of Δe.

7.2 Results and validation

The performance of AWE-GEN in simulating metrics of air humidity is evaluated by investigating several statistical properties of vapor pressure, e_a $[Pa]$, relative humidity, U $[-]$, and dew point temperature, T_{dew} $[^\circ C]$. The capability of the model to reproduce the first two moments of vapor pressure at the time aggregation

periods of 1 hour and 24 hours is shown in Figure 20. Overall the performance is quite remarkable. The vapor pressure probability density function (Figure 21b) is also well simulated. In hydrologic applications, the simulation of the daily cycle of relative humidity is an important feature that affects evaporative flux estimation. In Figure 21a, the comparison between simulations and observations highlights a good overlap of the daily cycles, especially during day-time hours.

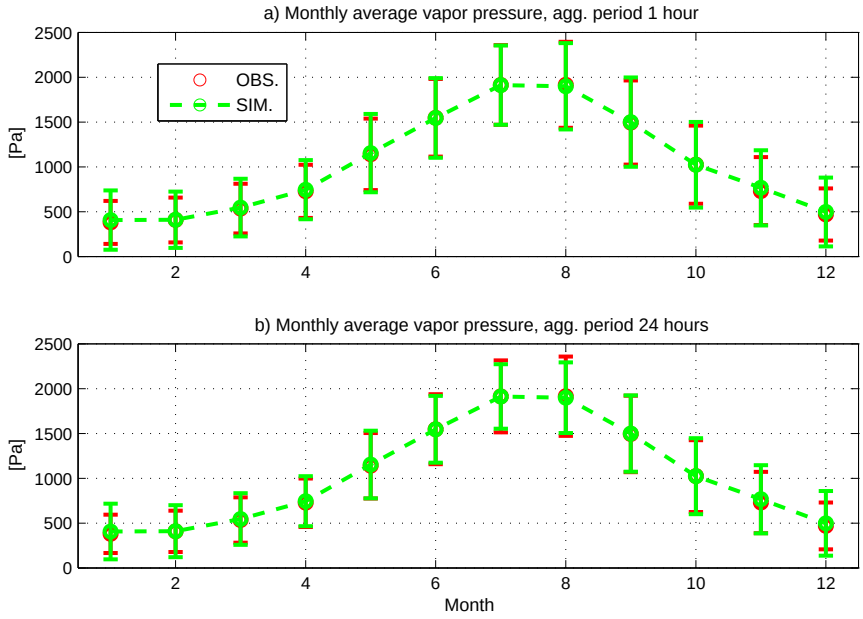


Figure 20: A comparison between the observed (red) and simulated (green) mean monthly vapor pressure for 1 $[h]$ (a) and 24 $[h]$ (b) aggregation time periods. The vertical bars denote the standard deviations of the monthly values.

The fitting of the probability density functions of relative humidity, U, and dew point temperature, T_{dew}, are shown in Figure 22. AWE-GEN reproduces these quantities satisfactorily. The relative humidity probability density in the upper and lower limits of the feasible range is overestimated. This holds true for several locations and is due to the overshot approximation. Differences in the tails of the T_{dew} distributions are rather frequent due to non-linearities in the transformation of e_a to T_{dew}.

The seasonality of mean relative humidity and its variance are also well captured (Figure 23), with only a slight underestimation of the 24 hours variance, especially in dry climates. Conversely, the dew point temperature is poorly simulated. The monthly variances simulated by AWE-GEN are usually larger than the observed ones (Figure 24a). This holds true for all the analyzed metrics of daily dew point temperature, i.e., mean, maximum, and minimum. Furthermore, the mean of daily maximum and minimum dew point temperature are typically overestimated and underestimated, respectively (Figure 24). Shortcomings in reproducing dew point temperature should be not an issue in many applications, since vapor pressure, e_a, and relative humidity, U, are typically required. When T_{dew} is the variable of inter-

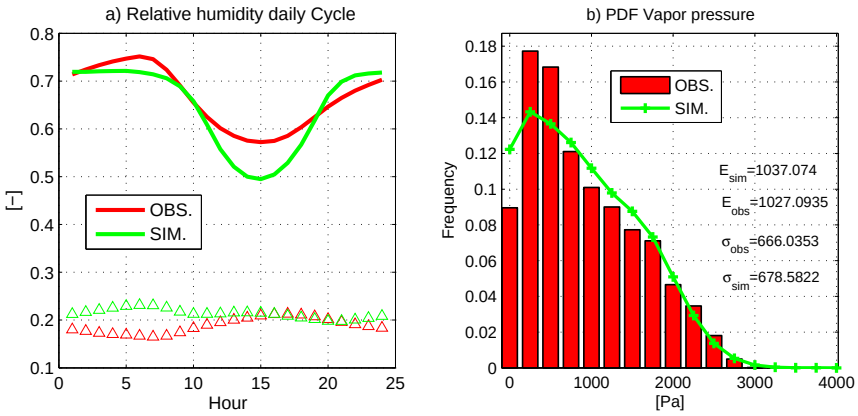


Figure 21: A comparison between the observed (red) and simulated (green) relative humidity daily cycle (a) and vapor pressure probability density function (b). The triangles in (a) represent the daily cycle of relative humidity standard deviation. E_{obs} and σ_{obs} are the observed mean and standard deviation and E_{sim} and σ_{sim} are the simulated ones.

est, the above limitations may become important and the suitability of simulated T_{dew} values must be checked according to the scope of the study.

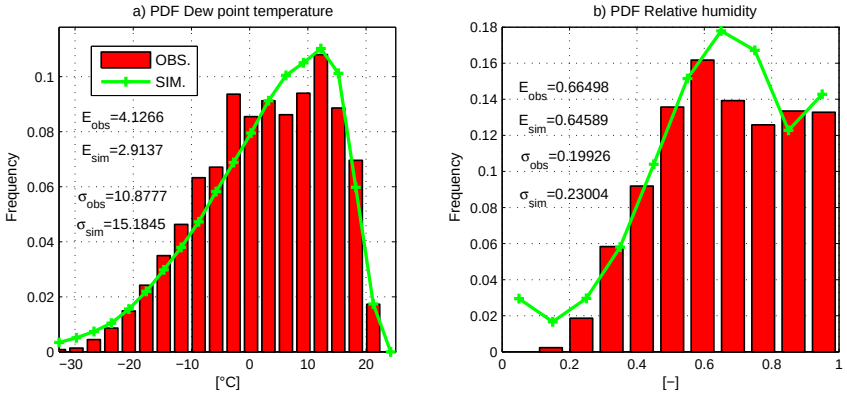


Figure 22: A comparison between the observed (red) and simulated (green) dew point temperature (a) and relative humidity (b) probability density functions. E_{obs} and σ_{obs} are the observed mean and standard deviation and E_{sim} and σ_{sim} are the simulated ones.

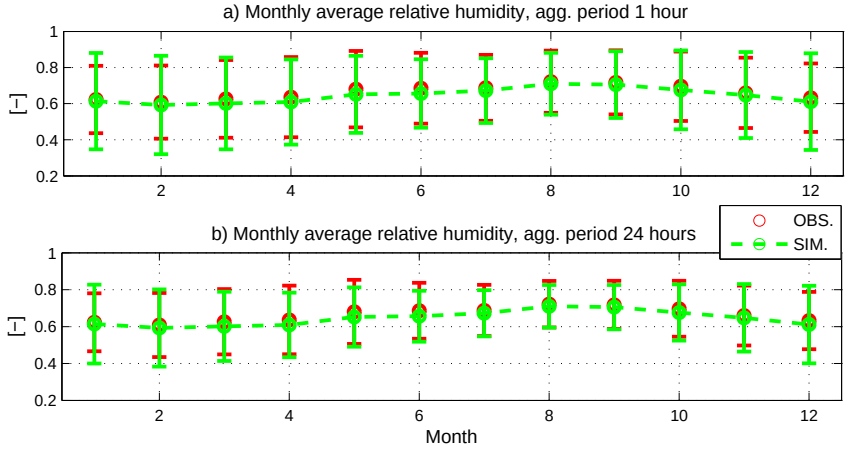


Figure 23: A comparison between the observed (red) and simulated (green) mean monthly relative humidity for aggregation periods of 1 hour (a) and 24 hours (b). The vertical bars denote the standard deviations of the monthly value.

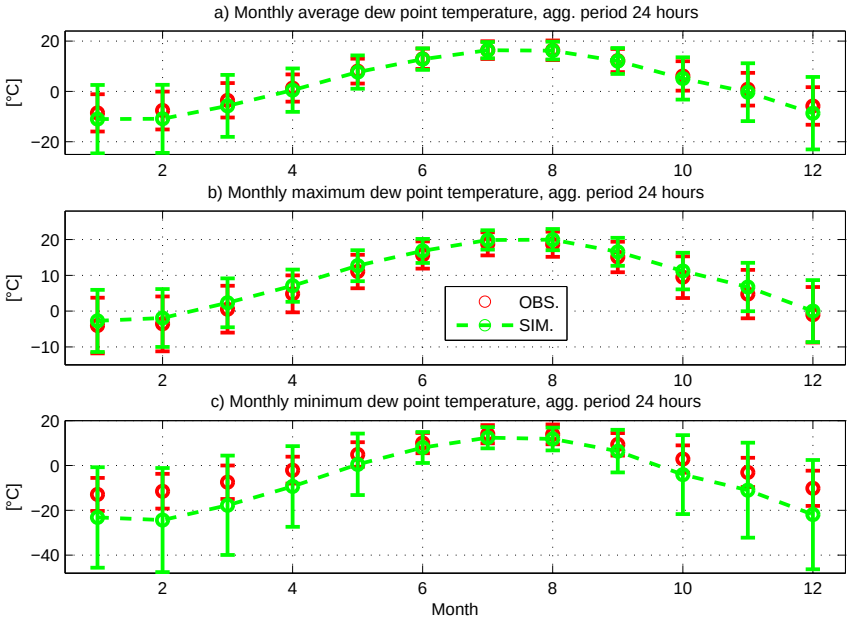


Figure 24: A comparison between the observed (red) and simulated (green) monthly dew point temperature for aggregation periods of 24 hours. a) Mean dew point temperature. b) Maximum dew point temperature. c) Minimum dew point temperature. The vertical bars denote the standard deviations of the monthly value.

8 Wind speed

Several studies highlight that cross-correlation between wind speed and other variables is typically very weak (*Curtis and Eagleson*, 1982; *Parlange and Katz*, 2000; *Ivanov et al.*, 2007) and thus wind speed is usually modeled as an independent variable. Nevertheless, in some locations wind speed exhibits a marked daily cycle and therefore the assumption of independence may need to be questioned. The inclusions of correlations among wind speed and other variables can be important because it allows the generator to capture wind speed intra-daily variations. The daily cycle of wind speed may affect the estimation of quantities such as the sensible and latent heat and is often required in hydrological and eco-hydrological modeling. The wind speed daily cycle at the ground surface is mainly related to the turbulent fluxes occurring in the surface boundary layer that are enhanced during the day-time by the the dissipation of sensible heat. The wind daily cycle is thus more pronounced in dry climates where the Bowen ratio is higher. Starting with this physical concept, the relation between the global solar radiation and wind speed has been investigated. It was found that the maximum correlation between the two cycles is usually shifted by several hours, possibly because of the different thermal properties of the ground surface and air. Correlation between different lags of global solar radiation and wind speed have been checked. The assumption of correlation between radiation and wind speed may become invalid for sites with strongly advective regime, e.g., when a site is located in a sea proximity, where the differential heating of surface affects the average daily cycle through sea breeze. Therefore, the weather generator is not expected to yield robust results for such locations with very complex wind speed daily cycles.

8.1 Model

The correlation coefficients found between wind speed, W_s $[m\ s^{-1}]$, and time shifted global solar radiation, R_{sw} $[W\ m^{-2}]$, are usually very small, however they are significant enough to induce a daily cycle in the wind speed component. Similarly to previously discussed approaches, the method to simulate the wind speed, W_s $[m\ s^{-1}]$, is based on representing the process as a sum of the deterministic, $\widehat{W_s}$, and the stochastic component, dW_s:

$$W_s(t) = \widehat{W_s}(t) + dW_s(t)\,. \tag{38}$$

The deterministic component, $\widehat{W_s}(t)$, relates the wind speed to the incident global shortwave radiation, R_{sw}. The correlation is shifted by several hours and the shift strongly depends on the site location. Lags up to three hours of R_{sw} are used to calculate the deterministic component of wind speed:

$$\widehat{W_s}(t) = c_0 + c_1 R_{sw}(t) + c_2 R_{sw}(t-1) + c_3 R_{sw}(t-2) + c_4 R_{sw}(t-3)\,, \tag{39}$$

where c_i ($i = 0, 1, \ldots, 4$) are the regression coefficients. The stochastic component, $dW_s(t) = W_s(t) - \widehat{W_s}(t)$, is modeled with the autoregressive AR(1) model including the Wilson-Hilferty transformation (*Wilson and Hilferty*, 1931; *Fiering and Jackson*, 1971). This transformation is necessary to represent the generally positive skewness exhibited by hourly wind speed data. A Weibull distribution is indeed very often assumed to model mean wind speed (*Takle and Brown*, 1978; *Deaves and Lines*, 1997). The stochastic component dW_s becomes:

$$dW_s(t) = \overline{dW_s} + \rho_{dW_s}\big(dW_s(t-1) - \overline{dW_s}\big) + \eta(t)\sigma_{dW_s}\sqrt{(1-\rho_{dW_s}^2)}\,, \tag{40}$$

where $\overline{dW_s}$ is the average wind speed deviation, σ_{dW_s} is the standard deviation, and ρ_{dW_s} is the lag-1 autocorrelation of the process. The term $\eta(t)$ represents the random deviate of the process that is skewed according to the Wilson-Hilferty transformation:

$$\eta(t) = \frac{2}{\gamma_n}\Big(1 + \frac{\gamma_n \varepsilon(t)}{6} - \frac{\gamma_n^2}{36}\Big)^3 - \frac{2}{\gamma_n}\,, \tag{41}$$

where the skewness of $\eta(t)$ is $\gamma_n = (1-\rho_{dW_s}^3)\gamma_{dW_S}/(1-\rho_{dW_s}^2)^{1.5}$; γ_{dW_s} is the skewness of the observed wind speed, and $\varepsilon(t)$ are the standard normal deviates.

The proposed approach also remains valid when the wind speed is an independent process; in this case the simulation will be dominated by the stochastic component, still producing consistent results.

The parameters c_i ($i = 0, 1, \ldots, 4$) of the deterministic component are estimated with conventional regression techniques. The parameters of the stochastic component: $\overline{dW_s}$, σ_{dW_s}, ρ_{dW_s}, and γ_{dW_S} are evaluated from the time series of $dW_s(t)$ after removing the deterministic component from the observed series of W_s. Wind speed generally does not present marked differences throughout the year, therefore the parameters are derived and assumed to be valid for all months.

8.2 Results and validation

The probability density function of wind speed is well captured in AWE-GEN as well as are the first two statistical moments of the process (Figure 25a). The wind speed daily cycle is reproduced correctly (Figure 25b) for almost all of the analyzed stations, and surprisingly also for stations located near sea. Nonetheless, it would not be surprising if the weather generator would perform poorly in some cases. The weather generator is unable to reproduce extremes, as seen in Figure 26. The hourly and daily extremes of wind speed are strongly underestimated, even for return periods lower than one year. This is not a problem for ecological or hydrological purposes but it implies that AWE-GEN cannot be used to generate meteorological forcing for structural design purposes where wind extremes are required.

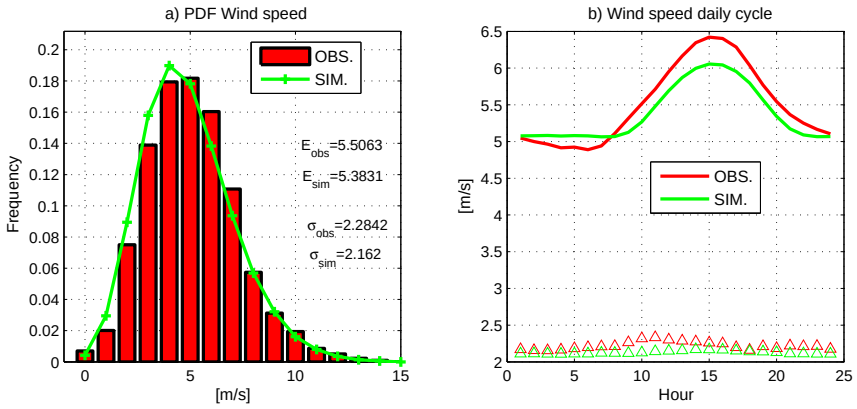


Figure 25: A comparison between the observed (red) and simulated (green) wind speed probability density function (a) and daily cycle of wind speed (b). E_{obs} and σ_{obs} are the observed mean and standard deviation and E_{sim} and σ_{sim} are the simulated ones.

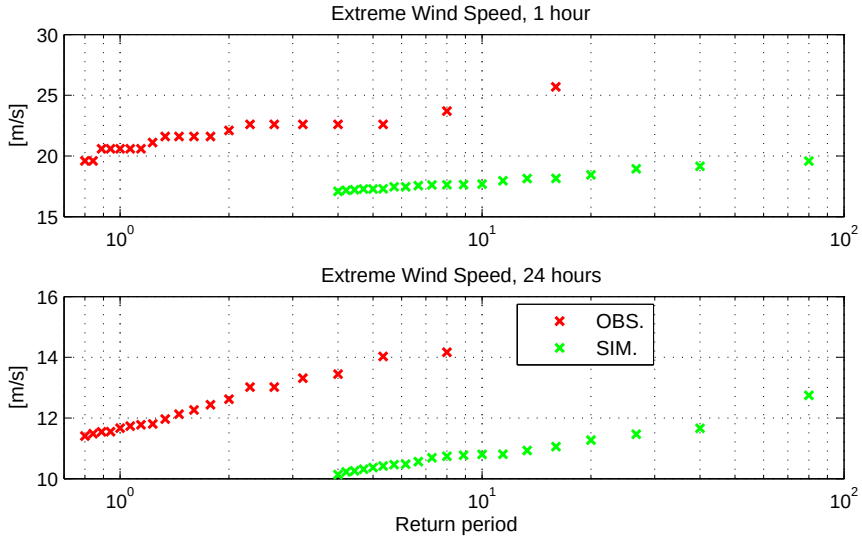


Figure 26: A comparison between the observed (red) and simulated (green) extremes of wind speed at aggregation periods of 1 hour (a) and 24 hours (b).

9 Atmospheric pressure

The atmospheric pressure, P_{atm} $[mbar]$, is generally neglected in weather generators given its relatively low impact on hydrological and ecological processes. However, P_{atm} is useful in many non-linear equations describing physical phenomena, such as evaporation. This observation implies that using a constant value of atmospheric pressure is theoretically incorrect. In AWE-GEN, a simple autoregressive model AR(1) is employed with parameters estimated to be valid for the entire year, thus neglecting the seasonal distribution of this variable.

$$P_{atm}(t) = \overline{P_{atm}} + \rho_{P_{atm}} \left(P_{atm}(t-1) - \overline{P_{atm}} \right) + \varepsilon(t) \sigma_{P_{atm}} \sqrt{(1 - \rho^2_{P_{atm}})}, \quad (42)$$

where $\overline{P_{atm}}$ is the average atmospheric pressure, $\sigma_{P_{atm}}$ is the standard deviation, $\rho_{P_{atm}}$ is the lag-1 autocorrelation of the process, and $\varepsilon(t)$ are the standard normal deviates. The shape of the atmospheric pressure distribution is perfectly reproduced as well as are the principal statistics (Figure 27).

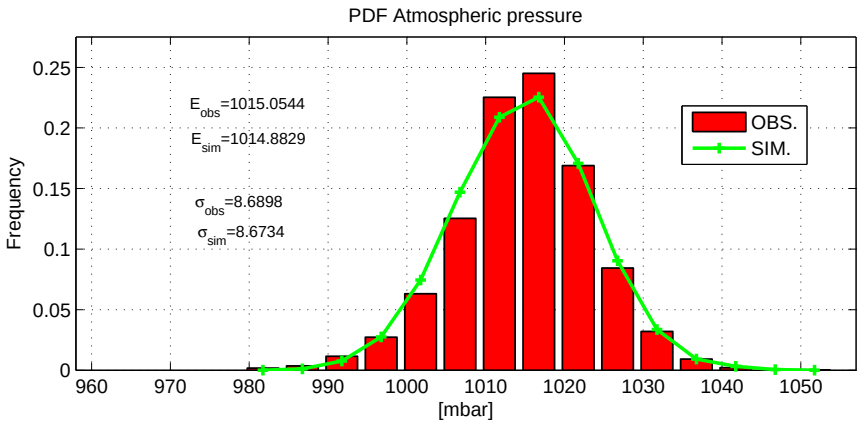


Figure 27: A comparison between the observed (red) and simulated (green) atmospheric pressure probability density function. E_{obs} and σ_{obs} are the observed mean and standard deviation and E_{sim} and σ_{sim} are the simulated ones.

The parameters of the model $\overline{P_{atm}}$, $\sigma_{P_{atm}}$, and $\rho_{P_{atm}}$ are evaluated from the time series of $P_{atm}(t)$ with conventional procedures.

10 Covariance between variables

The use of an intermediate physically-stochastic weather generator allows one to take into account mechanistic dependence between the meteorological variables, e.g precipitation vs cloudiness and, at the same time, directly introduces statistical correlations, e.g., vapor pressure vs temperature. Figure 28 illustrates a qualitative comparison of the interdependence between different variables. Precipitation occurrence affects cloud cover realizations, the latter process controls solar radiation and daily temperature range. Consequently, solar radiation and air temperature influence vapor pressure and wind speed calculation, generating a cascade of causal feedbacks that, starting from precipitation, affect all of the other variables.

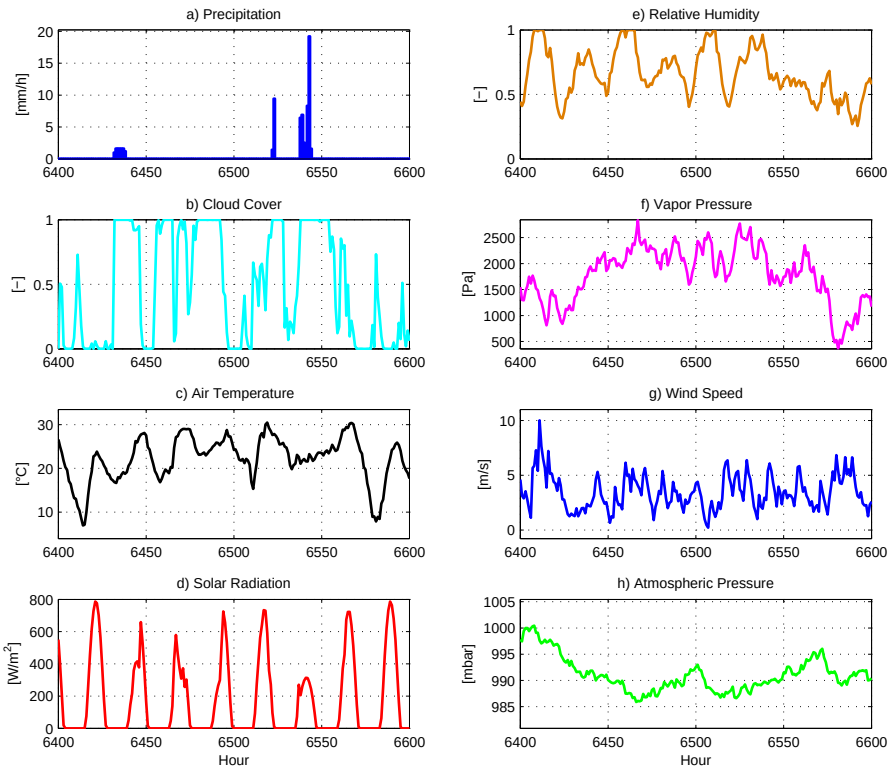


Figure 28: Simulated hourly values of hydro-meteorological variables with AWE-GEN: a.) precipitation; b.) cloud cover, c.) air temperature, d.) global shortwave radiation, e.) relative humidity, f.) vapor pressure, g.) wind speed, and h.) atmospheric pressure.

An explicit analysis of cross-correlation between climate variables is provided here. In Figure 29 a comparison between the observed and simulated mean monthly cloudiness and the number of wet days is shown. The interdependence between these two variables is generally captured in AWE-GEN but it cannot be appreciated for the discussed location. The performance in reproducing such a feature is well discernable for climates with a strong cloud cover seasonality, such as the climate corresponding to the location of San Francisco. Table 3 shows a synthesis of the mean values

Table 3: A comparison between the observed and simulated means of daily temperature amplitude, ΔT_{day}, daily global shortwave radiation, R_{sw}, and daily relative humidity, U, during rainy and rainless days.

Variable	**Rain days**		**Rainless days**	
	Measured	Simulated	Measured	Simulated
$\Delta T_{day}[°C]$	6.8	8.0	8.4	9.0
$R_{sw}[W\ m^{-2}]$	111	143	181	186
$U[-]$	0.80	0.65	0.61	0.64

of daily temperature amplitude, ΔT_{day}, daily global shortwave radiation, R_{sw}, and daily relative humidity, U, during rainy and rainless days. AWE-GEN tends to overestimate daily temperature amplitude and to underestimate the relative humidity during rainy days. These shortcoming are a consequence of the structure of the weather generator that only implicity accounts for the influence of rain in air temperature simulation. A cloud attenuation coefficient, $K(t)$, and incoming longwave radiation, L_{atm}, are the only variables connecting T_a to N. Moreover, this linkage is explicitly accounted for at the hourly and not at the daily scale. These dependencies in a humid temperate climate such as the one that characterizes the Boston area cannot fully explain the marked difference in ΔT_{day} between rainy and rainless days. The underestimation of daily relative humidity, U, in rainy days can be also explained by the same reasons and it is further affected by errors of daily temperature amplitude.

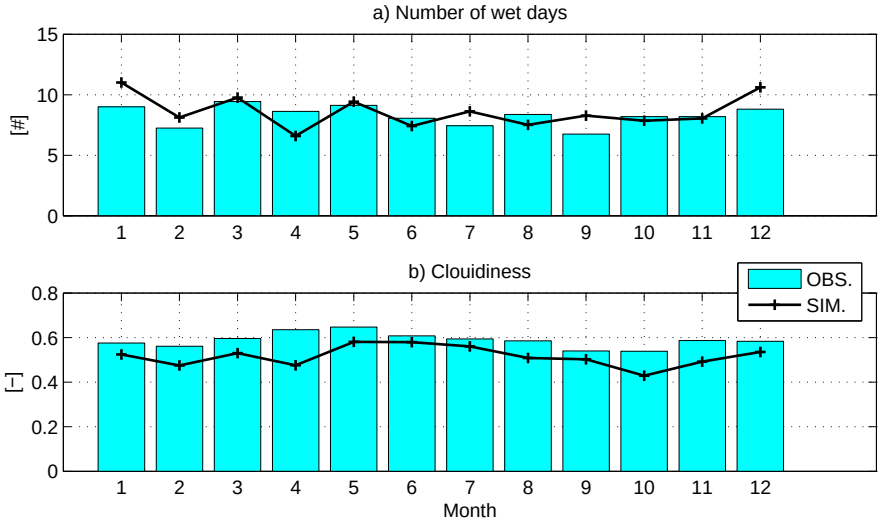


Figure 29: A comparison between the observed (cyan) and simulated (black) monthly number of wet days (a) and cloud cover (b).

Figure 30 shows an overview of cross-correlations at the daily scale between some of the variables. The principal cross-correlations are captured at lag-0. The hourly weather generator, unlike the empirical statistical weather generator, reproduces the cross correlations only implicity, especially at the daily time scale. Although overall some differences can be noticed, the results are considered quite satisfactory.

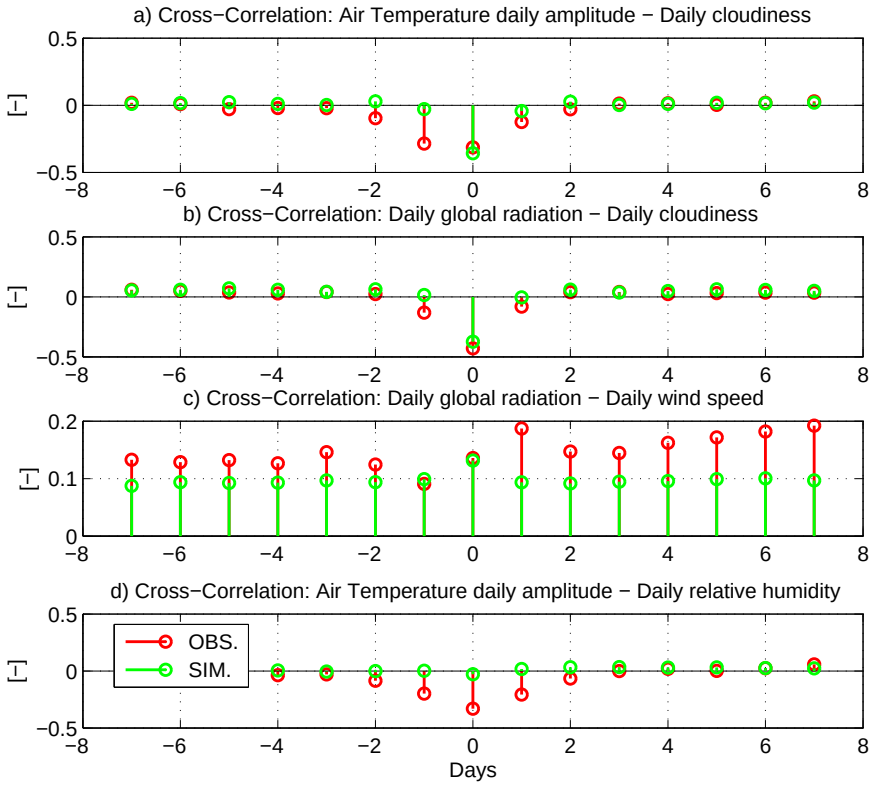


Figure 30: A comparison between the observed (red) and simulated (green) cross-correlation between: daily temperature amplitude and cloudiness (a), global solar radiation and cloudiness (b), global solar radiation and wind speed (c), and global solar radiation and relative humidity (d).

11 Application of AWE-GEN in climate change studies

The developed weather generator can be also used for climate change studies. There is the possibility to parameterize AWE-GEN on the basis of climate statistics that are not calculated from observations but are derived from a stochastic downscaling methodology. A detailed discussion about the use of the weather generator for the simulation of future climate scenarios, as inferred from climate models, can be found in *Fatichi et al.* (2011). Briefly, the stochastic downscaling procedure derives distributions of factors of change for several climate statistics from a multi-model ensemble of outputs of General Circulation Models (GCM) using a Bayesian approach. The methodology infers factors of change for precipitation and air temperature statistics, comparing realizations of climate models for two intervals of time. The two investigated periods are usually representing 20-40 years of simulated present climate conditions and 20-40 years of simulated future climate conditions. The factors of change are subsequently applied to the statistics derived from the observations to calculate statistics representative of the future climate conditions. Once all of the statistical properties are calculated for the future climate, these are used to re-evaluate the parameters of the weather generator. A new set of modified parameters of AWE-GEN is estimated. AWE-GEN is then used to simulate a scenario corresponding to future climate conditions (see *Fatichi et al.* (2011)).

For several reasons explained in *Fatichi et al.* (2011) the methodology is able to modify only a limited number of AWE-GEN parameters. Specifically, a new set of precipitation parameters: λ^{FUT}, β^{FUT}, η^{FUT}, μ_c^{FUT}, α^{FUT}, θ^{FUT} can be calculated as the final result of the stochastic downscaling procedure. Corrections are also provided for the inter-annual variability of precipitation, i.e., re-estimating $\overline{P}_{yr}^{FUT}$, $\sigma_{P_{yr}}^{FUT}$, and $\gamma_{P_{yr}}^{FUT}$. Finally, new parameters $\Delta T = T_{mon}^{FUT} - T_{mon}^{OBS}$ are introduced to account for changes in the air temperature. The new parameters ΔT_i $i = 1, ..., 12$ represent the variation of mean monthly temperature, T_{mon}, between the present and projected future climate conditions and can be used by AWE-GEN in the simulation of the the air temperature component (*Fatichi et al.*, 2011). Note that precipitation and air temperature affects directly or indirectly all of the other variables due to the imposed linkages within the weather generator. In such a way the information about the climate change can be also transferred to variables not directly accounted for in the downscaling.

APPENDIX

A Statistical properties of the *NSRP* model

Referring to Section 3.1 the coefficient $A(h,l)$ and $B(h,l)$ necessary to calculate the second moment of the Neymann-Scott rectangular pulse model are defined in *Cowpertwait* (1998) as:

$$\begin{aligned} A(h,l) &= h\eta + e^{-\eta h} - 1\,, \qquad \text{if } l = 0\,, \\ A(h,l) &= 0.5\left(1 - e^{-\eta h}\right)^2 e^{-\eta h(l-1)}\,, \quad \text{if } l > 0\,, \end{aligned} \tag{43}$$

$$\begin{aligned} B(h,l) &= h\beta + e^{-\beta h} - 1\,, \qquad \text{if } l = 0\,, \\ B(h,l) &= 0.5\left(1 - e^{-\beta h}\right)^2 e^{-\beta h(l-1)}\,, \quad \text{if } l > 0\,, \end{aligned} \tag{44}$$

where h is the time aggregation and $l \geq 0$ is a integer lag of the autocorrelation. The third moment, $\xi_h = E\{[Y_h - E\{Y_h\}]^3\}$, is also defined in *Cowpertwait* (1998):

$$\begin{aligned} \xi_h &= E\{[Y_h - E\{Y_h\}]^3\} = 6\lambda\mu_c\, E\{X^3\}(\eta h - 2 + \eta h e^{-\eta h} + 2e^{-\eta h})/\eta^4 \\ &\quad +3\,\lambda E\{X\}E\{X^2\}E\{C(C-1)\}f(\eta,\beta,h)/[2\eta^4\beta(\beta^2-\eta^2)^2] \\ &\quad +\lambda E\{X\}^3 E\{(C^2-C)(C-2)\}g(\eta,\beta,h) \\ &\quad /[2\eta^4\beta(\eta^2-\beta^2)(\eta-\beta)(2\beta+\eta)(\beta+2\eta)]\,, \end{aligned} \tag{45}$$

where the function $f(\eta,\beta,h)$ and $g(\eta,\beta,h)$ are listed below:

$$\begin{aligned} f(\eta,\beta,h) &= -2\eta^3\beta^2 e^{-\eta h} - 2\eta^3\beta^2 e^{-\beta h} + \eta^2\beta^3 e^{-2\eta h} + 2\eta^4\beta e^{-\eta h} \\ &\quad +2\eta^4\beta e^{-\beta h} + 2\eta^3\beta^2 e^{-(\eta+\beta)h} - 2\eta^4\beta e^{-(\eta+\beta)h} - 8\eta^3\beta^3 h + 11\eta^2\beta^3 - 2\eta^4\beta \\ &\quad +2\eta^3\beta^2 + 4\eta\beta^5 h + 4\eta^5\beta h - 7\beta^5 - 4\eta^5 + 8\beta^5 e^{-\eta h} - \beta^5 e^{-2\eta h} \\ &\quad -2h\eta^3\beta^3 e^{-\eta h} - 12\eta^2\beta^3 e^{-\eta h} + 2h\eta\beta^5 e^{-\eta h} + 4\eta^5 e^{-\beta h}\quad , \end{aligned} \tag{46}$$

$$\begin{aligned} g(\eta,\beta,h) &= 12\eta^5\beta e^{-\beta h} + 9\eta^4\beta^2 + 12\eta\beta^5 e^{-\eta h} + 9\eta^2\beta^4 + 12\eta^3\beta^3 e^{-(\eta+\beta)h} \\ &\quad -\eta^2\beta^4 e^{-2\eta h} - 12\eta^3\beta^3 e^{-\beta h} - 9\eta^5\beta - 9\eta\beta^5 - 3\eta\beta^5 e^{-2\eta h} \\ &\quad -\eta^4\beta^2 e^{-2\beta h} - 12\eta^3\beta^3 e^{-\eta h} + 6\eta^5\beta^2 h - 10\eta^3\beta^4 h + 6\eta^2\beta^5 h \\ &\quad -10\eta^4\beta^3 h + 4\eta\beta^6 h - 8\beta^2\eta^4 e^{-\beta h} + 4\beta\eta^6 h + 12\beta^3\eta^3 \\ &\quad -8\beta^4\eta^2 e^{-\eta h} - 6\eta^6 - 6\beta^6 - 2\eta^6 e^{-2\beta h} - 2\beta^6 e^{-2\eta h} \\ &\quad +8\eta^6 e^{-\beta h} + 8\beta^6 e^{-\eta h} - 3\beta\eta^5 e^{-2\beta h}\,. \end{aligned} \tag{47}$$

The probability that an arbitrary interval of length h is dry, $\Phi(h) = P(Y_h = 0)$, was derived from *Cowpertwait* (1991); *Cowpertwait et al.* (1996). It is here modified to take into account the use of the Geometrical distribution instead of the Poisson

distribution for the generation of the random number of cells:

$$\Phi(h) = \exp\Big(-\lambda\, h + \lambda\, \beta^{-1}\mu_c^{-1}[1 - e^{(-\mu_c + \mu_c e^{-\beta h})}] - \lambda \int_0^{\infty} \big[1 - p_h(t)\big] dt\Big), \quad (48)$$

where $p_h(t)$ is function of h, β, η and μ_c:

$$p_h(t) = \Big[e^{-\beta(t+h)} + 1 - (\eta e^{-\beta t} - \beta e^{-\eta t})/(\eta - \beta)\Big] \cdot \exp\Big[-\mu_c\beta(e^{-\beta t} - e^{-\eta t})/(\eta - \beta) - \mu_c e^{-\beta\, t} + \mu_c e^{-\beta(t+h)}\Big]. \quad (49)$$

B Cloud cover parameter estimation

This description of cloud cover parameter estimation follows *Ivanov et al.* (2007). The parameters used by the cloud cover model are M_0, σ_m^2, $\rho_m(1)$, $\gamma = \varsigma$, J_1, a, and b.

The existence of a stationary interstorm fairweather cloud cover process is the central assumption of the model. The identification of sequences of the fairweather periods in series of meteorological data therefore becomes essential. The methodology proposed by *Curtis and Eagleson* (1982) employs an iterative approach that uses records of the total cloud cover during periods between successive precipitation events. The essence of the method is in estimating the mean value of cloud cover for some sub-region Δt within an interstorm period (Figure 31).

Each interstorm period of length $T_{is} = \Delta t_0$ $[h]$ (Figure 31) is considered to be constrained by the last hour of the first rainfall event and by the first hour of the following rainfall event. By successively eliminating one hour from both ends of any given interstorm period ($\Delta\tau_1 = 1$ *hour*, $\Delta\tau_2 = 2$ *hours*, ...), a number of sub-regions, not exceeding in total $(T_{is}/2 - 1)$, can be defined for each interstorm period. For any given sub-region, Δt_k, corresponding to k number of eliminated hours from each end (Figure 31), a mean value of the cloud cover is estimated over all interstorm periods in the considered precipitation record whose duration exceeds $2k$ hours. Since $k \in [0, T_{is\,\max}/2 - 1]$, where $T_{is\,\max}$ is the maximum duration of an interstorm period in the considered record, a vector of the mean values of cloud cover of length $(T_{is\,\max}/2 - 1)$ is obtained.

Curtis and Eagleson (1982) argue that with the increasingly larger number of eliminated hours, the estimated mean value stabilizes, reaching some constant, or the fairweather mean value, M_0. The number of hours, T_r, eliminated from both ends of all interstorm periods (whose duration exceeds $2T_r$) after which there is no significant change in the mean cloudiness value, is considered to be the length of the *transition* period. Consequently, a necessary condition for an interstorm period to contain a fairweather cloud cover sequence is to be of duration $T_{is} > 2T_r$ $[h]$.

A note has to be made regarding a particular case of sub-regions within certain interstorm periods for which the described approach fails. Sometimes, passing atmo-

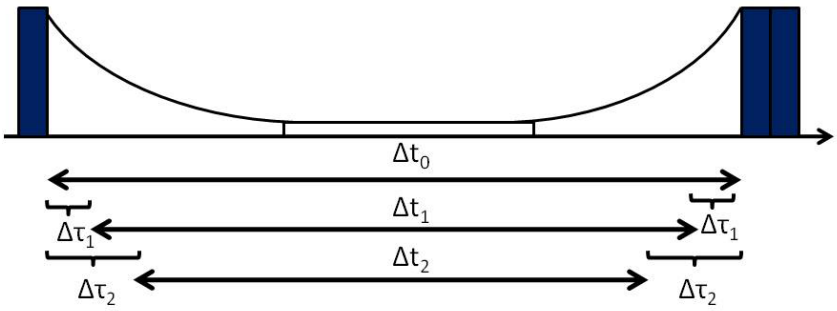


Figure 31: An illustration of the procedure used to identify the *fairweather* cloud cover period.

spheric precipitation systems do not necessarily result in rainfall at a given location. However, the cloud cover process is obviously non-stationary during such periods and the estimated mean value can be significantly affected. The discussed approach cannot identify such periods, which would, perhaps, require auxiliary information about cloud vertical structure and spatial information about the precipitation process. Nonetheless, the procedure is efficient for most of interstorm periods and results in reasonable estimates of the transition period as long as the above situation does not occur often. Caution has to be taken when interpreting the results of this method. Figure 32 illustrate the outlined procedure.

For the selected values of T_r, both the empirical and observed transition function, $J(t)$, are plotted in Figure 33. Some differences in comparison to the work of *Ivanov et al.* (2007) are discernible. The exponential form of $J(t)$ fits the observed cloud cover transition quite well in many months. The determination of the critical length, T_R, of the transition period in *Ivanov et al.* (2007) was left to the subjectivity of the user. T_R, indeed, is the length after which the fair weather region could be identified. In AWE-GEN, T_R is identified with an objective criterion. A threshold on the derivative of the smoothed mean cloud cover, $\widetilde{E\{N(t)\}}$, defines the begin of the fair weather region (Figure 32).

Once T_r is established, the fairweather sequences contained in the interstorm periods of length $T_{is} > 2T_r$ are combined in a new time series containing only fairweather cloud cover values. For these series, created for each month or the entire period of analysis, the parameters M_0, σ_m^2, $\rho_m(1)$ and J_1 are determined by conventional methods. The parameter $\gamma = \varsigma$, is estimated according to the equation proposed by *Curtis and Eagleson* (1982) (see also Section 4):

$$\gamma = \frac{4.61}{T_r}. \tag{50}$$

The parameters a and b are estimated by analyzing random deviates, $\varepsilon(t)$, which are computed from the observed cloud cover series by inverting equation (11) and (14). The estimation of $\varepsilon(t)$ is conditioned by the cloud cover at time $(t-1)$. Therefore, 11 vectors of deviates are composed from the cloud cover records in the different

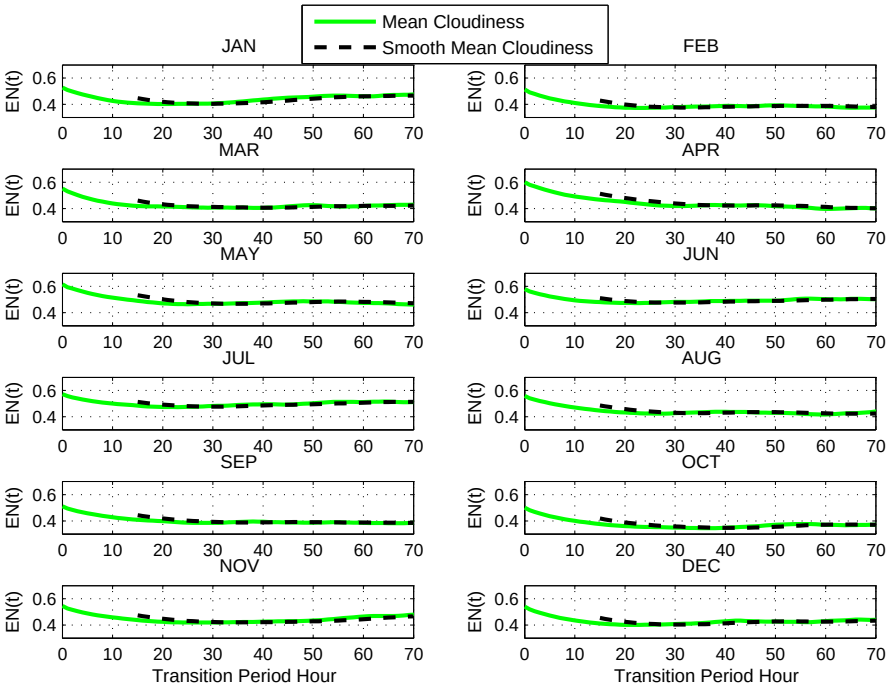


Figure 32: Estimated mean cloud cover value $E\{N(t)\}$ (continuous line) and the smoothed function $\widetilde{E\{N(t)\}}$ (dashed line) as a function of the length of transition period.

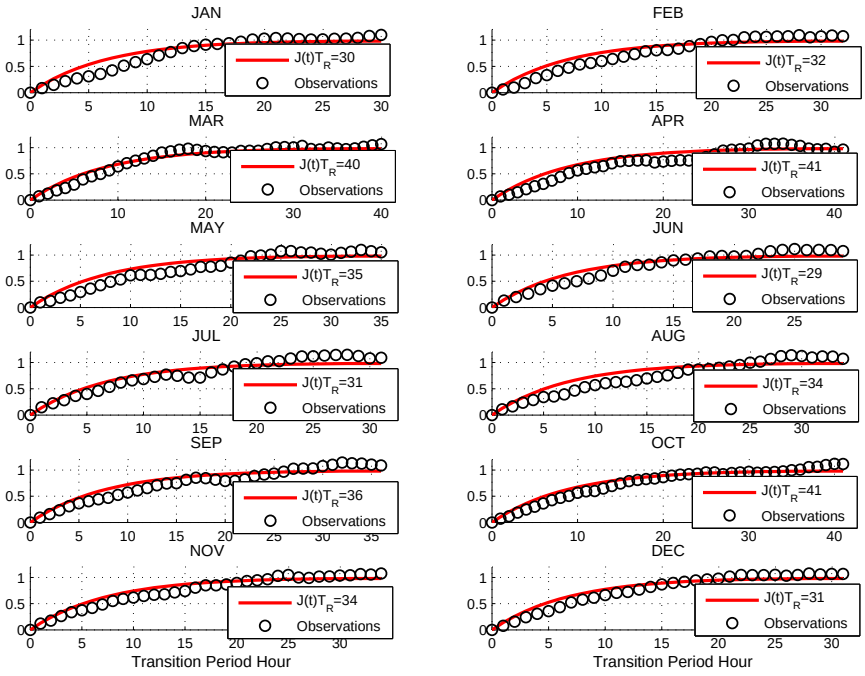


Figure 33: Analytical (continuous line) and observed (circles) transition functions $J(t)$ corresponding to the estimated transition period lengths.

months. Each vector corresponds to one of the values of $N(t-1)$: 0.0, 0.1, ..., 1.0. For each $N(t-1)$, the corresponding distribution of deviates is approximated by the *Beta* distribution with parameters a and b estimated from these deviates. The mean and standard deviation of the PDFs are essentially constant throughout the entire range of $N(t-1)$ values, the skewness of the deviates varies significantly, changing its sign from positive to negative. As can also be seen in Figure 34, the probability density functions of the *Beta* distribution, corresponding to the 11 $N(t-1)$ have significant different shapes. Moreover, since the variability is quite substantial for most months (for all stations), the values of a and b are estimated on a monthly basis.

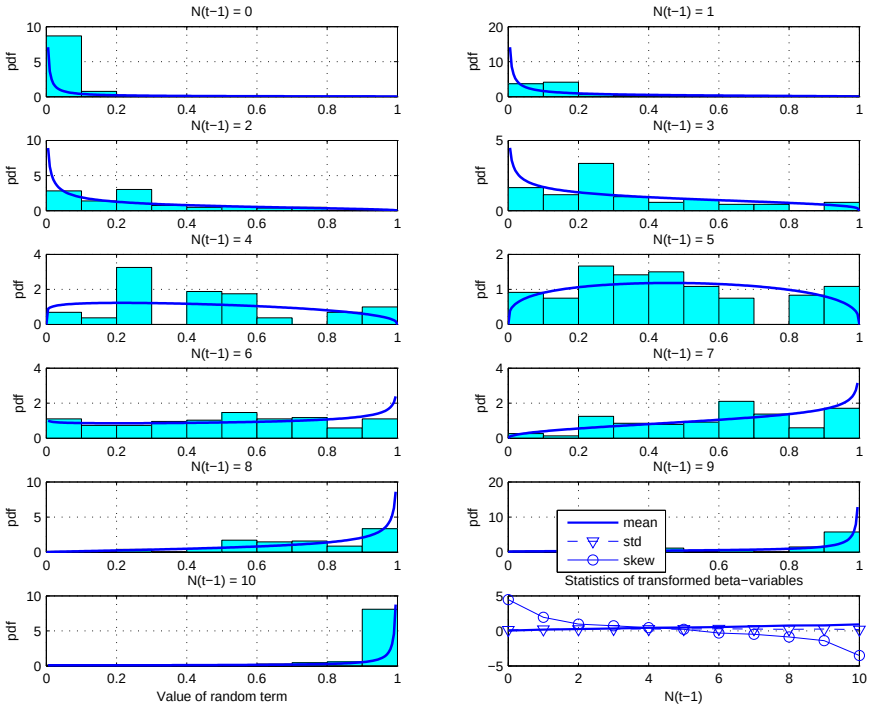


Figure 34: An histogram of deviates $\varepsilon(t)$ in the cloud cover model and the corresponding probability density function (solid line) approximated with the *Beta* distribution. The cloud cover $N(t-1)$ for the month of *November* is given on a [0, 10] basis.

C Definition of sun variables

Equations to define solar variables are taken from the auxiliary material of *Ivanov et al.* (2007), with some adjustment concerning the limits of integration of solar hour angle, altitude and azimuth.

Several variables are introduced that define the Sun's position with respect to a location on Earth. The declination of the Sun, δ $[rad]$, i.e., the angular distance between the celestial equator plane and the Sun, measured from the former (and positive when the Sun lies north of the equator) and along the hour circle (*Eagleson*, 2002) is defined as:

$$\delta = \frac{23.45\pi}{180} \cos\left[\frac{2\pi}{365}(172 - J_{Day})\right] . \tag{51}$$

The angular distance between the planes of the meridian and the Sun's hour circle (*Eagleson*, 2002) is known as the hour angle of the Sun, $\tau_S(t)$ $[rad]$:

$$\tau_S(t) = \frac{15\pi}{180}(t + 12 - \Delta t_{SL}) , \qquad \text{if} \qquad t < 12 + \Delta t_{SL} , \tag{52}$$

$$\tau_S(t) = \frac{15\pi}{180}(t - 12 - \Delta t_{SL}) , \qquad \text{if} \qquad t > 12 + \Delta t_{SL} , \tag{53}$$

where t $[h]$ is the standard time in the time zone of the observer counted from midnight and Δt_{SL} $[h]$ is the time difference between the standard and local meridian:

$$\Delta t_{SL} = \frac{\xi}{15} \left[\, 15|\Delta GMT| - |\Phi'| \, \right] , \tag{54}$$

where ΔGMT $[h]$ is the time difference between the local time zone and Greenwich Mean Time, Φ' $[angular\ degree]$ is the local longitude, and ξ is equal to -1 for west longitude and +1 for east longitude. The solar altitude, i.e., an angle of radiation with respect to an observer's horizon plane, h_S $[rad]$, is defined as

$$\sin h_S = \sin\Phi \sin\delta + \cos\Phi \cos\delta \cos\tau_S , \tag{55}$$

where Φ $[rad]$ is the local latitude. The mean value of solar altitude $h_{S,\,\Delta t}$ over a time interval Δt $[h]$ is often needed in practical applications. It is obtained integrating equation (55) in the interval $\Delta t = [t - t_{bef}] , [t + t_{aft}]$:

$$h_{S,\,\Delta t} = \int_{t-t_{bef}}^{t+t_{aft}} \arcsin[\sin h_S] \, dt , \tag{56}$$

where t_{bef} $[h]$ and t_{aft} $[h]$ are the backward and forward difference between the standard time in the time zone t $[h]$ and the limits of integration of the sun variables. Note the implicit dependence of h_S from the standard time within τ_S.

The Sun's azimuth ζ_S $[rad]$ is obtained from the "hour angle method" as the clockwise angle from north:

$$\zeta_S = \arctan\left(\frac{-\sin\tau_S}{\tan\delta\cos\Phi - \sin\Phi\cos\tau_S}\right) . \tag{57}$$

Note that also ζ_S should be integrated in the interval $\Delta t = [t - t_{bef}]\,, [t + t_{aft}]$ to obtain the average value.

The sunrise time, $T_{H\,rise}$ [*local hour*], the sunset time, $T_{H\,set}$ [*local hour*], and the total day length D_{LH} [h] are also required in applications:

$$T_{H\,rise} = \frac{180}{15\pi}[2\pi - \arccos(-\tan\delta\tan\Phi)] - 12\,, \tag{58}$$

$$T_{H\,set} = \frac{180}{15\pi}\arccos(-\tan\delta\tan\Phi) + 12\,, \tag{59}$$

$$D_{LH} = \frac{360}{15\pi}\arccos(-\tan\delta\tan\Phi)\,. \tag{60}$$

D Solution of the ODE for deterministic air temperature

Equation (17) in Section 5.1 is a first order differential equation, the solution to which can be found if the initial condition, i.e., the initial temperature, $\widetilde{T}(t^*)$, is given. *Curtis and Eagleson* (1982) provide the following equation:

$$\widetilde{T}(t) = \widetilde{T}(t^*)e^{-b_1(t-t^*)} + e^{-b_1 t}G(t,t^*)\,, \tag{61}$$

where:

$$\begin{aligned} G(t,t^*) &= b_0\int_{t^*}^{t} e^{b_1\tau}d\tau + b_2\int_{t^*}^{t} e^{b_1\tau}K(\tau)s(\tau)d\tau + \\ &\quad b_3\int_{t^*}^{t} e^{b_1\tau}K(\tau)r(\tau)d\tau + b_4 q(t-1)\int_{t^*}^{t} e^{b_1\tau}d\tau \\ &= I_1(t) + I_2(t) + I_3(t) + I_4(t)\,. \end{aligned} \tag{62}$$

By using the *full*, non-zero expressions for $s(t)$ and $r(t)$ (the system of equations 19) *Curtis and Eagleson* (1982) derived the following expressions for the terms of $G(t,t^*)$:

$$I_1(t) = b_0\int_{t^*}^{t} e^{b_1\tau}d\tau = \frac{b_0}{b_1}\left[e^{b_1 t} - e^{b_1 t^*}\right]\,, \tag{63}$$

$$\begin{aligned} I_2(t) &= b_2\int_{t^*}^{t} e^{b_1\tau}K(\tau)s(\tau)d\tau \\ &= K(t)\,[\,K_2\,(e^{b_1 t} - e^{b_1(t-1)}) - K_3\,e^{b_1 t}\cos\frac{\pi t}{12} - K_4\,e^{b_1 t}\sin\frac{\pi t}{12} + \\ &\quad K_3\,e^{b_1(t-1)}\cos\frac{\pi(t-1)}{12} + K_4\,e^{b_1(t-1)}\sin\frac{\pi(t-1)}{12}\,] + I_2(t-1)\,, \end{aligned} \tag{64}$$

$$
\begin{aligned}
I_3(t) &= b_3 \int_{t^*}^{t} e^{b_1\tau} K(\tau) r(\tau) d\tau \\
&= K(t)\,[\, K_6\, e^{b_1 t} \sin\frac{\pi t}{12} - K_5\, e^{b_1 t} \cos\frac{\pi t}{12} - \\
&\quad K_6\, e^{b_1(t-1)} \sin\frac{\pi(t-1)}{12} + K_5\, e^{b_1(t-1)} \cos\frac{\pi(t-1)}{12}\,] + I_3(t-1)\,, \quad (65) \\
I_4(t) &= b_4 \int_{t^*}^{t} e^{b_1\tau} q(\tau) d\tau \;=\; \frac{b_4}{b_1} q(t-1)(1-e^{b_1}) e^{b_1 t} + I_4(t-1)\,,
\end{aligned}
\quad (66)
$$

where

$$
\begin{aligned}
p &= \frac{\pi}{12}\,, \qquad K_1 = \frac{b_0}{b_1}\,, \qquad K_2 = \frac{b_2}{b_1}\sin\delta\sin\Phi\,, \\
K_3 &= \frac{b_1 b_2}{b_1^2+p^2}\cos\delta\cos\Phi\,, \qquad K_4 = \frac{p b_2}{b_1^2+p^2}\cos\delta\cos\Phi\,, \\
K_5 &= \frac{p^2 b_3}{b_1^2+p^2}\cos\delta\cos\Phi\,, \qquad K_6 = \frac{p b_1 b_3}{b_1^2+p^2}\cos\delta\cos\Phi\,.
\end{aligned}
\quad (67)
$$

Equation (66) linearizes the integral $I_4(t)$ that contains $q(t)$, which is a non-linear function of the temperature, by using the value from the previous hour $q(t-1)$. Besides, the one-hour integration interval is considered short enough to allow variables $K(t)$ and $q(t-1)$ to be brought outside their respective integrals (equations 64-66).

The full, non-zero expressions for $s(t)$ and $r(t)$ (the system of equations 19) were used to obtain the above general equations (64) - (65). Since $s(t)$ and $r(t)$ can be equal to zero during certain periods of the day, it can be seen that the integrals $I_2(t)$ and $I_3(t)$ may have different forms depending on time of the day. The ranges over which each form is valid are delimited by several critical times. *Curtis and Eagleson* (1982) identify five critical times: 1) t_0 is the value of t in local time corresponding to midnight in standard time; 2) t_R is the earliest standard hour that does not precede local sunrise $T_{H\,rise}$, ($t_R \geq T_{H\,rise}$); 3) t_{12} is the value of t at the earliest standard hour that does not precede local noon ($t_{12} \geq 12$); 4) t_S is the value of t at the earliest standard hour that does not precede local sunset, $T_{H\,set}$ ($t_S \geq T_{H\,set}$); 5) t_{23} is the value corresponding to 23.00 local standard time. The integrals $I_2(t)$ and $I_3(t)$ are evaluated according to the above time ranges using the system of equations (19), which leads to different forms for $G(t, t*)$.

E Air temperature parameter estimation

The parameters of the air temperature component are: the regression coefficients b_i ($i = 0, 1, \ldots, 4$), $\overline{dT}_h$, $\sigma_{dT,h}$, and ρ_{dT}. The procedure of parameter estimation follows *Curtis and Eagleson* (1982). The same is described also in the auxiliary material of *Ivanov et al.* (2007).

According to *Curtis and Eagleson* (1982), equation (61) can be re-written to ob-

tain:

$$\widetilde{T}(t) \;=\; e^{-b_1}\widetilde{T}(t-1) + e^{-b_1 t}G(t,t-1)\,. \tag{68}$$

The hourly temperature change, $Y(t) = T(t) - T(t-1)$, is obtained if temperature $T(t-1)$ is subtracted from both sides of equation (68). *Curtis and Eagleson* (1982) show that an equation for $Y(t)$ can be represented in the regression form:

$$Y(t) \;=\; a_0 + a_1 X_1(t) + \ldots + a_4 X_4(t)\,, \tag{69}$$

where the coefficients a_i-s ($i = 0,\, 1,\, \ldots,\, 4$) are:

$$\begin{aligned} a_1 &= e^{-b_1} - 1\,, \\ a_i &= -\frac{a_1}{b_1} b_i\,, \qquad i \;=\; 0, 2, \ldots, 4\,, \end{aligned} \tag{70}$$

and the predictors $X_i(t)$ are:

$$\begin{aligned} X_1(t) &= \widetilde{T}(t-1)\,, \\ X_2(t) &= \int_{t-1}^{t} K(\tau)s(\tau)d\tau\,, \\ X_3(t) &= \int_{t-1}^{t} K(\tau)r(\tau)d\tau\,, \\ X_4(t) &= q(t-1)\,. \end{aligned} \tag{71}$$

As above, the one-hour integration interval is considered to be short enough to allow variable $q(t-1)$ to be brought outside its integral. Similarly to the previous discussion, the terms $X_2(t)$ and $X_3(t)$ containing $s(t)$ and $r(t)$ have different form depending on time of the day. From a set of linear equations (69), the regression coefficients a_i-s ($i = 0,\, 1,\, \ldots,\, 4$) can be found with conventional methods. Once a_i-s ($i = 0,\, 1,\, \ldots,\, 4$) have been estimated, the regression parameters, b_i, can be easily obtained from (70). The b_i-s are estimated on a monthly basis.

Once the regression parameters have been estimated, equation (17) can be used to simulate the deterministic component of the hourly temperature model. Equation (17) is applied each day to compute temperatures for each hour starting from midnight ($t = 0$). The initial temperature, $\widetilde{T}(t^*)$, is taken as the deterministic temperature component estimated at 23 h of the previous day. According to (16), the difference between the observed and estimated deterministic temperature components defines the temperature random deviates. Consequently, series of deviates can be estimated for each period of interest, e.g., for each month, season, and also hour of the day. The parameters $\overline{dT}_h$, $\sigma_{dT,h}$, and ρ_{dT}, are obtained using conventional estimation techniques.

F Clear sky radiation parameterizations

The transmittances in band Λ1 and band Λ2 required to estimate direct beam radiation at normal incidence, $R_{Bn,\Lambda}$, and the incedent diffuse irradiance, $R_{Dp,\Lambda}$, are calculated as in *Gueymard* (2008).

F.1 Direct beam irradiance

The ozone absorption transmittances, $T_{o,\Lambda}$, are:

$$\begin{aligned} T_{o,\Lambda 1} &= (1 + f_1\, m_O + f_2\, {m_O}^2)/(1 + f_3\, m_O)\,, \\ T_{o,\Lambda 2} &= 1.0\,, \end{aligned} \tag{72}$$

where m_O is the ozone absorption air mass and the other parameter are function of the ozone amount in atmospheric column, u_o $[cm]$:

$$\begin{aligned} f_1 &= u_o(10.979 - 8.5421 u_o)/(1 + 2.0115 u_o + 40.189 u_o^2)\,, \\ f_2 &= u_o(-0.027589 - 0.005138 u_o)/(1 - 2.4857 u_o + 13.942 u_o^2)\,, \\ f_3 &= u_o(10.995 - 5.5001 u_o)/(1 + 1.6784 u_o + 42.406 u_o^2)\,. \end{aligned} \tag{73}$$

The nitrogen dioxide absorption transmittances, $T_{n,\Lambda}$, are:

$$\begin{aligned} T_{n,\Lambda 1} &= \min\left[1, (1 + g_1\, m_W + g_2\, m_W^2)/(1 + g_3\, m_W)\right], \\ T_{n,\Lambda 2} &= 1.0\,, \end{aligned} \tag{74}$$

where m_W is water vapor air mass and the other parameters are function of the nitrogen dioxide amount in atmospheric column, u_n $[cm]$:

$$\begin{aligned} g_1 &= (0.17499 + 41.654 u_n - 2146.4 u_n^2)/(1 + 22295.0 u_n^2)\,, \\ g_2 &= u_n(-1.2134 + 59.324 u_n)/(1 + 8847.8 u_n^2)\,, \\ g_3 &= (0.17499 + 61.658 u_n + 9196.4 u_n^2)/(1 + 74109.0 u_n^2)\,. \end{aligned} \tag{75}$$

The Rayleigh scattering transmittances, $T_{R,\Lambda}$, are:

$$\begin{aligned} T_{R,\Lambda 1} &= 1 + 1.8169\, m'_R - 0.033454\, m'^2_R)/(1 + 2.063\, m'_R + 0.31978\, m'^2_R)\,, \\ T_{R,\Lambda 2} &= (1 - 0.010394\, m'_R)/(1 - 0.00011042\, m'^2_R)\,, \end{aligned} \tag{76}$$

where $m'_R = (P_{atm}/P_{atm,0}) m_R$ is calculated from the Rayleigh scattering and uniformly mixed gas air mass, m_R, after correcting atmospheric pressure for the difference in pressures between the reference point, P_{atm} $[mbar]$, and sea level, $P_{atm,0} = 1013.25$ $[mbar]$. The equation to scale atmospheric pressure with elevation is $P_{atm}/P_{atm,0} = \exp\left[-g Z_{ref}/(R_d\, T_m)\right]$, with $g = 9.81$ $[m\ s^{-2}]$ acceleration of gravity, $R_d = 287.05$ $[J\ kg^{-1}; K^{-1}]$, air gas constant, T_m average value of air temperature between sea level and Z_{ref}, where Z_{ref} $[m]$ is the elevation of the reference point. Assuming on average $T_m = 288.15$ $[K]$ we have $P_{atm}/P_{atm,0} = \exp\left[-g Z_{ref}/8434.5\right]$.

The uniformly mixed gas absorption transmittances, $T_{g,\Lambda}$, are:

$$\begin{aligned} T_{g,\Lambda 1} &= (1 + 0.95885\, m'_R - 0.012871\, m'^2_R)/(1 + 0.96321\, m'_R + 0.015455\, m'^2_R)\,, \\ T_{g,\Lambda 2} &= (1 + 0.27284\, m'_R - 0.00063699\, m'^2_R)/(1 + 0.30306\, m'_R)\,. \end{aligned} \tag{77}$$

The water vapor absorption transmittances, $T_{w,\Lambda}$, are:

$$\begin{aligned} T_{w,\Lambda 1} &= (1 + h_1\, m_W)/(1 + h_2\, m_W)\,, \\ T_{w,\Lambda 2} &= (1 + c_1\, m_W + c_2\, m_W^2)/(1 + c_3\, m_W + c_4\, m_W^2)\,, \end{aligned} \tag{78}$$

where m_W is again the water vapor air mass and the other parameters are function of precipitable water in atmospheric column, w $[cm]$:

$$\begin{aligned} c_1 &= w(19.566 - 1.6506w + 1.0672w^2)/(1 + 5.4248w + 1.6005w^2)\,, \\ c_2 &= w(0.50158 - 0.14732w + 0.047584w^2)/(1 + 1.1811w + 1.0699w^2)\,, \\ c_3 &= w(21.286 - 0.39232w + 1.2692w^2)/(1 + 4.8318w + 1.412w^2)\,, \\ c_4 &= w(0.70992 - 0.23155w + 0.096514w^2)/(1 + 0.44907w + 0.75425w^2)\,, \\ h_1 &= w(0.065445 + 0.00029901w)/(1 + 1.2728w)\,, \\ h_2 &= w(0.065687 + 0.0013218w)/(1 + 1.2008w)\,. \end{aligned} \tag{79}$$

Since the precipitable water in atmospheric column, w, is a variable not routinely measured, it is estimated from the dew point temperature, T_{dew} $[^\circ C]$, according to an empirical model of *Iqbal* (1983): $w = \exp(0.07\, T_{dew} - 0.075)$ $[cm]$.

Aerosol extinction transmittances, $T_{a,\Lambda}$, are modeled as in *Gueymard* (2008). The band-average spectral aerosol optical depth, $\tau_{a\Lambda}$, is expressed with the same formalism of the original Ångström law, linearized for discrete aerosol channels (see also Section 6.1), but considering an effective wavelength for the entire bands $\Lambda 1_e$ and $\Lambda 2_e$:

$$\begin{aligned} \tau_{a\Lambda 1} &= \beta_{\Lambda 1} \Lambda 1_e^{-\alpha_{\Lambda 1}}\,, \\ \tau_{a\Lambda 2} &= \beta_{\Lambda 2} \Lambda 2_e^{-\alpha_{\Lambda 2}}\,, \end{aligned} \tag{80}$$

where $\alpha_{\Lambda 1}$, $\alpha_{\Lambda 2}$, $\beta_{\Lambda 1}$, and $\beta_{\Lambda 2}$, are the Ångström turbidity parameters for the two bands $\Lambda 1$ and $\Lambda 2$:

$$\begin{aligned} \beta_{\Lambda 1} &= \beta_A 0.7^{\alpha_{\Lambda 1} - \alpha_{\Lambda 2}}\,, \\ \beta_{\Lambda 2} &= \beta_A\,. \end{aligned} \tag{81}$$

As in *Gueymard* (2008) no distinction is made between the two α_Λ, that are taken equal to the reference Ångström turbidity α_A: $\alpha_{\Lambda 1} = \alpha_{\Lambda 2} = \alpha_A$, consequently also $\beta_{\Lambda 1} = \beta_{\Lambda 2} = \beta_A$. The effective wavelength for the entire bands, $\Lambda 1_e$, and $\Lambda 2_e$, are essentially function of a parameter $u_A = \ln[1 + m_A\, \beta_\Lambda]$ (*Gueymard*, 1989), where m_A is the air mass for aerosol extinction. The aerosol extinction transmittances $T_{a,\Lambda}$

for each band are thus:

$$\begin{aligned} T_{a,\Lambda 1} &= e^{-m_A \tau_{a,\Lambda 1}}\,, \\ T_{a,\Lambda 2} &= e^{-m_A \tau_{a,\Lambda 2}}\,. \end{aligned} \tag{82}$$

The revised functions used here to obtain u_A are as in *Gueymard* (2008):

$$\begin{aligned} \Lambda 1_e &= (d_0 + d_1\, u_A + d_2\, u_A^2)/(1 + d_3\, u_A^2)\,, \\ \Lambda 2_e &= (e_0 + e_1\, u_A + e_2\, u_A^2)/(1 + e_3\, u_A)\,, \end{aligned} \tag{83}$$

where:

$$\begin{aligned} d_0 &= 0.57664 - 0.024743\alpha_{\Lambda 1}\,, \\ d_1 &= (0.093942 - 0.2269\alpha_{\Lambda 1} + 0.12848\alpha_{\Lambda 1}^2)/(1 + 0.6418\alpha_{\Lambda 1})\,, \\ d_2 &= (-0.093819 + 0.36668\alpha_{\Lambda 1} - 0.12775\alpha_{\Lambda 1}^2)/(1 - 0.11651\alpha_{\Lambda 1})\,, \\ d_3 &= \alpha_{\Lambda 1}(0.15232 - 0.087214\alpha_{\Lambda 1} + 0.012664\alpha_{\Lambda 1}^2)/(1 - 0.90454\alpha_{\Lambda 1} + 0.26167\alpha_{\Lambda 1}^2)\,, \\ e_0 &= (1.183 - 0.022989\alpha_{\Lambda 2} + 0.020829\alpha_{\Lambda 2}^2)/(1 + 0.11133\alpha_{\Lambda 2})\,, \\ e_1 &= (-0.50003 - 0.18329\alpha_{\Lambda 2} + 0.23835\alpha_{\Lambda 2}^2)/(1 + 1.6756\alpha_{\Lambda 2})\,, \\ e_2 &= (-0.50001 + 1.1414\alpha_{\Lambda 2} + 0.0083589\alpha_{\Lambda 2}^2)/(1 + 11.168\alpha_{\Lambda 2})\,, \\ e_3 &= (-0.70003 - 0.73587\alpha_{\Lambda 2} + 0.51509\alpha_{\Lambda 2}^2)/(1 + 4.7665\alpha_{\Lambda 2})\,. \end{aligned} \tag{84}$$

In the above equations individual optical masses, m_R, m_O, m_W, and m_A, are used for Rayleigh (molecular) scattering and uniformly mixed gases absorption, ozone absorption, water vapor absorption, and aerosol extinction, respectively (*Gueymard*, 2008). Individual optical masses rather than a single air mass are considered to better characterize the solar rays' pathlength through the atmosphere. The values of the optical masses are obtained from the sun's solar altitude, h'_S [*angular degree*], with the same functions of the *REST* model (*Gueymard*, 2003). Note that molecular optical mass, m_R, sometimes is called "relative air mass", or simply "air mass":

$$\begin{aligned} m_R &= \left[\sin(h'_S) + (0.48353\; Z^{0.09584})/(96.741 - Z^{1.1754})\right]^{-1}\,, \\ m_O &= \left[\sin(h'_S) + (1.0651\; Z^{0.6379})/((101.8 - Z)^{2.2694})\right]^{-1}\,, \\ m_W &= \left[\sin(h'_S) + (0.10648\; Z^{0.11423})/((93.781 - Z)^{1.9203})\right]^{-1}\,, \\ m_A &= \left[\sin(h'_S) + (0.16851\; Z^{0.18198})/((95.318 - Z)^{1.9542})\right]^{-1}\,, \end{aligned} \tag{85}$$

where $Z = 90 - h'_S$ [*angular degree*] is the sun's zenith angle.

F.2 Diffuse irradiance

Aerosol extinction is mostly caused by scattering, and by absorption for the remaining part. The aerosol scattering transmittances are:

$$\begin{aligned} T_{as,\Lambda 1} &= e^{-m_A \omega_{\Lambda 1} \tau_{a\Lambda 1}}\,, \\ T_{as,\Lambda 2} &= e^{-m_A \omega_{\Lambda 2} \tau_{a\Lambda 2}}\,. \end{aligned} \tag{86}$$

where $\omega_{\Lambda 1}$ and $\omega_{\Lambda 2}$ are the single scattering albedos. The forward scattering fractions for Rayleigh extinction are indicated with $B_{R,\Lambda}$. In the absence of multiple scattering, they would be exactly 0.5 because molecules scatter equally in the forward and backward directions. Multiple scattering is negligible in $\Lambda 2$ (so that $B_{R,\Lambda 2} = 0.5$), but not in $\Lambda 1$. Using a simple spectral model to describe this effect $B_{R,\Lambda 1}$ is obtained after spectral integration and parametrization as in *Gueymard* (2008):

$$B_{R,\Lambda 1} = 0.5\,(0.89013 - 0.0049558\,m_R + 0.000045721\,m_R^2)\,. \tag{87}$$

The aerosol forward scatterance factor, B_a, is the same as *Gueymard* (2008):

$$B_a = 1 - \exp\left[-0.6931 - 1.8326\sin(h_S)\right]. \tag{88}$$

The correction factors, F_Λ, to compensate for multiple scattering effects and shortcomings for the simple approach are (*Gueymard*, 2008):

$$\begin{aligned} F_{\Lambda 1} &= (g_0 + g_1\tau_{a\Lambda 1})/(1 + g_2\tau_{a\Lambda 1})\,, \\ F_{\Lambda 2} &= (h_0 + h_1\tau_{a\Lambda 2})/(1 + h_2\tau_{a\Lambda 2})\,, \\ g_0 &= (3.715 + 0.368\,m_A + 0.036294\,m_A^2)/(1 + 0.0009391\,m_A^2)\,, \\ g_1 &= (-0.164 - 0.72567\,m_A + 0.20701\,m_A^2)/(1 + 0.0019012\,m_A^2)\,, \\ g_2 &= (-0.052288 + 0.31902\,m_A + 0.17871\,m_A^2)/(1 + 0.0069592\,m_A^2)\,, \\ h_0 &= (3.4352 + 0.65267\,m_A + 0.00034328\,m_A^2)/(1 + 0.034388\,m_A^{1.5})\,, \\ h_1 &= (1.231 - 1.63853\,m_A + 0.20667\,m_A^2)/(1 + 0.1451\,m_A^{1.5})\,, \\ h_2 &= (0.8889 - 0.55063\,m_A + 0.50152\,m_A^2)/(1 + 0.14865\,m_A^{1.5})\,. \end{aligned} \tag{89}$$

The sky albedo, $\rho_{s,\Lambda}$, parameterizations are again from *Gueymard* (2008):

$$\begin{aligned} \rho_{s,\Lambda 1} &= \frac{0.13363 + 0.00077358\alpha_{\Lambda 1} + \beta_{\Lambda 1}(0.37567 + 0.22946\alpha_{\Lambda 1})/(1 - 0.10832\alpha_{\Lambda 1})}{1 + \beta_{\Lambda 1}(0.84057 + 0.68683\alpha_{\Lambda 1})/(1 - 0.08158\alpha_{\Lambda 1})}\,, \\ \rho_{s,\Lambda 2} &= \frac{0.010191 + 0.00085547\alpha_{\Lambda 2} + \beta_{\Lambda 2}(0.14618 + 0.062758\alpha_{\Lambda 2})/(1 - 0.19402\alpha_{\Lambda 2})}{1 + \beta_{\Lambda 2}(0.58101 + 0.17426\alpha_{\Lambda 2})/(1 - 0.17586\alpha_{\Lambda 2})}\,. \end{aligned} \tag{90}$$

The reduction factor for direct beam, M_B, and global, M_G, radiation between the first band radiation and PAR adopted in equation (33) and (34) are (*Gueymard*,

2008):

$$\begin{aligned} M_B &= (t_0 + t_1\beta_e + t_2\beta_e^2)/(1 + t_3\beta_e^2)\,, \\ M_G &= (v_0 + v_1\beta_e + v_2\beta_e^2)/(1 + v_3\beta_e^2)\,, \end{aligned} \tag{91}$$

where the effective turbidity coefficient, β_e, is obtained from the previously defined $\alpha_{\Lambda 1}$, $\beta_{\Lambda 1}$, and $\Lambda 1_e$ as: $\beta_e = \beta_{\Lambda 1}(\Lambda 1_e^{1.3-\alpha_{\Lambda 1}})$ and the other parameters are function of $m_{15} = \min(m_R, 15)$:

$$\begin{aligned} t_0 &= \frac{0.90227 + 0.29\, m_{15} + 0.22928\, m_{15}^2 - 0.0046842\, m_{15}^3}{1 + 0.35474\, m_{15} + 0.19721\, m_{15}^2}\,, \\ t_1 &= \frac{-0.10591 + 0.15416\, m_{15} - 0.048486\, m_{15}^2 + 0.0045932\, m_{15}^3}{1 - 0.29044\, m_{15} + 0.026267\, m_{15}^2}\,, \\ t_2 &= \frac{0.47291 - 0.44639\, m_{15} + 0.1414\, m_{15}^2 - 0.014978\, m_{15}^3}{1 - 0.37798\, m_{15} + 0.052154\, m15^2}\,, \\ t_3 &= \frac{0.077407 + 0.18897\, m_{15} - 0.072869\, m_{15}^2 + 0.0068684\, m_{15}^3}{1 - 0.25237\, m15 + 0.020566\, m_{15}^2}\,, \end{aligned}$$

$$\begin{aligned} v_0 &= \frac{0.82725 + 0.86015\, m_{15} + 0.00713\, m_{15}^2 + 0.00020289\, m_{15}^3}{1 + 0.90358\, m_{15} + 0.015481\, m_{15}^2}\,, \\ v_1 &= \frac{-0.089088 + 0.089226\, m_{15} - 0.021442\, m_{15}^2 + 0.0017054\, m_{15}^3}{1 - 0.28573\, m_{15} + 0.024153\, m_{15}^2}\,, \\ v_2 &= \frac{-0.05342 - 0.0034387\, m_{15} + 0.0050661\, m_{15}^2 - 0.00062569\, m_{15}^3}{1 - 0.32663\, m_{15} + 0.029382\, m_{15}^2}\,, \\ v_3 &= \frac{-0.17797 + 0.13134\, m_{15} - 0.030129\, m_{15}^2 + 0.0023343\, m_{15}^3}{1 - 0.28211\, m_{15} + 0.023712\, m_{15}^2}\,. \end{aligned} \tag{92}$$

Note that this parametrization to estimate PAR was originally developed only for clear sky condition. In the weather generator is applied indifferently for clear and cloudy sky conditions.

G Overcast sky radiation parameterizations

According to *Stephens* (1978), the cloud optical thickness, τ_N, is one of the most important parameters needed to describe the radiative properties of clouds. Approximate range for τ_N is $5 < \tau_N < 500$. By considering a set of "standard" cloud types, *Stephens* (1978) derives that τ_N can be approximately parameterized in terms of the effective radius of cloud-droplet size distribution, r_e $[\mu m]$, and liquid water path, LWP $[g\ m^{-2}]$:

$$\tau_N \approx \frac{1.5\ LWP}{r_e}\,. \tag{93}$$

Liquid water path can be formally defined as the integral of the liquid water content from the cloud base to the cloud top. By considering two spectral intervals $[0.29 \div 0.75\,\mu m]$ and $[0.75 \div 4.0\,\mu m]$ for the set of "standard" cloud types, *Stephens*

(1978) also derives the following relationships:

$$\log_{10}(\tau_{N1}) = 0.2633 + 1.7095 \ln\big[\log_{10}(LWP)\big] , \quad (94)$$

$$\log_{10}(\tau_{N2}) = 0.3492 + 1.6518 \ln\big[\log_{10}(LWP)\big] , \quad (95)$$

where expression (94) refers to the first considered spectral band, where absorption by cloud droplets is extremely small, and expression (95) refers to the second band, where absorption is significant. It follows from equations (93), (94), and (95) that the knowledge of LWP allows one to obtain an approximate estimate of r_e. *Slingo* (1989) introduced a parametrization that provided an accurate estimate of cloud radiative properties based on r_e. This parametrization is used in the following.

Slingo (1989) considered four spectral bands, one in *UV/VIS*, $[0.25 \div 0.69\,\mu m]$, and three in *NIR* wavelength intervals: $[0.69 \div 1.19\,\mu m]$, $[1.19 \div 2.38\,\mu m]$, $[2.38 \div 4.0\,\mu m]$ with the following respective fractions, λ_i, $i = 1, \ldots, 4$, of solar irradiance at the top of the atmosphere: 0.460, 0.326, 0.181, 0.033. Note the slight differences with *Ivanov et al.* (2007). Following the parametrization of *Slingo* (1989), cloud transmittances and reflectances are estimated separately for each of these spectral intervals. The radiative fluxes computed for these four bands are then scaled to the two principal bands $\Lambda 1$ $[0.29 \div 0.7\,\mu m]$ and $\Lambda 2$ $[0.70 \div 4.0\,\mu m]$, considered in the model (Section 6.2).

G.1 Direct beam irradiance

For a given spectral interval, the single scattering properties of typical water clouds can be parameterized in terms of the liquid water path (provided r_e is known):

$$\tau_\lambda = LWP(a_\lambda + \frac{b_\lambda}{r_e}) , \quad (96)$$

$$\tilde{\omega_\lambda} = 1 - (c_\lambda + d_\lambda r_e) , \quad (97)$$

$$g_\lambda = e_\lambda + f_\lambda r_e , \quad (98)$$

where τ_λ is the cloud optical depth, $\tilde{\omega_\lambda}$ is the single scatter albedo, g_λ is the asymmetry parameter, and a_λ, b_λ, c_λ, d_λ, e_λ, f_λ are the coefficients of the parametrization (provided in Table 4).

Thus the transmissivity for the direct beam radiation, $T_{B,\lambda}$, is:

$$T_{B,\lambda} = e^{\left[-(1-\tilde{\omega_\lambda}\Upsilon_\lambda)\frac{\tau_\lambda}{\sin h_S}\right]} . \quad (99)$$

where h_S $[rad]$ is the solar height and $\Upsilon_\lambda = g_\lambda^2$.

G.2 Diffuse irradiance

Using the same notation as in G.1 and omitting the subscripts λ that denote a particular spectral band it is possible to introduce:

$$\beta_0 = \frac{3}{7}(1-g)\,, \tag{100}$$

$$\beta(h_S) = 0.5 - \frac{3\sin h_S\, g}{4(1+g)}\,, \tag{101}$$

$$\Upsilon = g^2\,, \tag{102}$$

$$U_1 = \frac{7}{4}\,, \tag{103}$$

$$U_2 = \frac{7}{4}\left[1 - \frac{(1-\tilde{\omega})}{7\tilde{\omega}\beta_0}\right]\,, \tag{104}$$

$$\alpha_1 = U_1[1-\tilde{\omega}(1-\beta_0)]\,, \tag{105}$$

$$\alpha_2 = U_2\tilde{\omega}\beta_0\,, \tag{106}$$

$$\alpha_3 = (1-\Upsilon)\tilde{\omega}\beta(h_S)\,, \tag{107}$$

$$\alpha_4 = (1-\Upsilon)\tilde{\omega}(1-\beta(h_S))\,, \tag{108}$$

$$\epsilon = \sqrt{\alpha_1^2 - \alpha_2^2}\,, \tag{109}$$

$$M = \frac{\alpha_2}{\alpha_1+\epsilon}\,, \tag{110}$$

$$E = e^{-\epsilon\tau}\,, \tag{111}$$

$$\gamma_1 = \frac{(1-\tilde{\omega}\Upsilon)\alpha_3 - \sin h_S\,(\alpha_1\alpha_3 + \alpha_2\alpha_4)}{(1-\tilde{\omega}\Upsilon)^2 - \epsilon^2\sin^2 h_S}\,, \tag{112}$$

$$\gamma_2 = \frac{-(1-\tilde{\omega}\Upsilon)\alpha_4 - \sin h_S\,(\alpha_1\alpha_4 + \alpha_2\alpha_3)}{(1-\tilde{\omega}\Upsilon)^2 - \epsilon^2\sin^2 h_S}\,, \tag{113}$$

where the U_1 and U_2 are the reciprocals of the effective cosines for the diffuse upward and downward fluxes respectively, β_0 is the fraction of the scattered diffuse radiation, which is scattered into the backward hemisphere, and $\beta(h_S)$ is the same for the direct radiation.

The diffuse transmissivity for direct beam and incident diffuse radiation are $T_{DB,\lambda}$ and $T_{DD,\lambda}$ respectively. The diffuse reflectivity for direct beam and diffuse incident radiation are $A_{B,\lambda}$ and $A_{D,\lambda}$ respectively, as defined in *Slingo* (1989). The diffuse reflectivity for diffuse incident radiation is:

$$A_{D,\lambda} = \frac{M_\lambda(1-E_\lambda^2)}{1-E_\lambda^2 M_\lambda^2}\,. \tag{114}$$

The diffuse transmissivity for diffuse incident radiation is:

$$T_{DD,\lambda} = \frac{E_\lambda(1-M_\lambda^2)}{1-E_\lambda^2 M_\lambda^2}\,. \tag{115}$$

Table 4: The values of coefficients in equations (96) - (98) (from *Slingo* (1989)).

Band	a_λ $[10^{-2}\, m^2\, g^{-1}]$	b_λ $[\mu m\, m^2\, g^{-1}]$	c_λ	d_λ $[\mu m^{-1}]$	e_λ	f_λ $[10^{-3}\, \mu m^{-1}]$
$[0.25 \div 0.69\ \mu m]$	2.817	1.305	-5.62×10^{-8}	1.63×10^{-7}	0.829	2.482
$[0.69 \div 1.19\ \mu m]$	2.682	1.346	-6.94×10^{-6}	2.35×10^{-5}	0.794	4.226
$[1.19 \div 2.38\ \mu m]$	2.264	1.454	4.64×10^{-4}	1.24×10^{-3}	0.754	6.560
$[2.38 \div 4.00\ \mu m]$	1.281	1.641	2.01×10^{-1}	7.56×10^{-3}	0.826	4.353

The diffuse transmissivity for direct beam incident radiation is:

$$T_{DB,\lambda} = -\gamma_{2,\lambda} T_{DD,\lambda} - \gamma_{1,\lambda} T_{B,\lambda} A_{D,\lambda} + \gamma_{2,\lambda} T_{B,\lambda}\,. \tag{116}$$

Finally, the diffuse reflectivity for direct beam radiation is:

$$A_{B,\lambda} = -\gamma_{2,\lambda} A_{D,\lambda} - \gamma_{1,\lambda} T_{B,\lambda} T_{DD,\lambda} + \gamma_{1,\lambda}\,. \tag{117}$$

H Terrain effects

Solar radiation originating from the sun travels through the atmosphere, and is modified by topography and other surface features. Solar radiation at the ground surface can be intercepted as direct beam, $R^T_{B,\Lambda}$, diffuse, $R^T_{D,\Lambda}$, and reflected radiation, $R^T_{R,\Lambda}$. As anticipated in Section 6.2, incoming solar radiation is function of the local topography through site aspect and slope, and of the surrounding terrain through sky view factor, $S_{vf}(\vec{x})$, and shadow effect, $S_h(\vec{x}, t)$, where $\vec{x}$ is the position and t is the local time. A brief description of incoming solar radiation components and topographic effects is provided in Figure 35. In this section all the symbols will refer to clear sky conditions, nevertheless results are valid also in cloudy conditions.

The importance of topographic variability in hydrological and biophysical processes is well known (*Bertoldi et al.*, 2006; *Ivanov et al.*, 2008b). For such a reason the quantities useful to take into account topographic influences on solar radiation are delineated in the following.

The principal variable controlling incident radiation on a slope, in mountainous terrain, is the local solar illumination angle, $\varphi_{S,T}$ $[rad]$, that is defined as the angle between the sun beam and the normal to the slope surface (*Dozier and Frew*, 1990), given by :

$$\cos\varphi_{S,T} = \cos\beta_T \sin h_S + \sin\beta_T cosh_S \cos(\zeta_S - \zeta_T)\,, \tag{118}$$

where β_T $[rad]$ is the slope of the site, ζ_T $[rad]$ is the local aspect (clockwise direction from north), and h_S $[rad]$, ζ_S $[rad]$ are the solar altitude and azimuth angles respectively.

Another important parameter is the sky view factor, S_{vf} for which two definitions

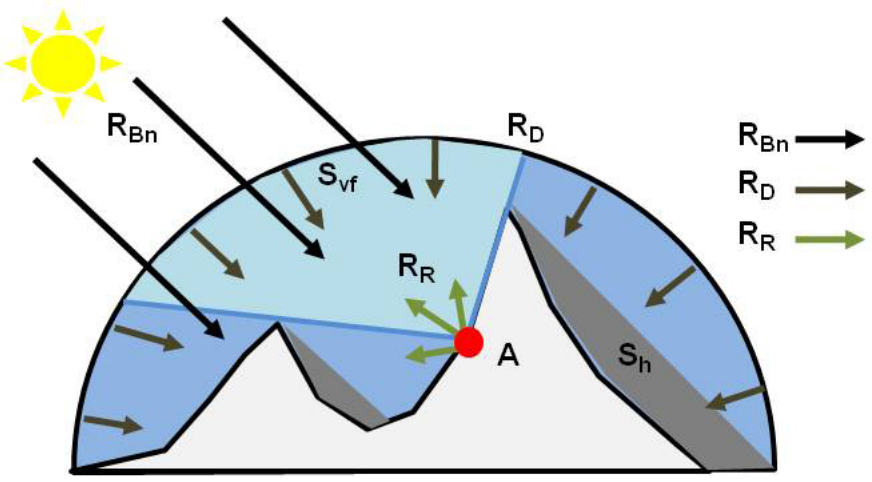


Figure 35: Components of incoming solar radiation on a slope: direct beam radiation at normal incidence, R_{Bn}, diffuse radiation, R_D, and diffuse and direct radiations reflected off by nearby terrain, R_R. The reflected contribution from a generic A location is shown as example. Sky view factor, S_{vf}, from A and shadow effects, S_h, in the represented landscape are also shown. The figure is adapted from *Dubayah and Loechel* (1997).

have been proposed (*Chen et al.*, 2006). The first one assumes a surface with a unique slope receiving diffuse radiation isotropically, and posits that total diffuse radiation should be proportional to the fraction of sky dome viewed by the inclined surface. If β_T is the surface slope angle, then this sky view factor is given by the following equation: $S'_{vf} = (1+\cos\beta_T)/2$ [−] (*Bonan*, 2002). However, the sky dome viewed by the slope surface in mountainous terrain can be obstructed by neighboring surfaces. *Dozier and Frew* (1990) provide a method to take this effect into account, defining the sky-view factor, S_{vf}, as:

$$S_{vf} \approx \frac{1}{2\pi}\int_0^{2\pi}\Big[\cos\beta_T \sin^2 H_\zeta + \sin\beta_T \cos(\zeta-\zeta_T)(H_\zeta - \sin H_\zeta \cos H_\zeta)\Big] d\zeta\,, \tag{119}$$

where H_ζ is the horizon angle (Figure 36), measured from the zenith downward to the local horizon, for direction ζ. Further details on the calculation of (119) are provided in *Dozier and Frew* (1990). Equation (119) includes the possibility to account for a variable horizon angle surrounding the point of interest, and not only for a constant horizon as assumed in the other sky-view factor derivation. Therefore, (119) is used to calculate S_{vf}.

Dozier and Frew (1990) derived also a terrain configuration factor, C_t [−], which approximates the total area between the point and the surrounding terrain for which the points are mutually visible:

$$C_t \approx \frac{1+\cos\beta_T}{2} - S_{vf}\,. \tag{120}$$

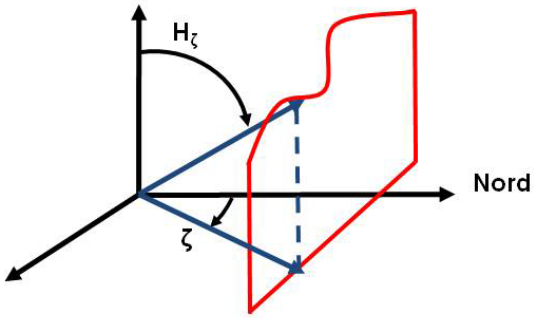


Figure 36: Horizon angle, H_ζ, for a direction ζ, adapted from *Dozier and Frew* (1990).

As counterpart of sky view factor, the terrain configuration factor, C_t, estimates the fraction of the surrounding terrain visible to the point and varies from 0 (only sky visible) to 1 (only terrain visible). Further details on the calculation of (120) are provided from *Dozier and Frew* (1990). The shadow effect, S_h [0/1], is finally calculated as a binary coefficient which value is zero when the sloping surface is shadowed by neighboring terrain, while equal to one otherwise (*Dubayah and Loechel*, 1997; *Chen et al.*, 2006).

The direct beam, $R_{dir,\Lambda} = R^T_{B,\Lambda}$, flux on a general slope is thus given by:

$$R^T_{B,\Lambda} \;=\; S_h \;\cos\varphi_{S,T}\; R_{Bn,\Lambda}\,. \tag{121}$$

Wherever $\cos\varphi_{S,T}$ is negative, the point is "self-shadowed", i.e. the sun is below the local horizon caused by the slope itself. When instead $S_h = 0$ is cast shadowed, i.e. the shadow is caused by nearby terrain blocking the sun (*Dubayah and Loechel*, 1997). Note that when there is no shadow effect and the surface is flat $\beta_T = 0$, equation 121 reduces to $R^T_{B,\Lambda} = \sin h_S \; R_{Bn,\Lambda}$. The latter is what the weather generator calculates by default.

The diffuse sky irradiance, $R^T_{D,\Lambda}$, on a surface oriented in space is composed of three components: the circumsolar, the circumzenith, and isotropic irradiation (*Olseth et al.*, 1995; *Olseth and Skartveit*, 1997), and for each of these components a specific topographic correction should be applied, see for example *Olseth and Skartveit* (1997) or a simplified version in the auxiliary material of *Ivanov et al.* (2007). Frequently, for simplicity the entire incident diffuse radiation $R_{D,\Lambda}$ is considered as isotropic (*Dozier and Frew*, 1990; *Dubayah and Loechel*, 1997; *Chen et al.*, 2006) and is given by:

$$R^T_{D,\Lambda} \;=\; S_{vf}\; R_{D,\Lambda}\,. \tag{122}$$

Another contribution to diffuse irradiance is given by reflected radiation, $R^T_{R,\Lambda}$, on surrounding topography. Incoming energy, in fact, may be reflected from nearby terrain toward the point of interest and can rarely be expected to be isotropic. In order to account for this effect, an approximate terrain configuration factor, C_t, is usually

employed (equation 120) (*Dozier and Frew*, 1990; *Dubayah and Loechel*, 1997). This is motivated by the complexity in determining the geometric relationships between a particular location and all the surrounding terrain elements. Therefore, the reflected radiation, $R^T_{R,\Lambda}$, from surrounding terrain is simply estimated as:

$$C_t \; R^T_{R,\Lambda} = C_t \; \rho_g \big(R_{Bn,\Lambda} \cos(\varphi_{S,T}) + (1 - S_{vf}) R_{D,\Lambda} \big) \,, \tag{123}$$

where ρ_g is the average ground albedo refereing to a large area of 5-50 $[km]$ radius around the point (*Gueymard*, 2008). Note that when an unobscured flat surface is considered $C_t = 0$, since $\beta_T = 0$ and $S_{vf} = 1$, i.e. all the sky dome is visible. Consequently the reflected radiation component is $R^T_{R,\Lambda} = 0$.

The sum of the diffuse shortwave radiation on a slope is the contribute of two components: $R_{dif,\Lambda} = R^T_{D,\Lambda} + C_t \; R^T_{R,\Lambda}$. Finally, the global shortwave radiation, $R_{sw,\Lambda}$, is:

$$R_{sw,\Lambda} = R_{dir,\Lambda} + R_{dif,\Lambda} = R^T_{B,\Lambda} + R^T_{D,\Lambda} + C_t \; R^T_{T,\Lambda} \,. \tag{124}$$

The parameters required to evaluate the previous equations, such as local site slope, $\beta_T(\vec{x})$ $[rad]$, local site aspect, $\zeta_T(\vec{x})$ $[rad]$, and horizon angle, $H_\zeta(\vec{x}, \zeta)$ $[rad]$, can be obtained from the analysis of Digital Elevation Models (DEM). Specifically, in order to calculate the horizon angle, $H_\zeta(\vec{x}, \zeta)$, the viewsheds for each cell $\vec{x}$ of an input DEM should be calculated. A viewshed is the angular distribution of sky visibility versus obstruction. This is similar to the view provided by upward-looking hemispherical (fisheye) photographs. A viewshed is calculated by searching in a specified set of directions around a location of interest. The resolution of the viewshed array must be sufficient to adequately represent all sky directions but small enough to enable rapid calculations, for the following examples an eight directions algorithm is used. Horizon angles for other directions are calculated using interpolation. The penumbral effects are neglected in the code, penumbral refers to decreased direct beam radiation at the edge of shadow due to partial obscuration of the solar disc, considering that the solar disc radius is 0.00466 $[rad]$.

An example of the values assumed by the above mentioned variables is provided in Figure 37 and in Figure 38 for the *Versilia* watershed in Tuscany (Italy). Sky-view factor, S_{vf}, terrain configuration factor, C_t, and shadow effect, S_h, in each cell and for a particular date and hour are calculated using the DEM.

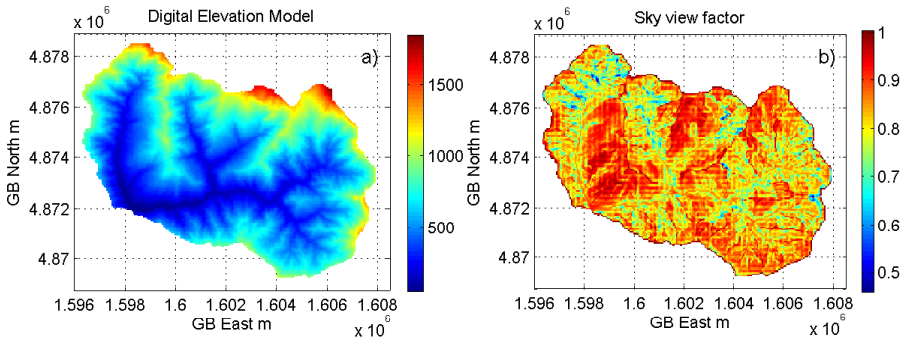


Figure 37: Digital Elevation Model (a), and sky-view factor, S_{vf}, (b) for the the *Versilia* watershed in Tuscany (Italy).

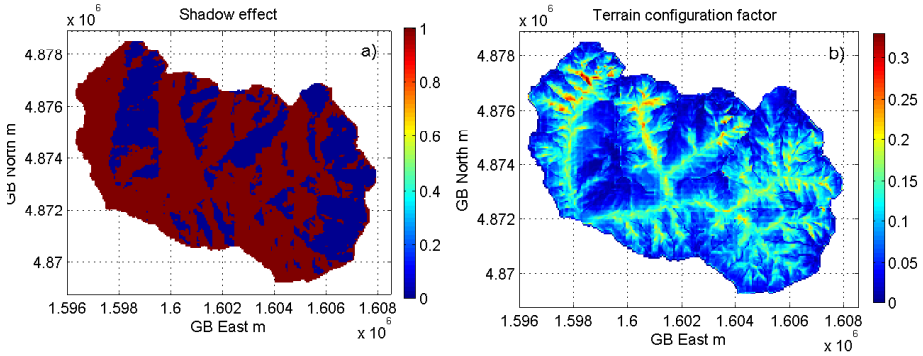


Figure 38: Shadow effect, S_h, (a) and terrain configuration factor, C_t, (b) for the the *Versilia* watershed in Tuscany (Italy). The shadow effect is calculated with sun height in the barycenter of the watershed, the 26 April 1982 at 8 am, local time.

References

Becchi, I., E. Caporali, L. Castellani, and F. Castelli (1994), Multiregressive analysis for the estimation of the spatial zero rainfall probability, in *Workshop on Climate change and hydrogeological hazards in the Mediterranean area*, Perugia, 27-28 June.

Becker, S. (2001), Calculation of direct solar and diffuse radiation in Israel, *International Journal of Climatology*, *21*(12), 1561–1576.

Bertoldi, G., R. Rigon, and T. M. Over (2006), Impact of watershed geomorphic characteristics on the energy and water budgets, *Journal of Hydrometeorology*, *7*, 389–403.

Bonan, G. (2002), *Ecological Climatology: Concept and Applications*, Cambridge Univ. Press, New York.

Bovard, B. D., P. S. Curtis, C. S. Vogel, H.-B. Su, and H. P. Schmid (2005), Environmental controls on sap flow in a northern hardwood forest, *Tree Physiology*, *25*, 31–38.

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

Chaiwiwatworakul, P., and S. Chirarattananon (2004), An investigation of atmospheric turbidity of Thai sky, *Energy and Buildings*, *36*, 650–659.

Chen, Y., A. Hall, and K. N. Liou (2006), Application of three-dimensional solar radiative transfer to mountains, *Journal of Geophysical Research*, *111*(D21111), doi:10.1029/2006JD007163.

Collins, D. B. G., R. L. Bras, and G. E. Tucker (2004), Modeling the effects of vegetation-erosion coupling on landscape evolution, *Journal of Geophysical Research*, *109*(F03004), doi:10.1029/2003JF000028.

Cowpertwait, P. S. P. (1991), Further developments of the Neyman-Scott clustered point process for modeling rainfall, *Water Resources Research*, *27*, 1431–1438.

Cowpertwait, P. S. P. (1994), A generalized point process model for rainfall, *Proceedings of the Royal Society of London, Series A*, *447*, 23–27.

Cowpertwait, P. S. P. (1995), A generalized spatial-temporal model of rainfall based on a clustered point process, *Proceedings of the Royal Society of London, Series A*, *450*, 163–175.

Cowpertwait, P. S. P. (1998), A Poisson-cluster model of rainfall: high order moments and extreme values, *Proceedings of the Royal Society of London, Series A*, *454*, 885–898.

Cowpertwait, P. S. P. (2004), Mixed rectangular pulses models of rainfall, *Hydrology and Earth System Sciences*, *8*, 993–1000.

Cowpertwait, P. S. P. (2006), A spatial-temporal point process model of rainfall for the Thames catchment, *Journal of Hydrology*, *330*, 586–595.

Cowpertwait, P. S. P., P. O'Connell, A. Metcalfe, and J. Mawdsley (1996), Stochastic point process modelling of rainfall: I. Single-site fitting and validation, *Journal of Hydrology*, *175*, 17–46.

Cowpertwait, P. S. P., C. Kilsby, and P. O'Connell (2002), A space-time Neyman-Scott model of rainfall: empirical analysis of extremes, *Water Resources Research*, *38*(8), 1–14.

Cowpertwait, P. S. P., V. Isham, and C. Onof (2007), Point process models of rainfall: developments for fine-scale structure, *Proceedings of the Royal Society of London, Series A*, *463*(2086), 2569–2587.

Cox, R., B. L. Bauer, and T. Smith (1998), A mesoscale model intercomparison, *Bulletin of the American Meteorological Society*, *79*(2), 265–283.

Cunnane, C. (1978), Unbiased plotting positions–a review, *Journal of Hydrology*, *37*, 205–222.

Curtis, D. C., and P. S. Eagleson (1982), Constrained stochastic climate simulation, *Tech. Rep. 274*, Mass. Inst. of Technol. Dep. of Civ. and Environ. Eng. Ralph M. Parsons Lab., Cambridge Mass. USA.

Darula, S., R. Kittler, and C. A. Gueymard (2005), Reference luminous solar constant and solar luminance for illuminance calculations, *Solar Energy*, *79*(5), 559–565.

Deaves, D. M., and I. G. Lines (1997), On the fitting of low mean windspeed data to the Weibull distribution, *Journal of Wind Engineering and Industrial Aerodynamics*, *66*, 169–178.

Dingman, S. L. (1994), *Physical Hydrology*, Prentice-Hall, New Jersey.

Dozier, J., and J. Frew (1990), Rapid calculation of terrain parameters for radiation modeling from digital elevation data, *IEEE Transactions on Geoscience and Remote Sensing*, *28*, 963–969.

Dubayah, R., and S. Loechel (1997), Modeling topographic solar radiation using GOES data, *Journal of Applied Meteorology*, *36*, 141–154.

Dubrovský, M., J. Buchtele, and Z. Žalud (2004), High-frequency and low-frequency variability in stochastic daily weather generator and its effect on agricultural and hydrologic modelling, *Climatic Change*, *63*, 145–179.

Eagleson, P. S. (2002), *Ecohydrology: Darwinian Expression of Vegetation Form and Function*, Cambridge University Press, Cambridge, U.K.

Elshamy, M. E., H. S. Wheater, N. Gedney, and C. Huntingford (2006), Evaluation of the rainfall component of a weather generator for climate impact studies, *Journal of Hydrology*, *326*, 1–24, doi:10.1016/j.jhydrol.2005.09.017.

Entekhabi, D., I. Rodriguez-Iturbe, and P. S. Eagleson (1989), Probabilistic representation of the temporal rainfall process by a modified Neyman-Scott rectangular pulses model: Parameter estimation and validation, *Water Resources Research*, *25*(2), 295–302.

Fatichi, S., V. Y. Ivanov, and E. Caporali (2011), Simulation of future climate scenarios with a weather generator, *Advances in Water Resources*, *In Press*, doi:10.1016/j.advwatres.2010.12.013.

Fiering, M. B., and B. B. Jackson (1971), Synthetic hydrology, *Water resourc. monogr.*, Americ. Geophys. Union.

Foufoula-Georgiou, E., and D. Lettenmaier (1987), A Markov renewal model for rainfall occurrences, *Water Resources Research*, *23*(5), 875–884.

Fowler, H. J., C. G. Kilsby, and P. E. O'Connel (2000), A stochastic rainfall model for the assessment of regional water resource systems under changed climatic conditions, *Hydrology and Earth System Science*, *4*(2), 263–282.

Fowler, H. J., C. G. Kilsby, P. E. O'Connel, and A. Burton (2005), A weather-type conditioned multi-site stochastic rainfall model for the generation of scenarios of climatic variability and change, *Journal of Hydrology*, *308*, 50–66.

Freidenreich, S. M., and V. Ramaswamy (1999), A new multiple-band solar radiative parameterization for general circulation models, *Journal of Geophysical Research*, *104*(D24), 31,389–31,409.

Gueymard, C. A. (1989), A two-band model for the calculation of clear sky solar irradiance, illuminance, and photosynthetically active radiation at the earth's surface, *Solar Energy*, *43*, 253–265.

Gueymard, C. A. (2001), Parameterized transmittance model for direct beam and circumsolar spectral irradiance, *Solar Energy*, *71*, 325–346.

Gueymard, C. A. (2003), Direct solar transmittance and irradiance predictions with broadband models. Part 1: Detailed theoretical performance assessment, *Solar Energy*, *74*, 355–379, corrigendum: *Solar Energy* 76, 513 (2004).

Gueymard, C. A. (2008), REST2: High-performance solar radiation model for cloudless-sky irradiance illuminance and photosynthetically active radiation - validation with a bechmark dataset, *Solar Energy*, *82*, 272–285.

Hansen, J. W., and T. Mavromatis (2001), Correcting low-frequency variability bias in stochastic weather generators, *Agricultural and Forest Meteorology*, *109*, 297–310.

Holben, B. N., T. F. Eck, I. Slutsker, D. Tanr, J. P. Buis, A. Setzer, E. Vermote, J. A. Reagan, Y. J. Kaufamn, T. Nakajima, F. Lavenu, I. Jankowiak, and A. Smirnov (1998), AERONET-A federated instrument network and data archive for aerosol characterization, *Remote Sensing of Environment*, *66*, 1–16.

Ineichen, P. (2006), Comparison of eight clear sky broadband models against 16 independent data banks, *Solar Energy*, *80*, 468–478.

Iqbal, M. (1983), *An Introduction to Solar Radiation*, Academic Press, Toronto.

Istanbulluoglu, E., D. G. Tarboton, and R. T. Pack (2004), Modeling of the interactions between forest vegetation, disturbances, and sediment yields, *Journal of Geophysical Research*, *109*(F01009), doi:10.1029/2003JF00004.

Ivanov, V. Y., R. L. Bras, and D. C. Curtis (2007), A weather generator for hydrological, ecological, and agricultural applications, *Water Resources Research*, *43*, W10,406, doi: 10.1029/ 2006WR005,364.

Ivanov, V. Y., R. L. Bras, and E. R. Vivoni (2008a), Vegetation-hydrology dynamics in complex terrain of semiarid areas: 1. A mechanistic approach to modeling dynamic feedbacks, *Water Resources Research*, *44*(W03429), doi:10.1029/2006WR005588.

Ivanov, V. Y., R. L. Bras, and E. R. Vivoni (2008b), Vegetation-hydrology dynamics in complex terrain of semiarid areas: 2. Energy-water controls of vegetation spatiotemporal dynamics and topographic niches of favorability, *Water Resources Research*, *44*(W03430), doi:10.1029/2006WR005595.

Kasten, F., and G. Czeplak (1980), Solar and terrestrial radition dependent on the amount and type of cloud, *Solar Energy*, *24*, 177–189.

Katz, R. W. (1996), Use of conditional stochastic models to generate climate change scenarios, *Climatic Change*, *32*, 237–255.

Katz, R. W., and M. B. Parlange (1993), Effects of an index of atmospheric circulation on stochastic properties of precipitation, *Water Resources Research*, *29*, 2335–2344.

Katz, R. W., and M. B. Parlange (1998), Overdispersion phenomenon in stochastic modeling of precipitation, *Journal of Climate*, *11*, 591–601.

Kiely, G., J. D. Albertson, M. B. Parlange, and R. W. Katz (1998), Conditioning stochastic properties of daily precipitation on indices of atmospheric circulation, *Meteorological Applications*, *5*.

Kilsby, C. G., P. Jones, A. Burton, A. Ford, H. Fowler, C. Harpham, P. James, A. Smith, and R. Wilby (2007), A daily weather generator for use in climate change studies, *Environmental Modelling & Software*, *22*, 1705–1719.

Kimball, J. S., S. W. Running, and R. Nemani (1997), An improved method for estimating surface humidity from daily minimum temperature, *Agricultural Forest Meteorology*, *85*, 87–98.

Koutsoyiannis, D. (2000), A generalized mathematical framework for stochastic simulation and forecast of hydrologic time series, *Water Resource Research*, *36*(6), 1519–1533.

Koutsoyiannis, D. (2003), Climate change, the Hurst phenomenon, and hydrological statistics, *Hydrological Sciences Journal*, *48(1)*(6), 3–24.

Koutsoyiannis, D., D. Kozonis, and A. Manetas (1998), A mathematical framework for studying rainfall intensity-duration-frequency relationships, *Journal of Hydrology*, *206*, 118–135.

Kumar, L., A. K. Skidmore, and E. Knowles (1997), Modeling topographic variation in solar radiation in a GIS environment, *International Journal of Geographic Information Science*, *11*, 475–497.

Kyselý, J., and M. Dubrovský (2005), Simulation of extreme temperature events by a stochastic weather generator: effects of interdiurnal and interannual variability reproduction, *International Journal of Climatology*, *25*, 251–269.

LeCam, L. (1961), A stochastic description of precipitation, in *Proceedings of Fourth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 3, edited by J. Neyman, pp. 165–176, Univ. of Calif. Press Berkeley.

Leonard, M., M. F. Lambert, A. V. Metcalfe, and P. S. P. Cowpertwait (2008), A space-time Neyman-Scott rainfall model with defined storm extent, *Water Resources Research*, *44*(W09402), doi:10.1029/2007WR006110.

Manning, L. J., J. W. Hall, H. J. Fowler, C. G. Kilsby, and C. Tebaldi (2009), Using probabilistic climate change information from a multimodel ensemble for water resources assessment, *Water Resources Research*, *45*(W11411), doi:10.1029/2007WR006674.

Mavromatis, T., and J. W. Hansen (2001), Interannual variability characteristics and simulated crop response of four stochastic weather generators, *Agricultural and Forest Meteorology*, *109*, 283–296.

McKague, K., R. Rudra, and J. Ogilvie (2003), CLIMGEN - A convenient weather generator tool for Canadian climate stations, in *Meeting of the CSAE/SCGR Canadian Society for engineering in agricultural food and biological systems, Montreal Quebec, July, 6-9*, 03-118, web-site: http://www.sipeaa.it/tools/ClimGen/CSAE03-118.pdf.

McMahon, T. A., and A. J. Miller (1971), Application of the Thomas and Fiering model to skewed hydrological data, *Water Resources Research*, *7*(5), 1338–1340.

Meyer, C. R., C. S. Renshler, and R. C. Vining (2007), Implementing quality control on a random number stream to improve a stochastic weather generator, *Hydrological Processes*, *doi: 10.1002/hyp.6668*.

Montanari, A., R. Rosso, and M. S. Taqqu (1997), Fractionally differenced ARIMA models applied to hydrologic time series: Identification, estimation and simulation, *Water Resources Research*, *33*(5), 1035–1044.

Muneer, T., M. S. Gul, and J. Kubie (2000), Models for estimating solar radiation and illuminance from meteorological parameters, *Journal of Solar Energy Engineering*, *122*, 146–153.

Nelder, J., and R. Mead (1965), A Simplex method for function minimization, *The Computer Journal*, *7*, 308–313.

Nicks, A. D., L. J. Lane, and G. A. Gander (1995), Weather generator, *USDA-Water Erosion Prediction Project: hillslope profile and watershed model documentation*, USDA-ARS National Soil Erosion Research Laboratory, chapter 2, 1-22.

Northrop, P. (1998), A clustered spatial-temporal model of rainfall, *Proceedings of the Royal Society of London, Series A*, *454*, 1875–1888.

Olseth, J. A., and A. Skartveit (1997), Spatial distribution of photosynthetically active radiation over complex topography, *Agricultural and Forest Meteorology*, *86*, 205–214.

Olseth, J. A., A. Skartveit, and H. Zou (1995), Spatially continuous mapping of solar resources in a complex high latitude topography, *Solar Energy*, *55*(6), 475–485.

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

Parlange, M. B., and R. W. Katz (2000), An extended version of the richardson model for simulating daily weather variables, *Journal of Applied Meteorology*, *39*, 610–622.

Ramanathan, V., P. J. Crutzen, J. T. Kiehl, and D. Rosenfeld (2001), Aerosols, climate, and the hydrological cycle, *Science*, *294*, 2119–2124, doi:10.1126/science.1064034.

Rebetez, M., H. Mayer, O. Dupont, D. Schindler, K. Gartner, J. P. Kropp, and A. Menzel (2006), Heat and drought 2003 in Europe: a climate synthesis, *Annals of Forest Science*, *63*, 569–577, doi: 10.1051/forest:2006043.

Richardson, C. W. (1981), Stochastic simulation of daily precipitation, temperature, and solar radiation, *Water Resources Research*, *17*, 182–190.

Richardson, C. W., and D. A. Wright (1984), WGEN: a model for generating daily weather variables, *ARS 8*, USDA-ARS.

Rigon, R., G. Bertoldi, and T. M. Over (2006), GEOtop: A distributed hydrological model with coupled water and energy budgets, *Journal of Hydrometeorology*, *7*(3), 371–388.

Robertson, A., J. Overpeck, D. Rind, E. Mosley-Thompson, G. Zielinski, J. Lean, D. Koch, J. Penner, I. Tegen, and R. Healy (2001), Hypothesized climate forcing time series for the last 500 years, *Journal of Geophysical Research*, *106*(D14), 14,783–14,803, doi:10.1029/2000JD900469.

Rodriguez-Iturbe, I., and P. S. Eagleson (1987), Mathematical models of rainstorm events in space and time, *Water Resources Research*, *23*(1), 181–190.

Rodriguez-Iturbe, I., D. Cox, and V. Isham (1987), Some models for rainfall based on stochastic point processes, *Proceedings of the Royal Society of London, Series A*, *410*, 269–288.

Rodriguez-Iturbe, I., D. Cox, and V. Isham (1988), A point process model for rainfall: further developments, *Proceedings of the Royal Society of London, Series A*, *417*, 283–298.

Russell, P. B., J. Redemann, B. Schmid, R. W. Bergstrom, J. M. Livingston, D. M. M. S. A. Ramirez, S. Hartley, P. V. Hobbs, P. K. Quinn, C. M. Carrico, M. J. Rood, E. strm, K. J. Noone, W. V. Hoyningen-Huene, and L. Remer (2002), Comparison of aerosol single scattering albedos derived by diverse techniques in two North Atlantic experiments, *Journal of Atmospheric Sciences*, *59*, 609–619.

Semenov, M. A. (2008), Simulation of extreme weather events by a stochastic weather generator, *Climate Research*, *35*, 203–212.

Semenov, M. A., and E. M. Barrow (1997), Use of a stochastic weather generator in the development of climate change scenarios, *Climatic Change*, *35*, 397–414.

Semenov, M. A., and E. M. Barrow (2002), *LARS-WG, A Stochastic Weather Generator for Use in Climate Impact Studies, Version 3.0*, user Manual. http://www.rothamsted.ac.uk/mas-models/download/LARS-WG-Manual.pdf.

Semenov, M. A., and J. R. Porter (1995), Climatic variability and the modelling of crop yields, *Agricultural and Forest Meteorology*, *73*, 265–283.

Semenov, M. A., R. J. Brooks, E. M. Barrow, and C. W. Richardson (1998), Comparison of the WGEN and LARS-WG stochastic weather generators for diverse climates, *Climate Research*, *10*, 95–107.

Sharpley, A. N., and J. R. Williams (1990), EPIC-Erosion/Productivity impact calculator: 1. Model documentation, *Technical Bulletin 1768*, US Department of Agriculture, 235 p.

Slingo, A. (1989), A GCM parameterization for the shortwave radiative properties of water clouds, *Journal of Atmospheric Sciences*, *46*(10), 1419–1427.

Srikanthan, R., and T. A. McMahon (1982), Simulation of annual and monthly rainfalls a preliminary study at five Australian stations, *Journal of Applied Meteorology*, *21*, 1472–1479.

Srikanthan, R., and T. A. McMahon (2001), Stochastic generation of annual, monthly and daily climate data: a review, *Hydrology and Earth System Science*, *5*, 653–670.

Stephens, G. L. (1978), Radiation profiles in extended water clouds: 2. Parameterization schemes, *Journal of Atmospheric Sciences*, *35*(11), 2123–2132.

Takle, E. S., and J. M. Brown (1978), Note on the use of Weibull statistics to characterize wind-speed data, *Journal of Applied Meteorology*, *17*, 556–559.

Todorovic, P., and D. A. Woolhiser (1975), A stochastic model of n-day precipitation, *Journal of Applied Meteorology*, *14*, 17–24.

Todorovic, P., and V. Yevjevich (1969), Stochastic process of precipitation, *Tech. rep.*, Colo. State Univ. Fort Collins, Fort Collins, Colorado.

TVA (1972), Heat and mass transfer between a water surface and the atmosphere., *Tech. Rep. 14*, Tennessee Valley Authority Water Resources Research Engineering Laboratory, Norris, TN.

Veneziano, D., and P. Furcolo (2002), Multifractality of rainfall and intensity-duration-frequency curves, *Water Resources Research*, *38*(12), doi:10.1029/2001WR000,372.

Veneziano, D., and V. Iacobellis (2002), Multiscaling pulse representation of temporal rainfall, *Water Resources Research*, *38*(8), doi: 10.1029/2001WR000,522.

Veneziano, D., P. Furcolo, and V. Iacobellis (2002), Multifractality of iterated pulse processes with pulse amplitudes generated by a random cascade, *Fractals*, *10*(2), 209–222.

Veneziano, D., A. Langousis, and P. Furcolo (2006), Multifractality and rainfall extremes: A review, *Water Resources Research*, *42*, W06D15, doi:10.1029/2005WR004,716.

Wallis, T. W. R., and J. F. Griffiths (1997), Simulated meteorological input for agricultural models, *Agricultural and Forest Meteorology*, *88*, 241–258.

Waymire, E. C., and V. K. Gupta (1981a), The mathematical structure of rainfall representation, 1. A review of the stochastic rainfall models, *Water Resources Research*, *17*(5), 1261–1272.

Waymire, E. C., and V. K. Gupta (1981b), The mathematical structure of rainfall representation, 2. A review of the point process theory, *Water Resources Research*, *17*(5), 1273–1285.

Waymire, E. C., and V. K. Gupta (1981c), The mathematical structure of rainfall representation, 3. Some applications of the point process theory to rainfall processes, *Water Resources Research*, *17*(5), 1287–1294.

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

Wilks, D. (1989), Conditioning stochastic daily precipitation models on total monthly precipitation, *Water Resources Research*, *23*, 1429–1430.

Wilks, D. S. (1992), Adapting stochastic weather generation algorithms for climate changes studies, *Climatic Change*, *22*, 67–84.

Wilks, D. S. (1999), Interannual variability and extreme-value characteristics of several stochastic daily precipitation models, *Agricultural and Forest Meteorology*, *93*, 153–169.

Wilks, D. S., and R. L. Wilby (1999), The weather generation game: a review of stochastic weather models, *Progress in Physical Geography*, *23*, 329–357.

Wilson, E. B., and M. M. Hilferty (1931), Distribution of Chi-square, *Proc. Natl. Acad. Sci. U.S.A.*, *17*, 684–688.