Research Papers
Issue RP0176
July 2013

ISC - Impacts on Soil and Coasts Division

Using a weather generator to simulate daily precipitation scenarios from seasonal weather forecasts

By **Renata Vezzoli**
Impacts on Soil and Coasts Division, CMCC - Capua
renata.vezzoli@cmcc.it

Silvano Pecora
SIMC, ARPA Emilia Romagna - Parma
specora@arpa.emr.it

Maria Paola Manzi
Impacts on Soil and Coasts Division, CMCC - Capua
mariapaola.manzi@cmcc.it

SUMMARY This work presents the conditioned upscaling procedure applied in the Po-FEWS system by Hydrology area of ARPA SIMC Emilia Romagna to select the rainfall daily scenarios which better fit seasonal meteorological forecasts. Seasonal meteorological forecast provides qualitatively useful information for management activities over a short-medium term horizon, but their temporal aggregation time is too coarse to be allow quantitatively considerations. For this reason, Hydrology area of ARPA SIMC Emilia Romagna developed a conditioned upscaling procedure to obtain daily scenarios coherent with the seasonal forecasts. The daily scenarios are the inputs to the hydrological/hydraulics chains available in the Po-FEWS system for short-medium term forecasts. The daily rainfall scenarios are generated through a Spatio-Temporal Neymann Scott Rectangular Pulse model calibrated using daily observations over the period 1987-2008. Seasonal forecasts provide also the expected anomalies of minimum and maximum temperatures. Minimum and maximum temperatures are simulated at daily scale using a multilinear regression model conditioned on rainfall scenarios.

Keywords: Weather Generator, conditioned bottom-up approach

The research leading to these results has received funding from the Italian Ministry of Education, University and Research and the Italian Ministry of Environment, Land and Sea under GEMINA and NEXTDATA projects.

Introduction

The study of climate change impacts on hydrology and activities like water managements requires the availability of data at a fine spatio-temporal resolution, that usually is not available from GCMs simulations. The problem of increasing the model resolution has been addressed through downscaling techniques. Downscaling methodologies can be divided into two main categories, dynamic (Regional Climate Models, RCMs) and statistical. In the dynamic downscaling, the GCM acts as boundary and initial conditions for the RCM simulation over the region of interest with a finer spatial resolution than the original one. Within GEMINA project the RCM is COSMO-CLM that provides outputs in a spatial range among 1 and 50 kms, [16]. This resolution range is adequate to capture the geographical variation of the atmospherical features, in particular temperature and precipitation, the main parameters of interest for hydrological impact studies. Statistical downscaling techniques are based on statistical models applied to historical data, [15]. Techniques like regression models, weather pattern classification schemes, weather time series generators are classified as statistical downscaling method. Statistical downscaling can reach a resolution finer that dynamical downscaling, depending on the observations availability, and it is generally less demanding from a computational point of view. On the other side, in absence of a large amounts of observations statistical downscaling can not be performed and the validity of the relationships between predictands (the variables to be downscaled) and predictors (the GCMs variables) is limited to the range of the data used for calibration, while future projections may not belong to that range, [3, 16]. Within GEMINA project the spatial downscaling of daily data will based either on dynamical downscaling approach with COSMO CLM as regional climate model and statistical downscaling tecniques like MOS analog, linear scaling and quantile mapping [13].

The climate modelling chain implemented in GEMINA provides through COSMO-CLM daily precipitation and temperature series that are used to run long term hydrological simulations, see the simulation scheme proposed in [14], with the aim to characterise the capability of the chain to reproduce hydrological extremes and mean values and to study their change under the scenarios RCP 4.5 and RCP 8.5 [5]. However, rainfall data at hourly time scale are requested to address the impacts of climate on floods, specially in small catchments characterised by a concentration time of few hours, [9]. The Hydrology area of ARPA SIMC has already developed a procedure of (statistical) temporal downscaling based on a weather generator approach. Currently the applied procedure is used to downscale seasonal (3 months) weather forecasts to a daily scale, to obtain a 3 months forecast of the hydrological variables for water management policies. This paper describes how the temporal downscaling is addressed using a conditioned bottom-up procedure based on a weather generator technique to simulate series of rainfall and temperature at Po river basin scale. The approach allows to obtain series that are consistent with observations and among themselves avoiding situation like, e.g., two sites closely located that, at a given instant, experience extremely different precipitation values.

The rainfall daily scenarios are generated by a Spatio-Temporal Neymann Scott Rectangular Pulse model, [1, 3], calibrated using daily observations from the period 1987-2001, that is the reference period for the meteorological center of ARPA SIMC. The minimum and maximum temperature daily fields are derived from rainfall scenarios through linear multi-regression relationships calibrated from the observed data. By the available dataset over the whole basin, the analysis performed suggests to simulate the temperature field using a simple AR(1) model or a slightly more complex model conditioned on rainfall occurrence like the one used by [10], see [6]. This kind of approach is of particular interest even for NEXTDATA project, one of whose aims is the usage of the scenarios for impact studies of the climate variability, especially concerning the hydrogeological response. In particular, the procedure used for temporal statistical downscaling is applicable within this project, since one of the focal point is the comparison of the climate regionalization methods on the Alpine area in terms of cumulative rainfall values on sub-daily time scale and spatial scale less than 10 km.

Weather Generator Literature review

A weather generator is a statistical tool able to simulate, from observed statistics, atmospheric variables like rainfall, temperature, relative humidity, solar radiation, etc., in one or more sites. The simulated variables reflects the statistical behaviour of the observed ones. Weather generators are usually structured into 2 steps:

- simulation of rainfall synthetic time series;
- simulation of the other climate variables eventually conditioned on rainfall occurrence (for the purpose of Hydrology area of ARPA SIMC the variables of interest are minimum and maximum temperature).

A "climate change" module can be added to incorporate/adapt the results of the weather generator to future climate scenarios.

The weather generators are calibrated at monthly scale to better reproduce data seasonality and the temporal correlation structure. The weather generators are divided into two main categories: Richardson-type, [10, 11] and serial generators [8, 12].

In the next paragraphs we provide a short description of Richardson-type, serial, and a multisite weather generator based upon Spatio Temporal Neymann Scott Rectangular Pulses.

RICHARDSON-TYPE weather generators simulate rainfall occurrence like a first order, two states Markov chain one for rainfall occurrence (wet), the other to indicate rainfall absence (dry), thus the Markov chain is characterized by the four transition probabilities p_{DD}, p_{DW}, p_{WD}, and p_{WW}. The rainfall model can be made more complex and include additional classes for precipitation (low/medium/high) or have a longer memory, in these cases to characterise the Markov chain a higher number of parameters is requested. Once that the Markov chain assigns the state “wet” the rainfall amount is extracted from a distribution probability like the Γ. The other climate variables are derived through a multivariate (linear) regression that depends on rainfall and temporal correlation between rainfall and the variables to be simulated.

The minimum and maximum temperature (and other derived variables) are obtained from a autoregressive model conditioned on precipitation [12]. Given the precipitation occurrence at the t-th day, the temperature (minimum/maximum) is assumed to be normal with mean and variance conditioned on month and wet/dry status. The randomly standardized variables

$$T_k^*(t) = \frac{T_{i,k}(t) - \mu_{i,k}}{\sigma_{i,k}} \tag{1}$$

with $i = W, D$ and $k = \min, \max$ can be simulated as

$$\begin{aligned} T^*_{\min}(t) &= a_{11}T^*_{\min}(t-1) + a_{12}T^*_{\max}(t-1) + \\ &\quad + b_{11}\epsilon(t) + b_{12}\delta(t) \\ T^*_{\max}(t) &= a_{12}T^*_{\min}(t-1) + a_{22}T^*_{\max}(t-1) + \\ &\quad + b_{21}\epsilon(t) + b_{22}\delta(t) \qquad (2) \end{aligned}$$

where ϵ and δ are random variables normally distributed with mean equal zero and unitary variance and $\{a_{ij}\}$, $\{b_{ij}\}$ are function of the lag 0 and lag 1 correlation matrices M_0 and M_1:

$$A = M_1 M_0^{-1} \qquad (3)$$

$$BB^T = M_0 - AM_1^T. \qquad (4)$$

Thus, once the normalized temperatures are simulated from equation (2), the synthetic temperature time series are obtained inverting the equation (1) using the appropriate (conditioned on precipitation status) mean and variance.

A disadvantage of Richardson-type weather generators is that they are not able to reproduce the persistence of dry/wet periods.

SERIAL GENERATORS overby the problem of the wet/dry persistence simulating the sequence of wet and dry periods. The length of these periods is randomly chosen from their probability distributions (conditioned on the month the series stars). Only after the simulation of the dry/wet sequence, the rainfall amount is associated to the wet days according to the distribution function of the precipitation intensity. The procedure to simulate temperature and other derived variables is analogous to the one presented for Richardson-type weather generators.

Richardson-type and serial weather generators have been developed for one-site application. If, there is the need to simulate more than one site it can happen that the results are statically correct, if considered site by site, but non coherent in their spatial distribution [7].

Instead the spatial coherence is a prerequisite for hydrological application. To overcome the lack of spatial coherence a spatio-temporal rainfall generator has been identified as a more adequate tool for hydrological tool. The model is based on Spatio-Temporal Neymann Scott Rectangular Pulse (STNSRP). The model has been developed and tested at the University of Newcastle upon Tyne [4, 1] and a software named RAINSIM is available. The model is able to produce rainfall series of any length and time resolution down to minutes, [4, 1].

THE SPATIO-TEMPORAL NEYMANN SCOTT RECTANGULAR PULSE for rainfall is a generalization of the Temporal Neymann Scott Rectangular Pulse (TNSRP) model [2]. In the TNSRP model the rainfall events are described through a cluster of rain cells. The time intervals between the storm origin and the birth of the individual cells is modelled by a set of independent and identically distributed random variables. The model structure is based on the following assumptions:

- storm origins is a Poisson process with parameter λ;
- each storm is associated to a random number of rain cells distributed as a Poissonian of parameter ν;
- the time interval between the storm origin and the birth of a rain cell is exponentially distributed with parameter β;
- the duration and intensity of a rain cell are exponentially distributed with parameters η and ξ, respectively;

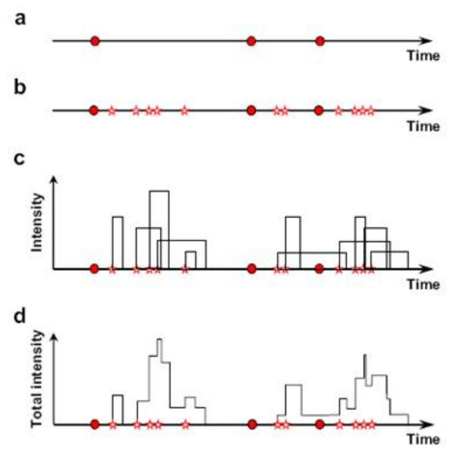

Figure 1:
A schematic representation of the TNSRP model from [4].

- the rainfall intensity is the sum of the contributes of all the rain cells that are active in a certain instant.

Figure 1 gives a graphical representation of the TNSRP model and Table 1 summarizes the parameters.

The STNSRP model has two additional parameters: γ, related to the rain cell dimension and carries the dimension of 1/km, and ρ, in 1/km^2, representing the spatial density of the centres of the rain cells and a scaling factor matrix, ϕ, adimensional. Thus the STNSRP model requires to estimate 7 parameters (λ, β, η, ξ, ν, γ, and ρ) for each month and the scaling factor matrix ϕ characterised by one value for each couple site-month.

[4] applied a four states (wet-dry, dry-wet, wet-wet, dry-dry, meaning a wet day followed by a dry day and so on) approach to characterised the relationships between rainfall and the derived variables, like temperature. Temperature field are expressed in term of mean temperature (T) and range (R). The mean temperature is defined as the average value of maximum and minimum temperature, the range is given by the difference between maximum and minimum temperature. The procedure to estimate the parameters of the linear regression (and, after, to apply it to simulate the variables) requires to (a) divide each month into two halves, (b) divide the variables into four groups : each group is associated to one state (wet-wet, dry-dry, wet-dry, and dry-wet), (c) normalise each group with respect to its own mean and variance, (d) estimate the parameters of the regression model associate to each state: for the wet-wet case, the model is

$$
\begin{aligned}
T^*(t) &= a_1 T^*(t-1) + b_1\delta(t) + c_1\epsilon(t) \\
R^*(t) &= a_2 R^*(t-1) + b_2\delta(t) + c_2\delta(t), \quad (5)
\end{aligned}
$$

for the dry-dry case:

$$
\begin{aligned}
T^*(t) &= a_3 T^*(t-1) + b_3\delta(t) + c_3\epsilon(t) \\
R^*(t) &= a_4 R^*(t-1) + b_4\delta(t) + c_4\delta(t), \quad (6)
\end{aligned}
$$

for the dry-wet case:

$$
\begin{aligned}
T^*(t) &= a_5 T^*(t-1) + a_6 P^*(t) + \\
&\quad + b_5\delta(t) + c_5\epsilon(t) \\
R^*(t) &= a_7 R^*(t-1) + a_8 P^*(t) + \\
&\quad + b_6\delta(t) + c_6\delta(t). \quad (7)
\end{aligned}
$$

for the wet-dry case:

$$
\begin{aligned}
T^*(t) &= a_9 T^*(t-1) + a_{10} P^*(t-1) + \\
&\quad + b_7\delta(t) + c_7\epsilon(t) \\
R^*(t) &= a_{11} R^*(t-1) + a_{12} P^*(t-1) + \\
&\quad + b_8\delta(t) + c_8\delta(t), \quad (8)
\end{aligned}
$$

where ϵ and δ are random variables normally distributed with mean equal zero and unitary

Table 1
List of parameters of the TNSRP model.

Parameter	Description	Unit
λ^{-1}	average time between two storm origins	h
β^{-1}	average waiting time for one rain cell after a storm origin	h
η^{-1}	average cell duration	h
ξ^{-1}	average cell intensity	mm/h
ν^{-1}	average number of cells per storm	adim

variance. Each regression parameter assumes 24 values (one for each possible half month).

In addition, [4] present a methodology based on change factors to project observed statistics and thus modify the STNSRP model parameters estimates according to climate scenarios. The change factors are estimated from the statistics of control and future scenarios and applied to the observed statistics, for rainfall the change factors are multiplicative factors while for the other variables are additive factors variable along the year months.

The change factor (α) for rainfall is defined as

$$\alpha = \frac{P^{Fut}}{P^{Obs}} = \frac{P^{C,Fut}}{P^{C,Ref}} \tag{9}$$

where the suffix "Obs" stands for observed values, "Fut" for forecasted variables, "C,Ref" for reference climate model and "C,Fut" for projected climate model, therefore $P^{Fut} = \alpha \times P^{obs}$. The change factor for the probability of having a dry day ($Pdry$) is derived as

$$\beta = \frac{X({P_{dry}}^{Fut})}{X({P_{dry}}^{Obs})} = \frac{X({P_{dry}}^{C,Fut})}{X({P_{dry}}^{C,Ref})} \tag{10}$$

where $X(P_{dry}) = P_{dry}/(1 - P_{dry})$ and

$${P_{dry}}^{Fut} = \frac{\beta X({P_{dry}}^{Obs})}{(1 + \beta X({P_{dry}}^{Obs}))}. \tag{11}$$

For temperature, the change factor (δ) is given by the difference between $T^{C,Fut}$ and $T^{C,Ref}$, thus $T^{Fut} = T^{Obs} + \delta$.

Figure 2:
The coverage of the grid of seasonal forecasts (·) of the Po river basin (blue line).

ARPA Emilia Romagna seasonal weather forecast

The meteorological center of ARPA SIMC provides to the Hydrology area, each month, a seasonal (3 months) weather forecast of the climate anomaly (precipitation, temperature) over Italy based on the forecasts of the European Center for Medium-Range Weather Forecast (ECMWF). The seasonal forecast are available over a grid of about 30 x 30 km, Figure 2.

For hydrological purposes the 3-months aggregation of the forecasts is too coarse and there is the need to downscale them to a finer temporal resolution (days/hours). The Hydrology area of ARPA SIMC approached the problem through a conditioned bottom-up procedure, [6]. The bottom-up procedure is articulated into four main steps:

1. Multisite generation of synthetic rainfall time series (spatio-temporal scenarios) through the RAINSIM software [4];

2. Temporal aggregation of the spatio-temporal scenarios at 3 months, i.e., the forecast temporal resolution;

3. Comparison between seasonal forecast and aggregated synthetic time series at basin scale to select those that have a total precipitation over the whole basin similar to the forecasted one (global criteria);

4. Among the scenarios that satisfy the global criteria, comparison between seasonal forecast and aggregated synthetic time series for each cell grid falling into the Po river basin; the simulations with a precipitation amount similar to the forecasted one in at least the 80% of the cells are considered as valid simulations (local criteria).

The first two step are independent from the seasonal forecasts while the satisfaction of global and local criteria is function of the seasonal forecasts. The main advantage of this procedure is that the computational efforts to generate synthetic scenarios is limited to the initial phase and is requests only once in a while. In addition the parameters of the rainfall generator account for the spatial relationship between rainfall among the different locations, this will guarantee the spatial coherence of the synthetic fields, [1].

Since steps 3 and 4 do not allow to select a prefixed number of scenarios, a ranking procedure has been developed and implemented. The ranking procedure computes, at monthly/seasonal scale and at basin/cell level, the absolute difference between the seasonal forecast and the synthetic scenarios, identifying those that are "more close" to the forecast. In this way, it is possible to select a certain prefixed number of scenarios knowing their distance/error from the forecast. If any of the scenarios is close enough to the seasonal forecast, steps (1) and (2) can be repeated expanding the scenarios database.

According to the available data, the temperatures field are simulated using a simple AR(1) model or a model conditioned on rainfall occurrence like the one used by [10] instead of the model proposed by [4]. To ensure the coherence with the anomalies identified in the seasonal forecast, in the temperature field generation the monthly mean value is substituted by the seasonally forecasted temperature.

Conclusion

The procedure individuated by the Hydrology area of ARPA SIMC requires minimal computational efforts and time to associate daily scenarios to the seasonal forecasts. The daily scenarios are used to feed the hydrological/hydraulics chains in the Po-FEWS system to obtain qualitative and quantitative forecast useful for short-medium term water management activities. The same approach applied to generate hourly scenarios from daily field can be useful for flood forecasting and management activities where the daily scale is too coarse.

Bibliography

[1] A. Burton, C.G. Kilsby, H.J. Fowler, P.S.P Cowpertwait, and P.E. O'Connell. RainSim: A spatial-temporal stochastic rainfall modelling system. *Environmental Modelling and Software*, 23:1356–1369, 2007.

[2] P.S.P Cowpertwait, C.G. Kilsby, and P.E. O'Connell. A spatial temporal Neyman-Scott model of rainfall: empirical analysis of extremes. *Water Resources Research*, 38(8):1311, 2002.

[3] H.J. Fowler, S. Blenkinsop, and C. Tebaldi. Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling. *International Journal of Climatology*, 27:1547–1578, 2007.

[4] C.G. Kilsby, P.D. Jones, A. Burton, A.C. Ford, H.J. Fowler, C. Harpham, P. James, A. Smith, and R.L. Wilby. A daily weather generator for use in climate change studios. *Environmental Modelling and Software*, 22:1705–1719, 2007.

[5] M. Meinshausen, S.J. Smith, K. Calvin, J.S. Daniel, J.-F. Lamarque, K. Matsumoto, S. Montzka, S.C.B. Raper, K. Riahi, A. Thomson, G.J.M. Velders, and D.P. van Vuuren. *The RCP Greenhouse Gas Concentrations and their Extensions from 1765 to 2500*, chapter Climatic Change. 2011.

[6] C. De Michele, R. Rosso, and R. Vezzoli. Relazione di sintesi - Parte II - Serie sintetiche di precipitazione e temperatura congruenti con previsioni trimensili. Technical report, Politecnico di Milano, ARPA Emilia Romagna, Autorità di bacino del fiume Po, Agenzia Interegionale per il fiume Po, 2009.

[7] M.B. Parlange and R.W. Katz. An extended version of the Richardson model for simulating daily weather variables. *Journal of Applied Meteorology*, 39:610–622, 2000.

[8] P. Racsko, L. Szeidl, and M. Semenov. A serial approach to local stochastic weather models. *Ecological Modelling*, 57(1-2):27–41, 1991.

[9] N. Rebora, L. Ferraris, J. von Hardenberg, and A. Provenzale. Rainfall downscaling and flood forecasting: a case study in the Mediterranean area. *Natural Hazards and Earth System Sciences*, 6:611–619, 2006.

[10] C.W. Richardson. Stochastic simulation of daily precipitation, temperature and solar radiation. *Water Resources Research*, 17:182–190, 1981.

[11] C.W. Richardson and D.A. Wright. WGEN: a model for generating daily weather variables. Technical Report 8, U.S. Dept. of Agriculture, Agricultural Research Service, Springfield, Va. : National Technical Information Service (NTIS), 1984.

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

[13] M. Turco, A.L. Zollo, G. Rianna, R. Vezzoli, L. Cattaneo, and P. Mercogliano. Post-processing methods for cosmo-clm precipitation over italy. *CMCC - Research Paper*, 2013.

[14] R. Vezzoli, P. Mercogliano, S. Pecora, M. Montesarchio, L. Cattaneo, M. del Longo, and F. Tonelli. Performance evaluation of an hydrological modelling chain forced by the regional climate model COSMO-CLM driven by ERA40 Reanalysis. *CMCC - Research Paper*, in prep.

[15] R.L. Wilby and T.M.L. Wigley. Downscaling general circulation model output: a review of methods and limitations. *Progress in Physical Geography*, 21:530–548, 1997.

[16] A.L. Zollo, P. Mercogliano, M. Turco, R. Vezzoli, G. Rianna, E. Bucchignani, M.P. Manzi, and M. Montesarchio. Architectures and tools to analyse the impacts of climate change on hydrogeological risk on Mediterranean area. *CMCC - Research Paper*, 2012.