SCENARIOS OF HYDROMETEOROLOGICAL VARIABLES BASED ON AUXILIARY DATA FOR WATER STRESS RETRIEVAL IN CENTRAL TUNISIA

N. Farhani, J. Carreau, G. Boulet, Z. Kassouk, B. Mougenot, M. Le Page, Z. L. Chabaane and R. Zitouna

N. Farhani is with GREEN-TEAM (U. Carthage / INAT), Tunis, Tunisia and CESBIO (IRD / U. Toulouse), Toulouse, France

ABSTRACT

Characterization of plant water use, generally determined from evapotranspiration, together with water stress, derived from remote sensing data in the thermal infrared domain, are needed to better manage water resources. Evapotranspiration and water stress can be simulated by a dual source energy balance model that combines satellite and in situ hydrometeorological information. Available hydrometeorological observations are often insufficient to account for the spatial and temporal variability of the area of interest. To address this issue, we developed a stochastic weather generator that relies on ERA5 reanalyses and provides spatio-temporal scenarios of multiple hydrometeorological variables. The generator is evaluated and compared with two bias correction methods in terms of their ability to reproduce both observed hydrometeorological variables and simulated evapotranspiration and water stress in central Tunisia. Our analyses show that the stochastic generator offers interesting advantages to perform gap-filling and to extend the hydrometeorological time series in the past.

***Index Terms*—** stochastic generator, bias correction, reanalyses, evapotranspiration, dual source energy balance model

1. DROUGHT MONITORING

In arid and semi-arid areas, water is a major limitation factor for agricultural production. The vulnerability to climate change and, in particular, to drought periods is high. Indeed, these areas are characterized by a short rainy season and strong irregularity in time and space of precipitation events. This induces more frequent annual and intra-seasonal droughts. Natural variations in climate water cycle affect the availability of water and form the main driver of droughts. Agriculture and irrigation are particularly influenced by water deficit.

An important issue in climate and hydrology is to improve the monitoring of droughts and the prediction of their occurrence in the future. This requires a better understanding of the physical mechanisms that lead to this phenomenon. To analyze the actual water use over a long period in the past is a key component of better drought management. It allows to establish a temporal analysis and a monitoring of agricultural practices. Another key component is the quantification of droughts to help analyze drought occurrences. Stress indices can be established to quantify periods of droughts according to their frequency, intensity, spatial extension and duration. To this end, long series of drought observations are required.

2. VEGETATION STRESS

In this work, we are interested in the vegetation stress. Indeed, the state of vegetation is generally representative of environmental water stress especially in arid and semi-arid areas [1]. Plant water use is generally computed based on evapotranspiration estimation which is the preponderant component of the terrestrial water balance and is a key factor for scarce water resources management. The quantitative state of drought is defined by a stress index. Owing to the temporal and spatial scales of climate variability, evapotranspiration and water stress index must be monitored at subdaily to daily scales. Therefore, we choose to compute evapotranspiration using energy balance methods that combine from medium to low resolution remote sensing (RS) data. RS data in the thermal infrared domain is particularly informative for monitoring agrosystem health and adjusting irrigation requirements. In water deficit condition, plants reduce their transpiration rates to preserve the remaining water. This reduced evaporative part in the leaf surfaces induces a detectable thermal signal of elevated canopy temperatures that can be measured from thermal infrared sensors [2]. Thus, surface energy balance uses surface temperature to solve the coupled equations of sensible, latent and heat energy [1].

We rely on a dual source energy balance model [3] that allows retrieval of separate estimates of evaporation and transpiration. High temporal and spatial resolution data are required. For this reason, we need long time series of hydrometeorological variables (air temperature, relative air humidity, global radiation and wind speed) and satellite information (NDVI, LAI, albedo and surface temperature obtained from the TERRA and AQUA sensors of the MODIS satellite). However, the available observations are often insufficient to account for the spatial and temporal variability of the area of

Thanks to LMI NAÏLA and PHC Utique AMANDE for funding.

interest due to the sparsity of gauged networks, the lack of long observation periods and the presence of numerous gaps.

3. HYDROMETEOROLOGICAL SCENARIOS

To perform imputation of missing data and to extend the observation period in the past, we developed a stochastic weather generator that relies on low resolution reanalysis data to generate scenarios of hydrometeorological variables at a temporal resolution of 30 min [4]. The stochastic weather generator is based on generalized linear models (GLM) with the following probability distributions : Gaussian for air pressure, wind speed (combined with the transformation $\log(1 + \exp(y))$), air temperature, relative humidity (combined with the transformation $\tan(\pi(y - 0.5))$) and for global radiation (combined with the transformation $\ln[\max(\ln(y)) - \ln(y)]$) ; Bernoulli for precipitation occurrence and Gamma for precipitation intensity. In addition to the low resolution counterpart of each hydrometeorological variable provided by the reanalysis data, a subset of hydrometeorological variables is used as covariates to introduce inter-variable dependencies, see Fig. 1. Further spatio-temporal effects can be taken into account through geographical coordinates, covariates to reproduce the seasonal cycle $(\cos(^{2\pi d}/_{k}), \sin(^{2\pi d}/_{k})$ with $k \in (183, 365, 91, 30)$ and where $1 \leq d \leq 365$ is the day of the year) and the diurnal cycle $(\cos(^{2\pi h}/_{k}), \sin(^{2\pi h}/_{k})$ with $k \in (24, 12, 6)$ and where $1 \leq h \leq 24$ is the hour of the day). Memory effects to account for temporal persistence can be included as lagged values, either directly from the variable of interest, from spatial averages (the average of the values at all the sites at the given time step), from moving averages with a window of one day (48 time steps) or less, or from a combination of a spatial and moving averages. Selection among the spatio-temporal and memory effects is performed in two steps : (1) LASSO [5] is applied as an approximate regression model to screen an initial large set of covariates and (2) goodness-of-fit criteria [6, 7] are used to remove further covariates without deteriorating the fit.

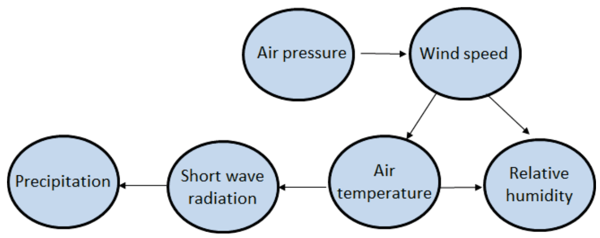


Fig. 1. Inter-variable dependence graph.

We compare the proposed stochastic weather generator with two bias correction methods which also exploit low resolution reanalysis data. Indeed, the reanalysis data provide a multivariate, spatially complete and coherent record of the global atmospheric circulation [8]. In addition, they are available for a long period in the past (from 1979 till now). Nevertheless, its spatial resolution is low from 31 km (ERA 5) to 75 km (ERA interim) thus local variability is not accounted for [9]. To use such reanalysis data in energy balance models, bias correction methods can be applied to account for the difference in spatial resolution between reanalyses and observations from the gauged network. By working on anomalies of diurnal cycles computed for three seasons (june to august, november to march and the remaining months), existing bias correction methods can be adapted to sub-daily temporal resolution. We consider CDFt [10] and MBCn [11], which are state-of-the-art univariate and multivariate bias correction methods respectively.

4. SCENARIOS EVALUATION

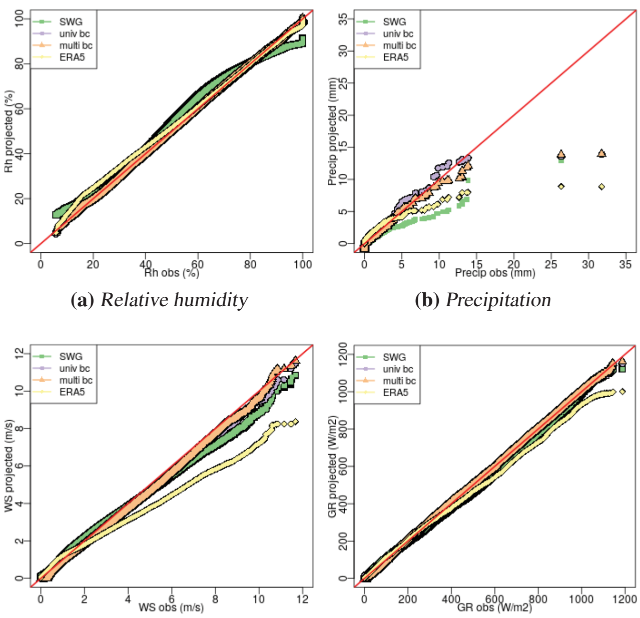
The evaluation is carried out on hydrometeorological time series from three gauged stations in the Kairouan area in central Tunisia which is subject to semi-arid climate. Observations are available at a 30 min time step over 2012-2016 (five years). First, the ability of the stochastic weather generator to reproduce seasonal and diurnal cycles is assessed. It is then evaluated in terms of its gap-filling ability by introducing artificially 15% missing values throughout the hydrometeorological time series. The performance of the generator is satisfactory concerning both the reproduction of the distribution of each variable, as assessed by quantile-quantile plots, and of the inter-variable dependency, as measured by Kendall's rank correlation coefficient [12] .

The stochastic weather generator is compared with the two bias correction methods in terms of their projection ability, i.e. to extend hydrometeorological time series over a time period that has no overlap with the calibration period. To this end, we rely on a cross-validation procedure in which each year is left aside in turn and statistical relationships are calibrated on the remaining years. The simulated scenarios obtained with the cross-validation procedure (either by the stochastic generator, the univariate bias correction method or the multivariate bias correction method) are compared with the observations and the raw ERA5 reanalyses.

On one hand, we evaluate how accurately the scenarios reproduced the observations from the gauged stations. Quantile-quantile plots for four hydrometeorological variables are shown in Fig. 2 (not shown for air temperature and air pressure because they are equivalent for all scenarios). Fig. 2 shows that, hydrometeorological variables derived from different scenarios are relatively comparable. However, the relative humidity and precipitation, Figs. 2a and 2b, are better reproduced by the two bias correction methods (as can be seen from the perfect alignment with the first bisector) than by the stochastic generator and reanalyses. In addition, reanalyses underestimate wind speed observations as can be seen from the fact that they are below the first bisector in

Fig. 2c. Finally, the global radiation distribution in Fig. 2d is well reproduced for the different scenarios.

Inter-variable dependence is measured with Kendall's rank correlation coefficient [12], see Fig. 3. This correlation analysis is used to assess the strength of the relationship between two variables. The aim is to evaluate whether scenarios preserve the relationships observed between measured variables, see Fig. 3a. For example, in the observed series, the rank correlation between the wind speed (WS) and the air temperature (AirT) is about 0.17. Using the weather generator and multivariate bias correction, the strength of the relationship between these two variables is well reproduced. However, this inter-dependence is less well reproduced with the univariate bias correction and completely fails using the reanalyses (Kendall's τ is about 0.05). A negative non linear correlation is also observed between the relative humidity (Rh) and global radiation (GR) using the different scenarios.



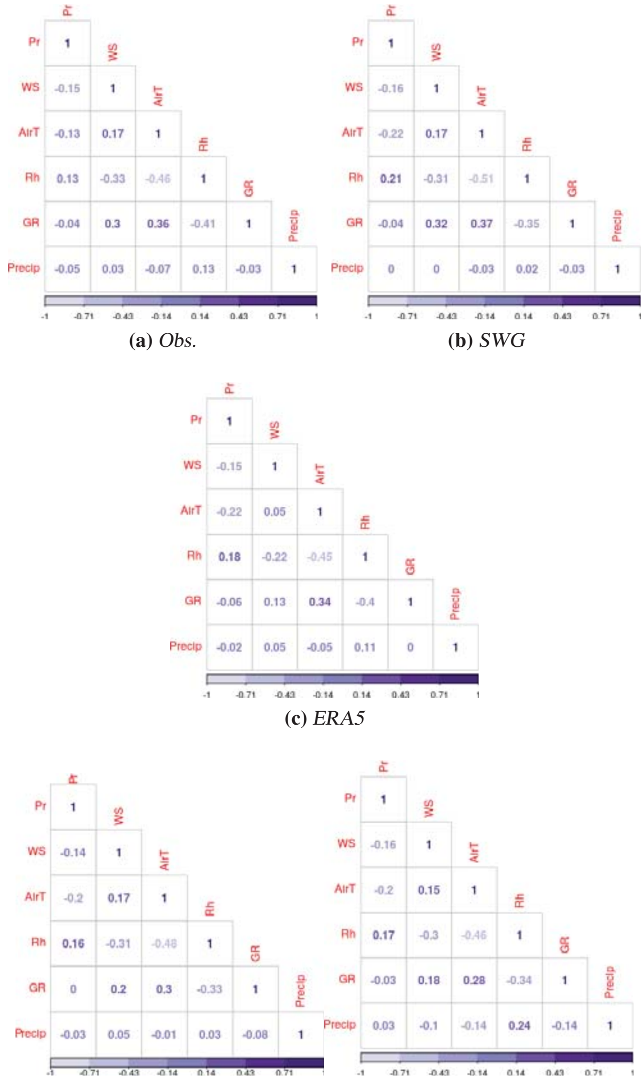
(a) *Relative humidity* **(b)** *Precipitation*

(c) *Wind speed* **(d)** *Global radiation*

Fig. 2. Cross-validation results : quantile-quantile plots at one of the gauged station in the Kairouan plain.

On the other hand, the scenarios are used to constrain the dual source energy balance model and are compared in terms of the simulation of the evapotranspiration and water stress index, see the quantile-quantile plots in Fig. 4a and 4b respectively. Results show that, simulations with the energy balance model using different hydrometeorological scenarios are comparable. However, for the stress index in Fig. 4b, both low and high extreme are deteriorated especially the low extreme values when using bias correction simulations and reanalyses.

A further comparison between simulations constrained by the different scenarios focuses on the water stress index. The exceedance probability, i.e. the probability that the index exceeds a given threshold, is considered. In Fig. 5, threshold values are represented on the x-axis, ranging from -0.5 to 1. The exceedance probabilities are on the y-axis, in black, as estimated by forcing the SPARSE model by the observations, with a 95% of confidence band in gray. When the model is forced by ERA5 and by the scenarios from the bias correction methods, the estimated exceedance probability falls outside the confidence interval between -0.5 and 0.5. Using scenarios from the generator, the estimated exceedance probability is almost always within the confidence interval.



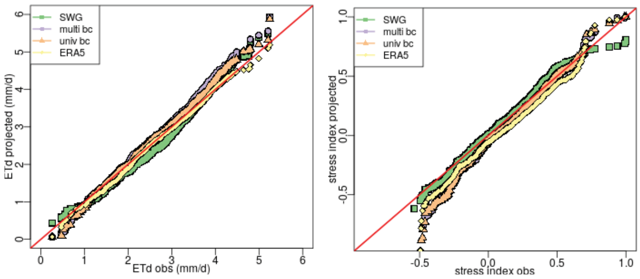
(a) *Obs.* **(b)** *SWG*

(c) *ERA5*

(d) *Multi. BC* **(e)** *Univ. BC*

Fig. 3. Cross-validation results : inter-variable dependencies measured with Kendall's rank correlation coefficient at one of the gauged station in the Kairouan plain.

The main contributions of this work are (1) the adaptation of a multi-variable stochastic weather generator to the sub-daily resolution, (2) its application in a semi-arid climate, (3) a comparison with two state-of-the-art bias correction methods and (4) the consideration of a wide range of performance criteria pertaining to the reproduction of the observations themselves and to the use of the scenarios to constrain the energy balance model.



(a) *Daily evapotranspiration* **(b)** *Water stress index*

Fig. 4. Cross-validation results : quantile-quantile plots of the simulation from the dual source enery balance model when constrained by the various scenarios.

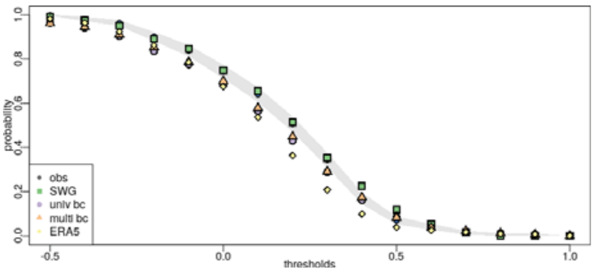


Fig. 5. Cross-validation results : probability that the stress index exceeds a given threshold when simulated by the various scenarios.

5. REFERENCES

[1] J. Sheffield and E. F. Wood, *Drought: past problems and future scenarios*, Routledge, 2012.

[2] H. G. Jones, R. Serraj, B. R. Loveys, L. Xiong, A. Wheaton, and A. H. Price, "Thermal infrared imaging of crop canopies for the remote diagnosis and quantification of plant responses to water stress in the field," *Functional Plant Biology*, vol. 36, no. 11, pp. 978–989, 2009.

[3] G. Boulet, B. Mougenot, J.-P. Lhomme, P. Fanise, Z. Lili-Chabaane, A. Olioso, M. Bahir, V. Rivalland, L. Jarlan, and O. Merlin, "The sparse model for the prediction of water stress and evapotranspiration components from thermal infra-red data and its evaluation over irrigated and rainfed wheat," *Hydrology and Earth System Sciences Discussions*, , no. 19, pp. 4653–4672, 2015.

[4] R. Chandler, "A multisite, multivariate daily weather generator based on Generalized Linear Models," *User guide : R package*, 2015.

[5] J. Friedman, T. Hastie, and R. Tibshirani, "Regularization paths for generalized linear models via coordinate descent," *Journal of Statistical Software*, vol. 33, pp. 1–22, 2010.

[6] H. Akaike, "Information theory and an extension of the maximum likelihood principle," in *Proc. 2nd Int. Symp. Inference Theory*, Petrov B.N. and Csàki F., Eds., 1973, pp. 267–281.

[7] G. Schwarz, "Estimating the dimension of a model," *Ann. Statist.*, vol. 6, no. 2, pp. 461–464, 03 1978.

[8] D. P. Dee, S. M. Uppala, A. J. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M. A. Balmaseda, G. Balsamo, P. Bauer, P. Bechtold, A. C. M. Beljaars, L. van de Berg, J. Bidlot, N. Bormann, C. Delsol, R. Dragani, M. Fuentes, A. J. Geer, L. Haimberger, S. B. Healy, H. Hersbach, E. V. Hólm, L. Isaksen, P. Kållberg, M. Köhler, M. Matricardi, A. P. McNally, B. M. Monge-Sanz, J. J. Morcrette, B. K. Park, C. Peubey, P. de Rosnay, C. Tavolato, J. N. Thépaut, and F. Vitart, "The ERA-Interim reanalysis: Configuration and performance of the data assimilation system," *Quarterly Journal of the royal meteorological society*, vol. 137, no. 656, pp. 553–597, 2011.

[9] J. Hooker, G. Duveiller, and A. Cescatti, "A global dataset of air temperature derived from satellite remote sensing and weather stations," *Scientific data*, vol. 5, pp. 180246, 2018.

[10] P.-A. Michelangeli, M. Vrac, and H. Loukos, "Probabilistic downscaling approaches: Application to wind cumulative distribution functions," *Geophysical Research Letters*, vol. 36, no. 11, 2009.

[11] A. J. Cannon, "Multivariate quantile mapping bias correction: an n-dimensional probability density function transform for climate model simulations of multiple variables," *Climate dynamics*, vol. 50, no. 1-2, pp. 31–49, 2018.

[12] H. Joe, *Multivariate models and multivariate dependence concepts*, CRC Press, 1997.