RMAWGEN: A software project for a daily Multi-Site Weather Generator with R

Emanuele Cordano, Emanuele Eccel

Department of Sustainable Agro-Ecosystems and Bioresources, Research and Innovation Center
Fondazione Edmund Mach San Michele all'Adige (TN), Italy

Abstract

Among the pre-existing approaches in literature, the R package **RMAWGEN** aims to generate future daily weather conditions by using the theory of vector auto-regressive models (VAR). The VAR model is used here because it is able to maintain the temporal and spatial correlations among the variables. In particular, historical time series of daily maximum and minimum temperature and precipitation are used to calibrate the parameters of a VAR model, which is stored as an object whose class inherits the `varest` S3 class defined in the package **vars** (Pfaff, 2008). Therefore the VAR model, coupled with monthly mean weather variables downscaled by Global Climate Model predictions, can generate several stochastic daily scenarios.

What is a VAR?

A set of K random variables can be described by a Vector Auto-Regressive Model (VAR(K,p)) as follows:

$$\mathrm{x_t} = \mathrm{A_1} \cdot \mathrm{x_{t-1}} + ... + \mathrm{A_p} \cdot \mathrm{x_{t-p}} + \mathrm{C} \cdot \mathrm{d_t} + \mathrm{u_t} \quad (1)$$

where $\mathrm{x_t}$ is a K-dimensional vector representing the set of weather variables generated at day t by the model, called "endogenous" variables, $\mathrm{d_t}$ is a set of known K-dimensional processes, whose components are called "exogenous" variables, $\mathrm{A_i}$ is a coefficient matrix $K \times K$ for $i = 1, ..., p$, C is a coefficient matrix for the exogenous variable and $\mathrm{u_t}$ is the VAR residual, i.e. K-dimensional stochastic process, $\mathrm{x_t}$ and $\mathrm{u_t}$ are usually normalized to have a null mean. u_t is a *Standard White Noise* (Luetkepohl, 2007,def. 3.1), i.e. a continuous random process with zero mean and $\mathrm{u_t}, \mathrm{u_s}$ independent for each $t \neq s$, consequently it has a time-invariant nonsingular covariance matrix. In **RMAWGEN**, the VAR model is saved as a `varest2` S4 object which contains `varest` S3 object.

The VAR models work correctly if the variable $\mathrm{x_t}$ is normally distributed. This requires a normalization procedure of the meteorological variable, which can be resumed as follows:

$$\mathrm{x_t} = \mathrm{G_m}(\mathrm{z_t}) \qquad \mathrm{z_t} = \mathrm{G_m}^{-1}(\mathrm{x_t}) \quad (2)$$

where $\mathrm{z_t}$ is the meteorological variable time-series and $\mathrm{G_m}$ is a suitably defined function (m is a month indicator) so that $\mathrm{x_t}$ is multi-normally distributed and can vary with time, month and season. In the univariate case, Gaussianization is done through the quantile estimation. In the multivariate case, the Gaussianization is done with an iterative method based on Univariate Gaussianization and Principal Component Analysis (PCA) rotation (GPCA) (Laparra , 2011). The final results tend to be normally distributed with a reduction of the Kullback-Leibler divergence at each iteration. During VAR model creation, **RMAWGEN** does an univariate Gaussianization for each month (see Fig. 1) and a Multivariate GPCA only if GPCA option is enabled. In case of GPCA the whole model is saved as an object of the sub-class `GPCAvarest2`.

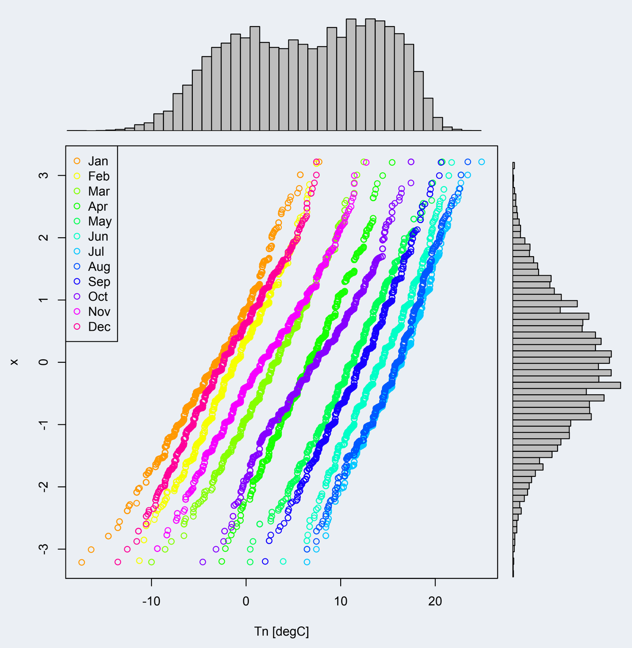

Figure 1: Monthly relationship between minimum daily temperature and X with the respective frequency histogram.

Discussion and Future developments

- **RMAWGEN** fairly reproduces generated time series of daily temperature statistically coherent with the observed ones (even if some problems should be investigated farther).
- **RMAWGEN** is quite flexible: thanks to Gaussianization (see Fig. 1), it can work with different probability distributions of weather variables.
- **RMAWGEN** contains also a stochastic daily precipitation generator (which will be improved and upgraded soon) based on VAR models. The use of VAR, as `varest` or `varest2` objects, allows the introduction of exogenous variables: this enables the generation of daily weather variables simultaneously at several sites.
- **RMAWGEN** was conceived to generate daily time series of temperature and precipitation for ecological and phenological modeling and is open to customized implementations.

Application: Calibration for Daily Minimum and Maximum Temperature (Trentino Dataset)

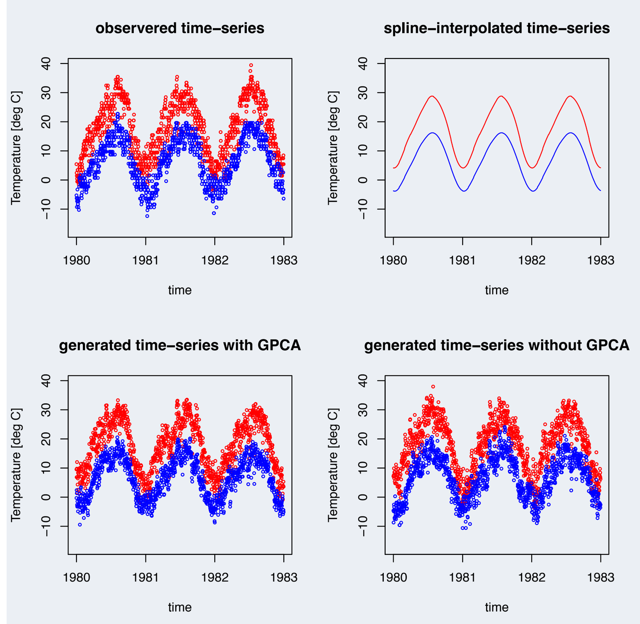

Figure 2: Time Series at T0090

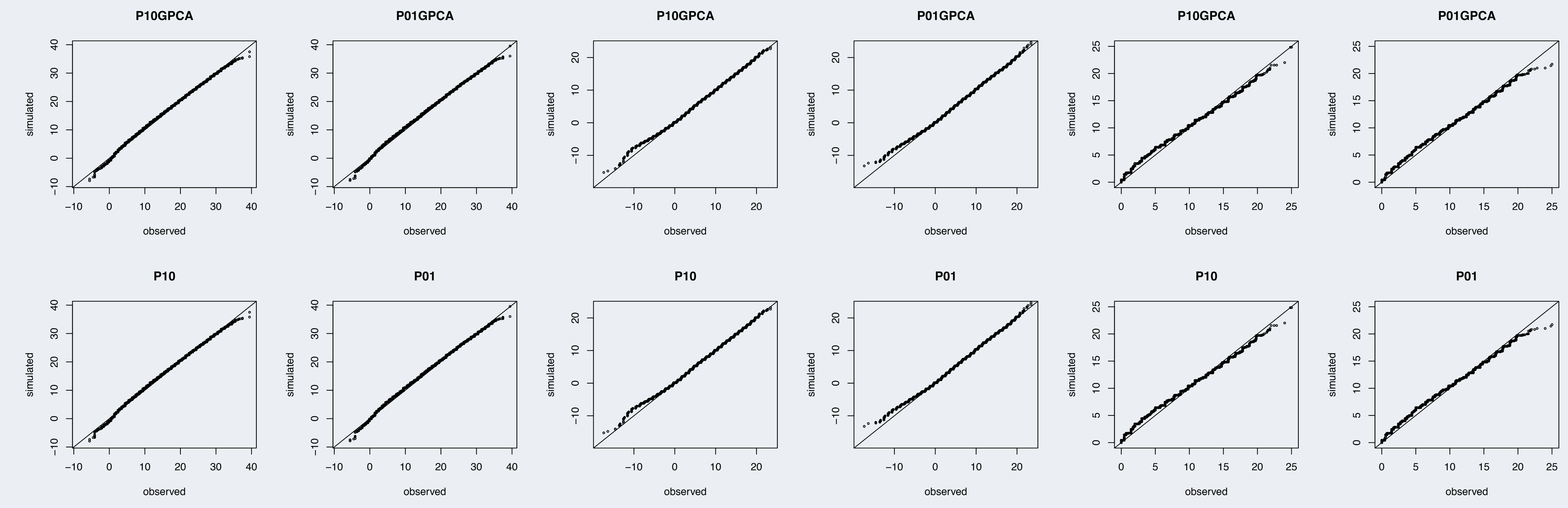

Figure 3: Q-Q (quantile-quantile) plot at T0090 Station: Maximum (left) and Minimum (middle) Temperature and Daily Thermal Range (right)

The observed and simulated time-series (provided by `trentino` dataset, Eccel =[, 2012) (see Fig. 2) are reported under different values of p (auto-regression order) and with or without GPCA (see Table 1). Q-Q plots show a satisfactory fitting between observations and simulations, as expected (see Fig. 3). However, the tests of acceptance of the VAR model, carried out on residuals, are successful only in case of GPCA. Unfortunately, heteroskedasticity of the residuals cannot be verified by the tests Is it significant for climatology applications? **Further validation of the model was successfully done by analyzing some Climate Indices and here presented by Di Piazza A et al**,*Use of a Weather Generator for analysis of projections of future daily temperature and its validation with climate change indices*, **Z98-EGU2012-5404**.

Table 1: Test on *whiteness* of residuals

	$p = 1$	p =10	p =1 (with GPCA)	p =10 (best AIC)(with GPCA)
Acronym	P01	P10	P01GPCA	P10GPCA
Normality Test	NO	NO	YES	YES
Seriality Test	NO	NO	NO	YES
Heteroskedasticity Test	NO	NO	NO	NO

Some References and Acknowledgments

Eccel E, Cau P, Ranzi R (2012). "Data reconstruction and homogenization for reducing uncertainties in high-resolution climate analysis in Alpine regions." *Theoretical and Applied Climatology*, pp. 1–14. ISSN 0177-798X. 10.1007/s00704-012-0624-z, URL `http://dx.doi.org/10.1007/s00704-012-0624-z`.

Laparra V, Camps-Valls G, Malo J (2011). "Iterative Gaussianization: from ICA to Random Rotations." *IEEE Transactions on Neural Networks*.

Luetkepohl H (2007). *New Introduction to Multiple Time Series Analysis*. 2nd printing edition. Springer-Verlag, Berlin Hedelberg, Germany.

Pfaff B (2008). "VAR, SVAR and SVEC Models: Implementation Within R Package vars." *Journal of Statistical Software*, **27**(4). URL `http://www.jstatsoft.org/v27/i04/`.

Work funded by projects ACE-SAP and ENVIROCHANGE (Provincia Autonoma di Trento) The authors thank Dr. Annalisa Di Piazza and the R Development Core Team. useR!

RMAWGEN is available with GPL on `http://cran.r-project.org/web/packages/RMAWGEN/index.html`

emanuele.cordano@gmail.com,emanuele.eccel@fmach.it