European Geosciences Union
General Assembly 2013

Vienna, Austria,
7 - 12 April 2013

Session HS7.5/NP8.4:
Hydroclimatic stochastics

The Castalia stochastic generator and its applications to multivariate disaggregation of hydro-meteorological processes

Anna Venediki[1], Savvas Giannoulis[1], Christos Ioannou[1],
Lamprini Malatesta[1], Georgios Theodoropoulos[1], Georgios Tsekouras[1],
Yannis G. Dialynas[2], Simon Michael Papalexiou[1],
Andreas Efstratiadis[1] and Demetris Koutsoyiannis[1]

[1]Department of Water Resources and Environmental Engineering,
National Technical University of Athens, Greece

[2]School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, USA

1. Abstract

Castalia is a software system that performs multivariate stochastic simulation preserving essential marginal statistics, specifically mean value, standard deviation and skewness, as well as joint second order statistics, namely auto- and cross-correlation. Furthermore, Castalia reproduces long-term persistence. It follows a disaggregation approach, starting from the annual time scale and proceeding to finer scales such as monthly and daily. To assess the performance of the Castalia system we test it for several hydrometeorological processes such as rainfall, sunshine duration and wind speed. To this aim we retrieve time series of these processes from a large database of daily records and we estimate their statistical properties, including long-term persistence. We generate synthetic time series using the Castalia software and we examine its efficiency in reproducing the important statistical properties of the observed data.

Acknowledgment: This research is conducted within the frame of the undergraduate course "Stochastic Methods in Water Resources" of the National Technical University of Athens.

2. Data Set - Preprocessing

□ We retrieve daily time series of wind speed, sunshine duration and precipitation from “KNMI Climate Explorer” (http://climexp.knmi.nl/), “European Climate Assessment & Data (ECA&D)” (http://eca.knmi.nl/) and “Deutscher Wetterdienst” (http://www.dwd.de).

□ Time series were selected based on the criteria below:

- Length of time series ≥ 70 years (for relatively reliable estimation of the Hurst coefficient);
- Available wind speed metadata (measurement height above ground).

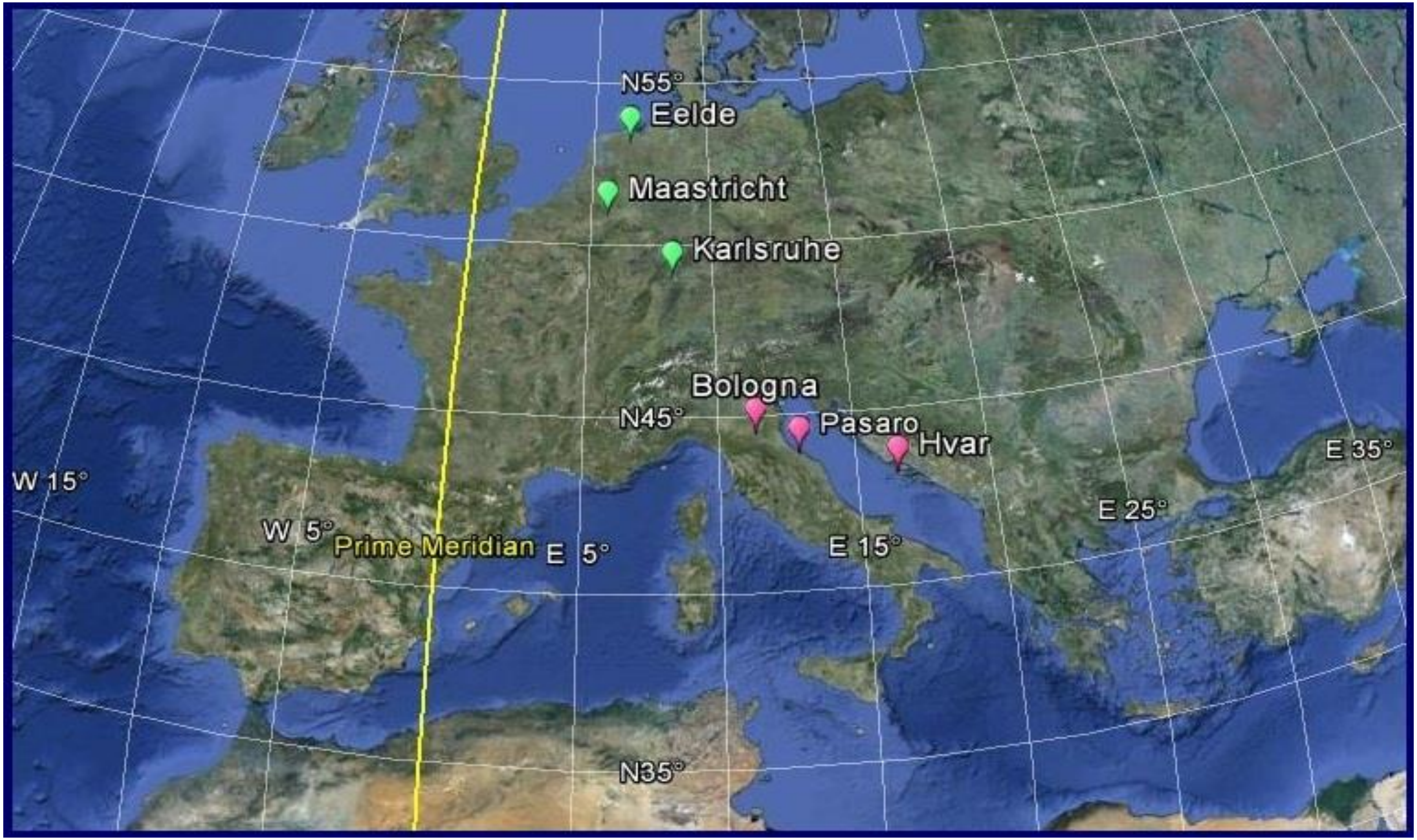
□ We examined three time series for each of the above variables.

□ Wind speed data with changes in the measurement height were adjusted, in order to refer to the most recent observation height.

□ The sunshine duration (*S*) is a variable bounded from both below and above with finite probability of zero value. Considering that the Gamma distribution is used in several software systems (including Castalia) for the generation of synthetic time series, it is useful to associate the distribution of the relative sunshine duration (*X)* with the Gamma distribution. That is achieved by the logarithmic transformation $Y = -\ln(1 - X)$ so that the domain of the transformed variable is $[0, +\infty)$ and the value $X = 0$ corresponds to $Y = 0$.

Station	Country	Latitude	Longitude	Variables	Time series period
Eelde	Netherlands	53.12N	6.58E	Sunshine duration	1906-2011
				Wind speed	1906-2011
Karlsruhe	Germany	49.04N	8.36E	Sunshine duration	1936-2011
				Wind speed	1945-2011
Maastricht	Netherlands	50.84N	5.68E	Sunshine duration	1906-2011
				Wind speed	1927-2011
Bologna	Italy	44.50N	11.35E	Precipitation	1813-2007
Hvar	Croatia	43.17N	16.45E	Precipitation	1857-2008
Pesaro	Italy	43.91N	12.90E	Precipitation	1871-2008

3. Locations of stations

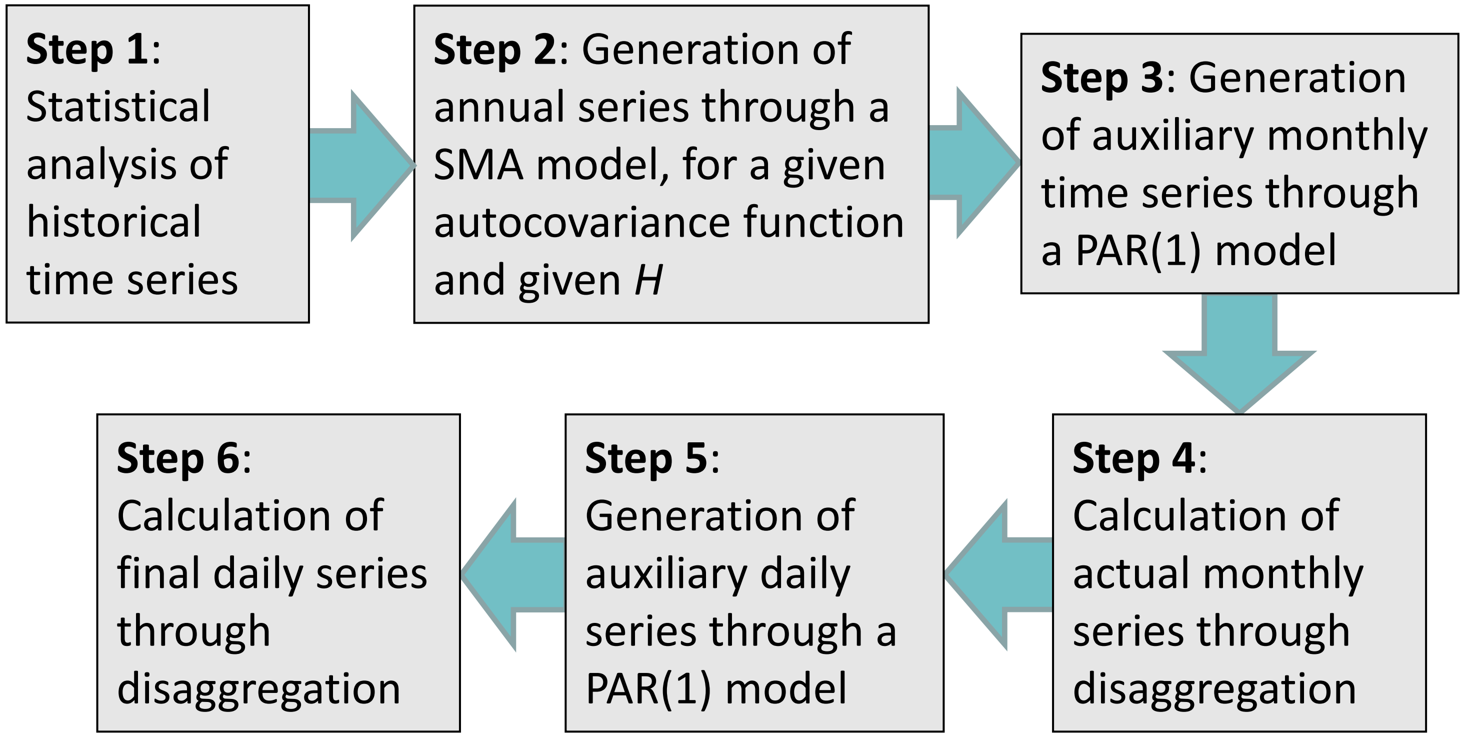


🟢 Sunshine Duration & Wind

🔴 Precipitation

4. The software system Castalia (1)

- The model performs multivariate stochastic simulation on annual, monthly and daily scales following a disaggregation approach (Koutsoyiannis and Efstratiadis, 2001; Dialynas, 2011; Dialynas et al., 2011).
- The system preserves :
  - At all scales the marginal statistics (mean value, standard deviation, skewness) and the joint second order statistics (auto- and cross-correlations);
  - At multi-year scales the long term persistence (LTP);
  - At sub-annual scales the periodicity;
  - At daily scale the intermittency.
- LTP is reproduced through a symmetric moving average (SMA) scheme for a generalized autocovariance function with user-specified parameters (Koutsoyiannis, 2000), allowing to represent from ARMA-type (H = 0.50) to highly persistent processes (H > 0.50).
- Auxiliary series are provided by a multivariate PAR(1) scheme, both for the monthly and daily scales (Koutsoyiannis, 1999).
- A disaggregation procedure is employed to ensure statistical consistency between the three temporal scales; first the monthly series are adjusted to the known annual ones, and next the daily time series are adjusted to the disaggregated monthly ones (Koutsoyiannis, 2001).



Step 1: Statistical analysis of historical time series

Step 2: Generation of annual series through a SMA model, for a given autocovariance function and given H

Step 3: Generation of auxiliary monthly time series through a PAR(1) model

Step 4: Calculation of actual monthly series through disaggregation

Step 5: Generation of auxiliary daily series through a PAR(1) model

Step 6: Calculation of final daily series through disaggregation

Flowchart of computational procedures of Castalia

5.The software system Castalia (2)

- ❑ LTP is modelled using a parametrically defined generalized autocovariance function:

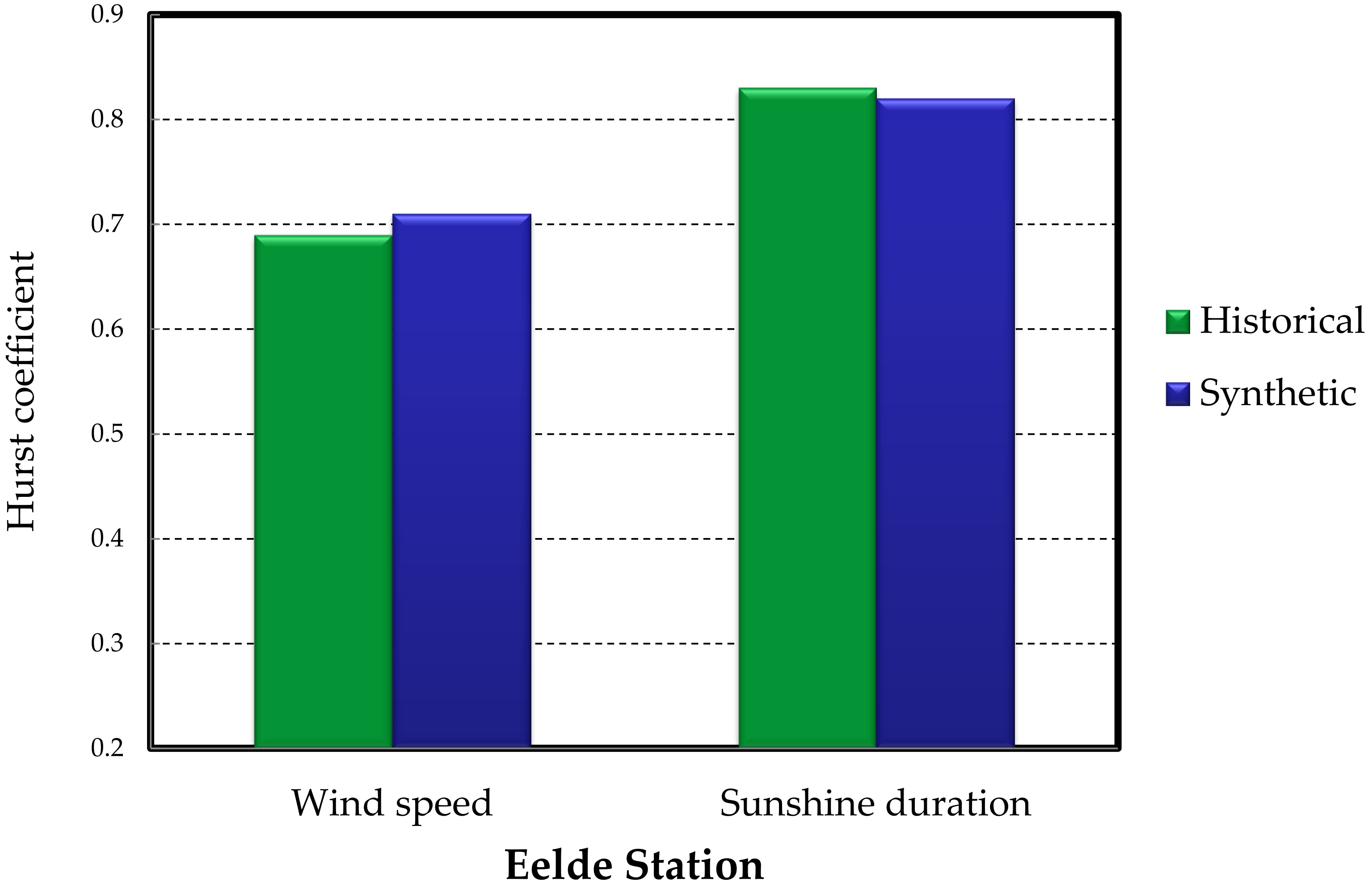
$$\gamma_j = \gamma_0[1 + \kappa\, \beta j]^{-1/\beta}$$

where γ_j is the autocovariance for lag j and κ and β are parameters.

- ❑ This theoretical autocovariance function is fitted to the sample autocovariance of each annual variable. The parameters of the Symmetric Moving Average (SMA) model are estimated using this theoretical autocovariance function and the annual marginal and joint second order statistics (Koutsoyiannis 2000).
- ❑ The Castalia program can also handle possible intermittent behaviour of variables on the daily time scale using the following parameters chosen by the user:
  - k_1 and k_2, which within a Markov chain model express the probability of a zero value occurring in the current time step if there is a zero (k_1) or a non-zero value (k_2) value in the previous time step;
  - k_3, which expresses the probability of the values of all variables being zero in the current time step, if at least one of them is zero;
  - π_0 and l_0 , which are parameters of an empirical round-off rule, according to which a percentage π_0 of the generated values below a threshold l_0 are converted to zero.
- ❑ All three procedures generate zero values and they all contribute to the final frequency of zero values.

6. Sunshine Duration & Wind Speed (1)

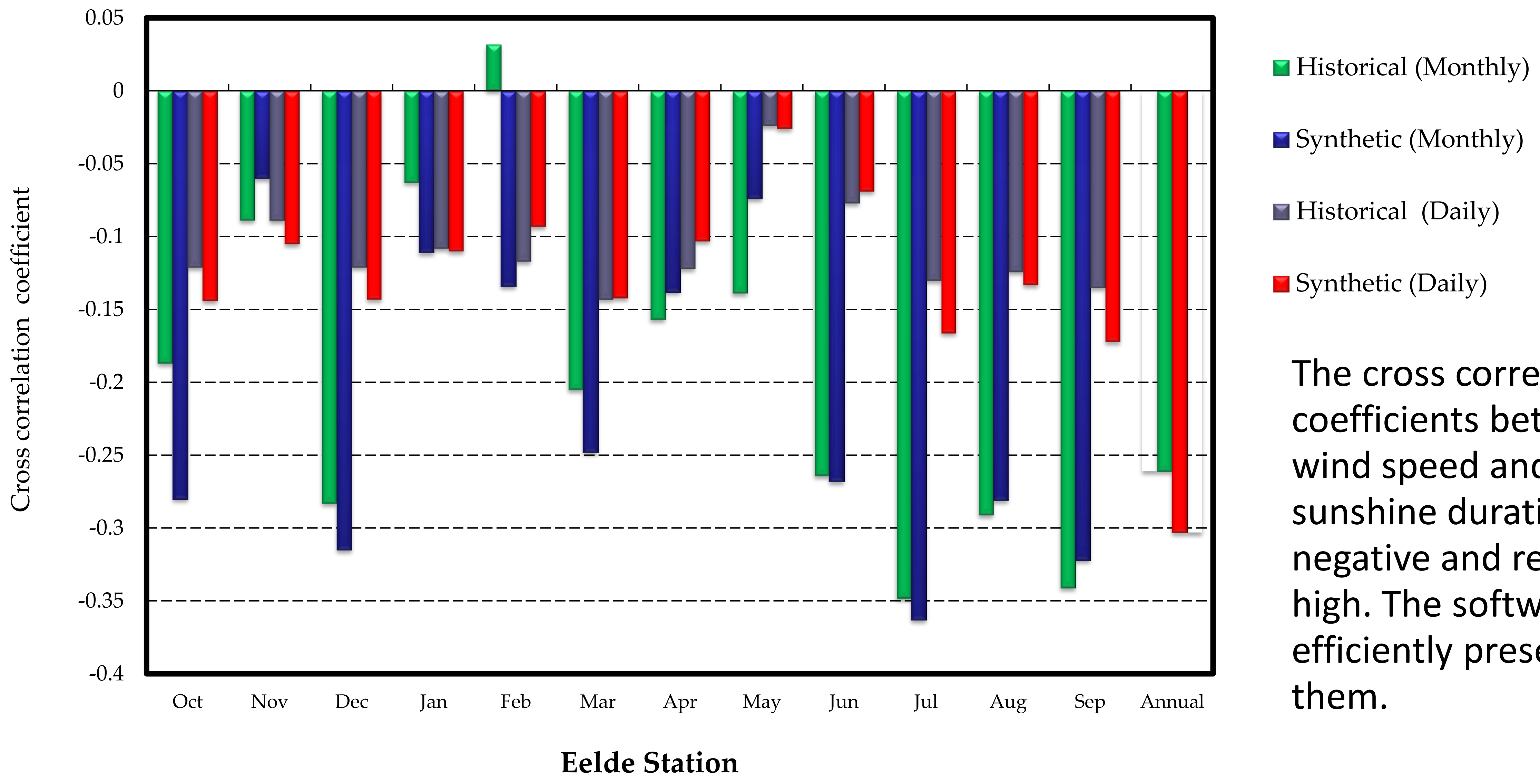
Synthetic time series with a length of 1000 years are generated in each of the 3 locations having both wind speed and sunshine duration records. The proposed logarithmic transformation of the sunshine duration data is used and, after the generation of the synthetic time series of the transformed variables *Y*, the synthetic daily time series of the sunshine duration *X* are calculated by applying the reverse transformation on the *Y* daily time series. To test the reliability of the software system, historical and synthetic statistics are intercompared.



The chart on the left depicts the histogram of the Hurst coefficient, as it is estimated from historical and synthetic time series. Apparently, the Hurst coefficient is preserved. Notice that both values exceed 0,65, which indicates long-term persistence.

7. Sunshine Duration & Wind Speed (2)

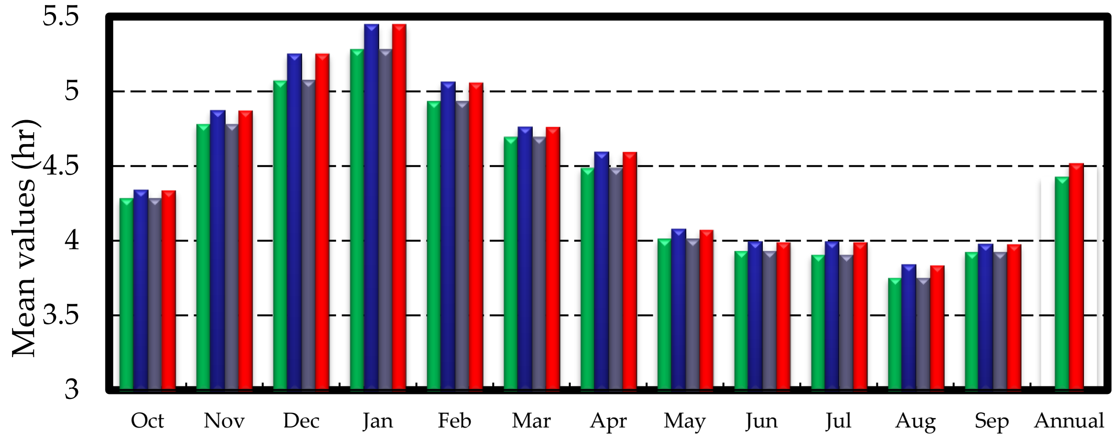
	January	February	March	April	May	June	July	August	September	October	November	December	Year	Station
Monthly	-0,294	-0,005	-0,204	-0,06	-0,014	-0,246	-0,379	-0,016	-0,322	-0,462	-0,305	-0,501	-0,008	Karlsruhe
Daily	-0,185	-0,138	-0,169	-0,17	-0,201	-0,285	-0,235	-0,166	-0,191	-0,268	-0,24	-0,231		
Monthly	-0,287	-0,296	-0,228	-0,192	-0,194	-0,219	-0,132	0,058	-0,095	-0,051	-0,129	-0,309	-0,077	Maastricht
Daily	-0,161	-0,152	-0,144	-0,16	-0,144	-0,193	-0,165	-0,056	-0,087	-0,146	-0,143	-0,18		
Monthly	-0,187	-0,089	-0,283	-0,063	0,032	-0,205	-0,157	-0,139	-0,264	-0,348	-0,291	-0,341	-0,303	Eelde
Daily	-0,121	-0,089	-0,121	-0,108	-0,117	-0,143	-0,122	-0,024	-0,077	-0,13	-0,124	-0,135		



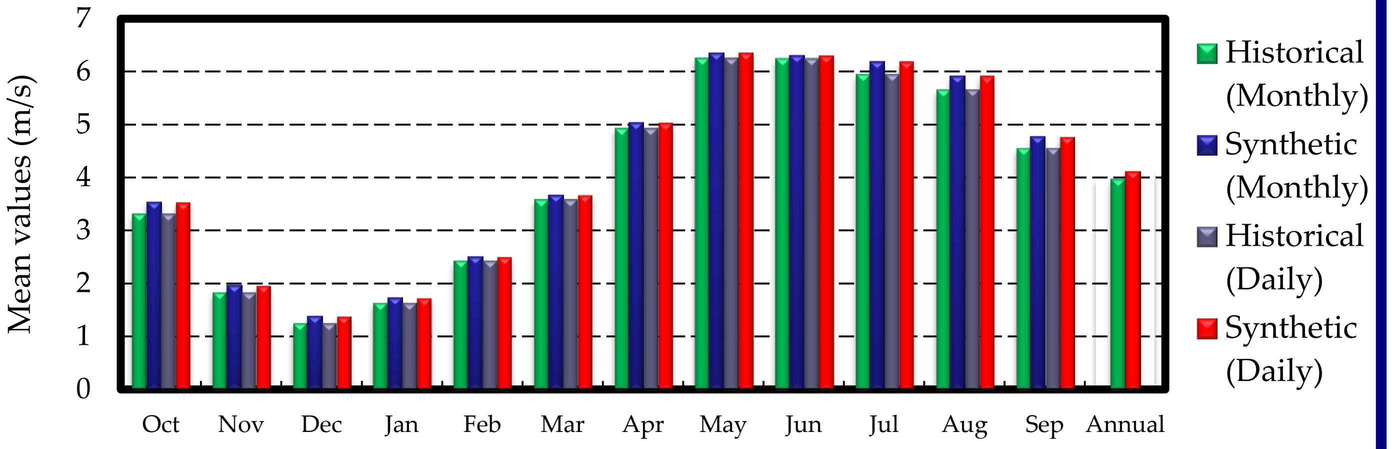
The cross correlation coefficients between wind speed and sunshine duration are negative and relatively high. The software efficiently preserves them.

8. Sunshine Duration & Wind Speed (3)

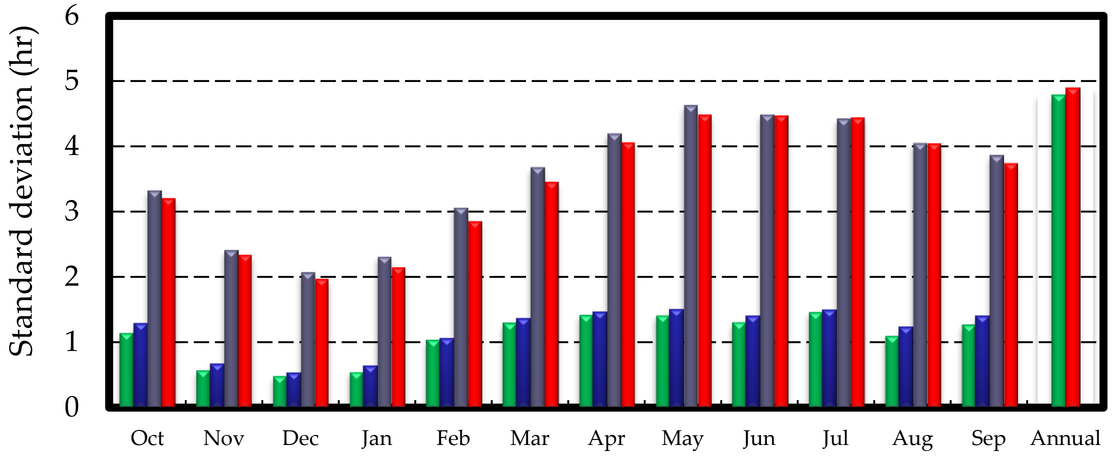
The historical marginal statistical characteristics such as mean value, standard deviation and skewness coefficient are preserved in the synthetic time series.



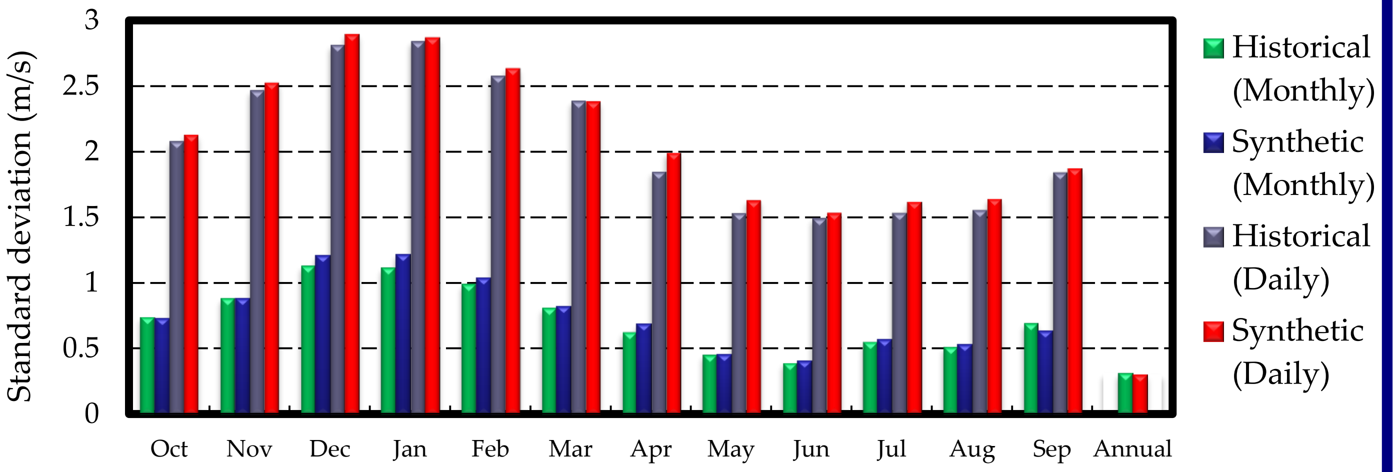
Maastricht Station



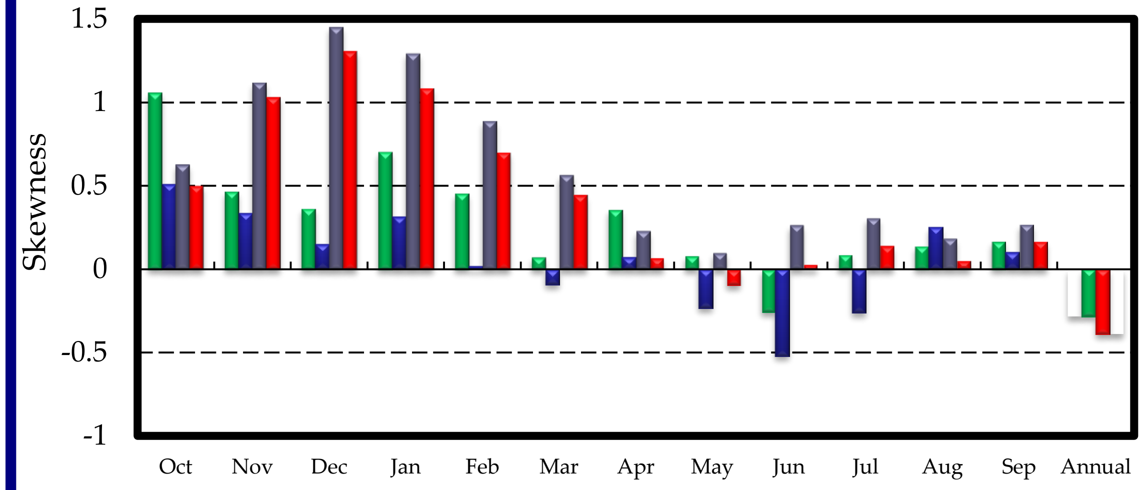
Maastricht Station



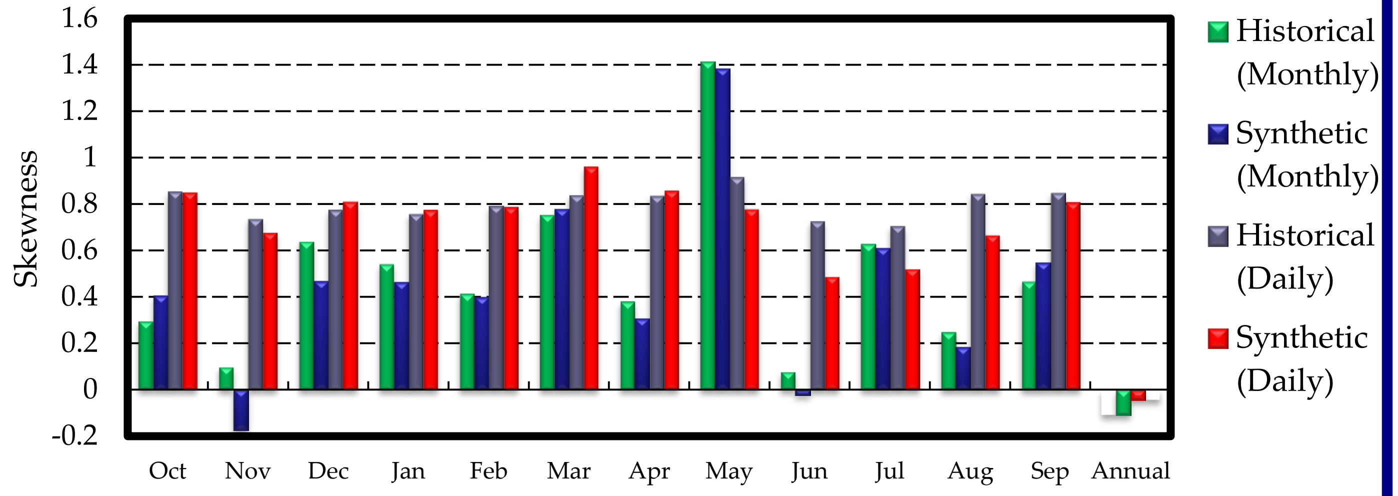
Karlsruhe Station



Karlsruhe Station

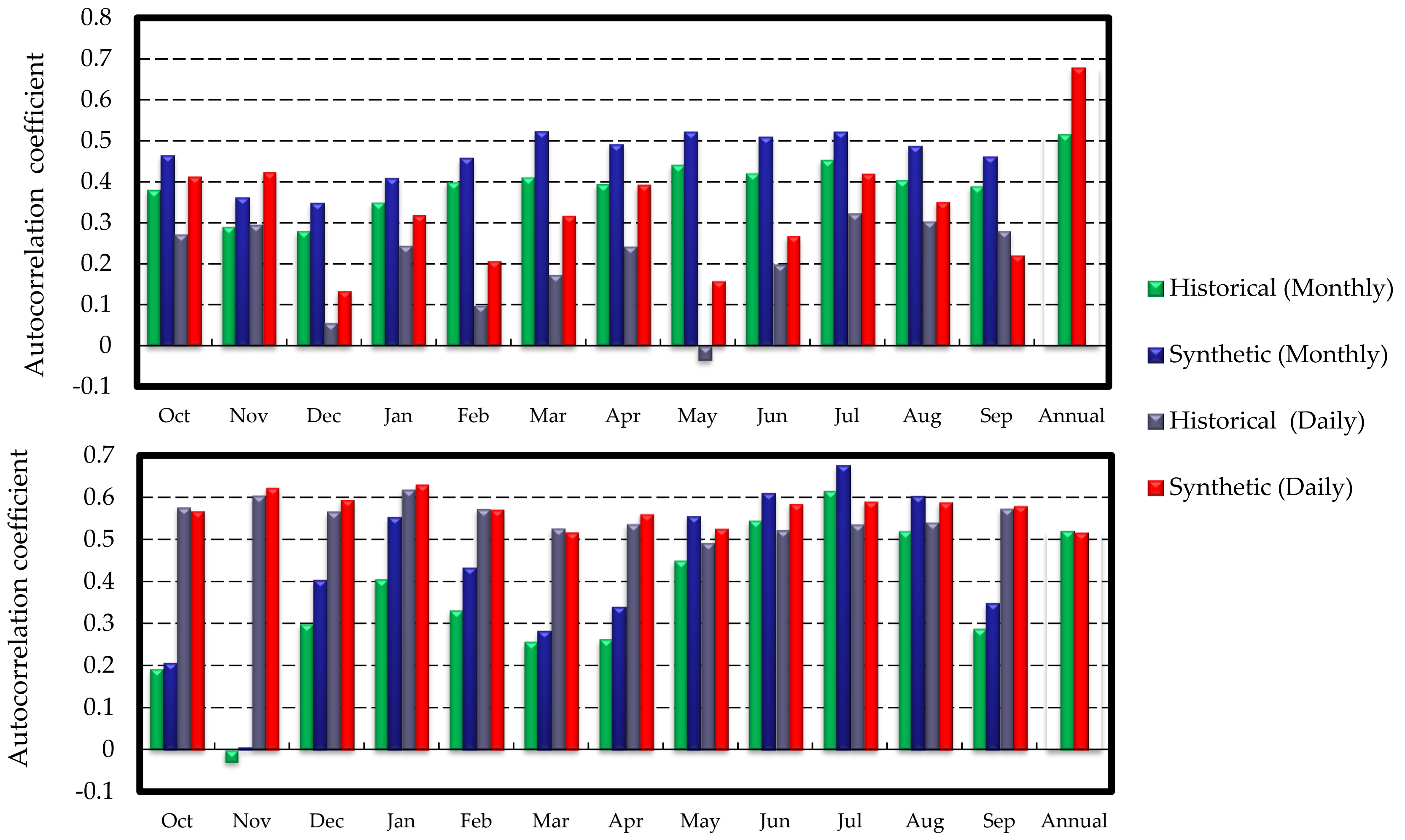


Eelde Station



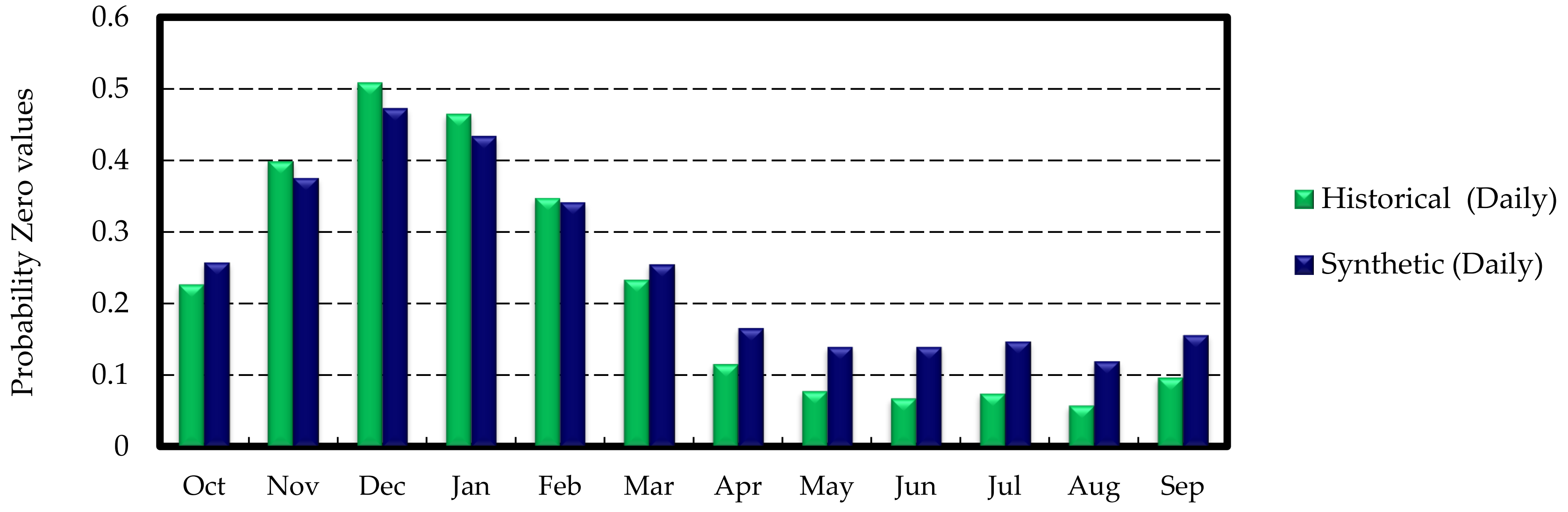
Eelde Station

9. Sunshine Duration & Wind Speed (4)



Maastricht Station

The autocorrelation coefficients of the synthetic time series produced by Castalia are similar to the ones of the historical time series. These results correspond to the transformed variable *Y*= –ln(1 – *X*).

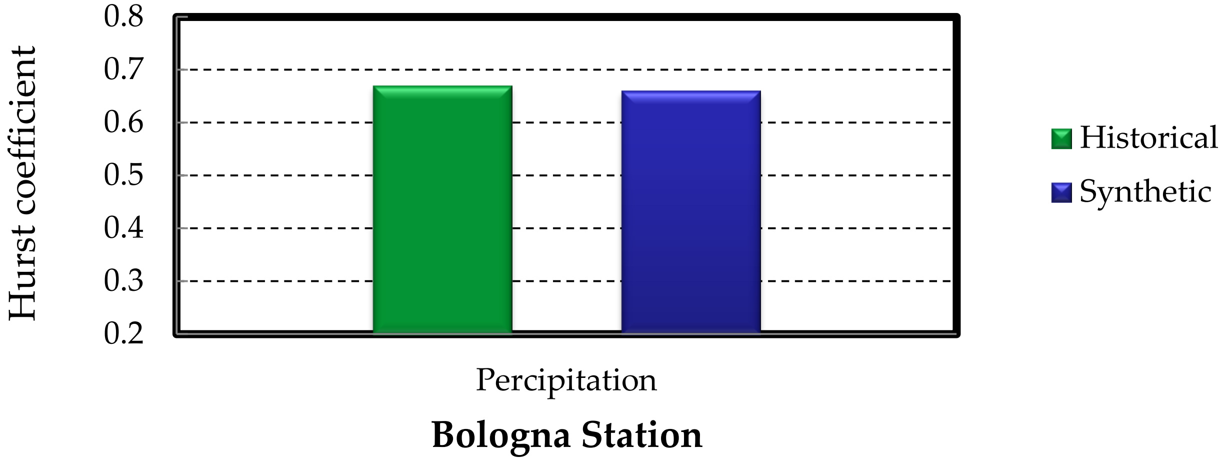


Eelde Station

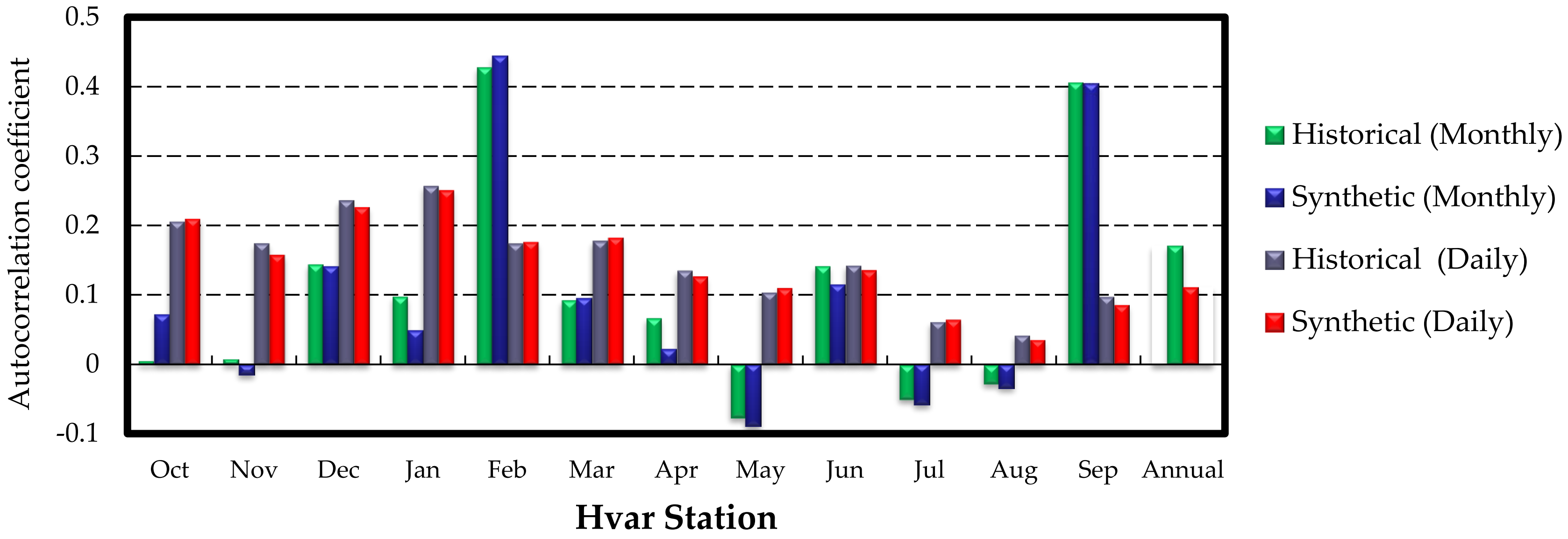
At the chart on the left the probability of no sunshine is presented; this probability is reproduced at a satisfactory level.

10. Precipitation (1)

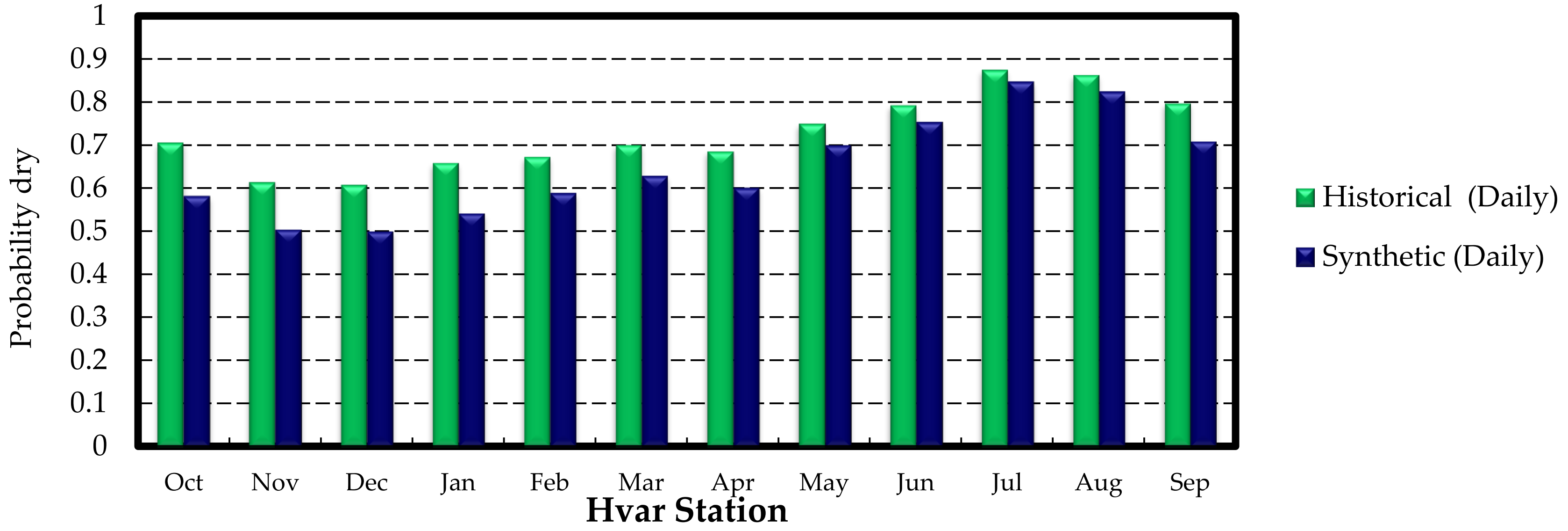
1000-year synthetic time series are also generated in each of the 3 locations with precipitation records. As shown in the histogram on the right, the Hurst coefficient of the historical time series is preserved in the synthetic time series produced by Castalia.



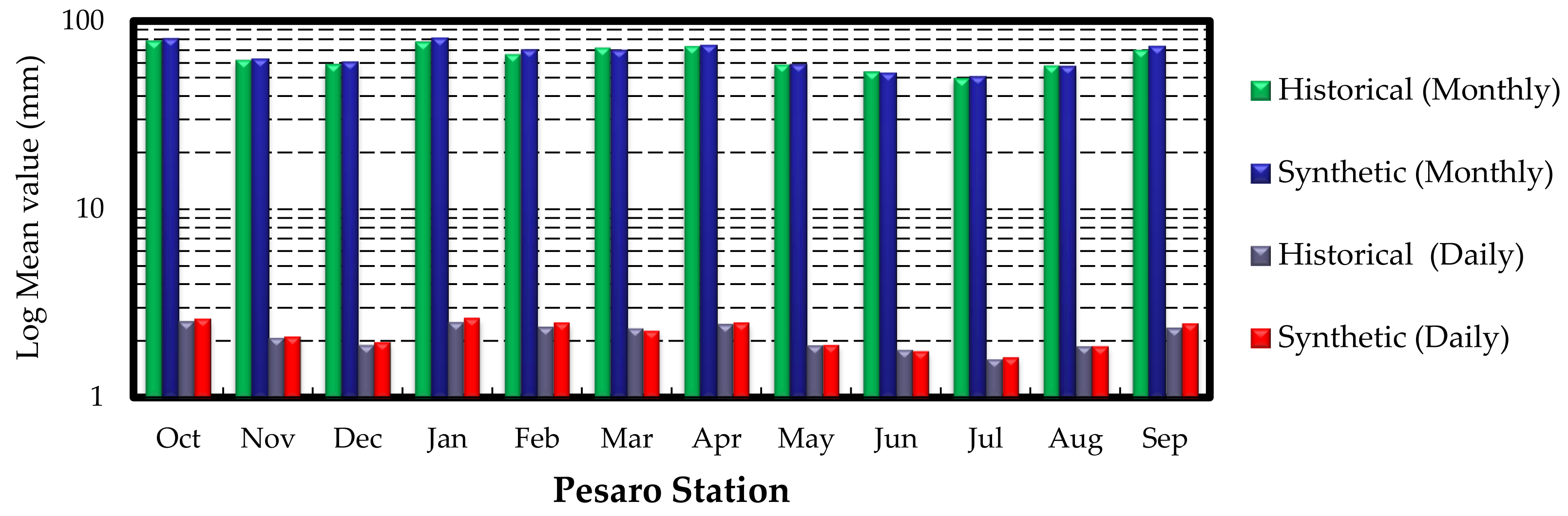
Also, the autocorrelation coefficient of the synthetic time series is very close to the one of the historical time series on monthly basis as well as on daily basis.

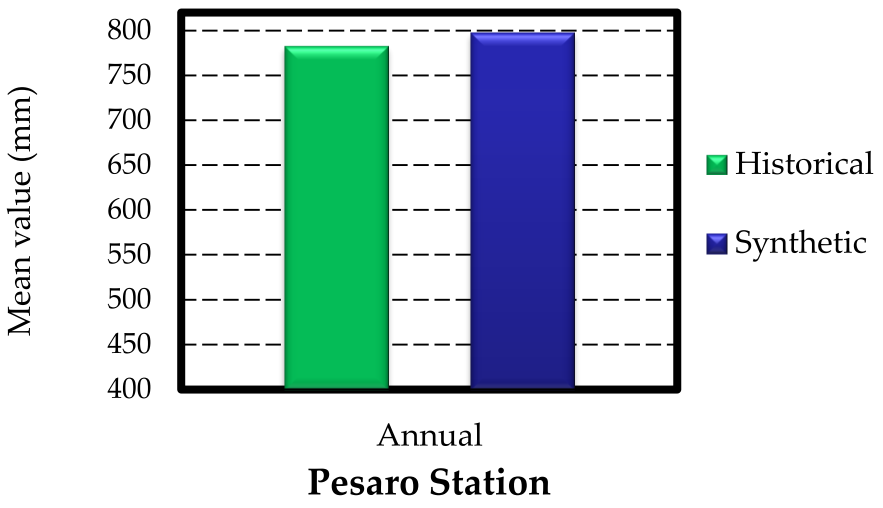


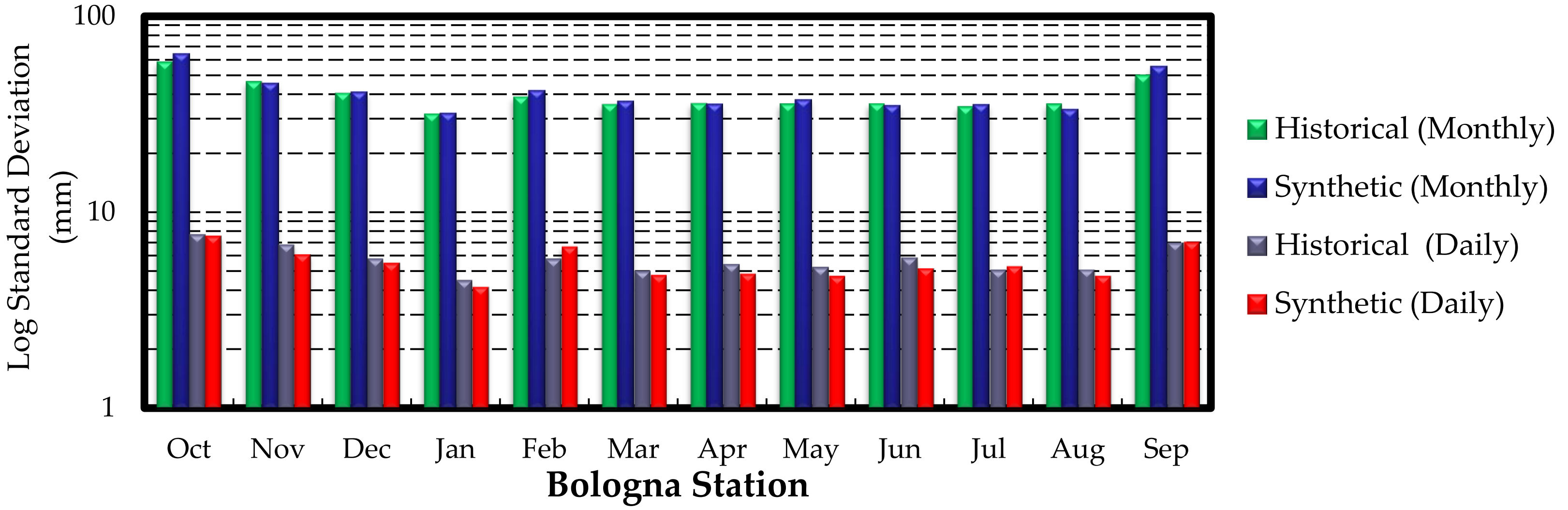
Furthermore, the probability dry of the synthetic time series is slightly underestimated, with respect to the historical one, but the overall performance is satisfactory.

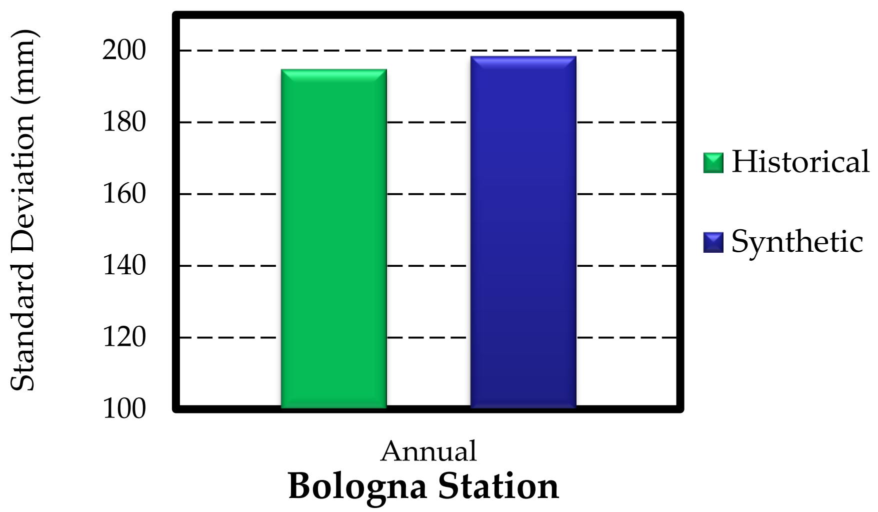


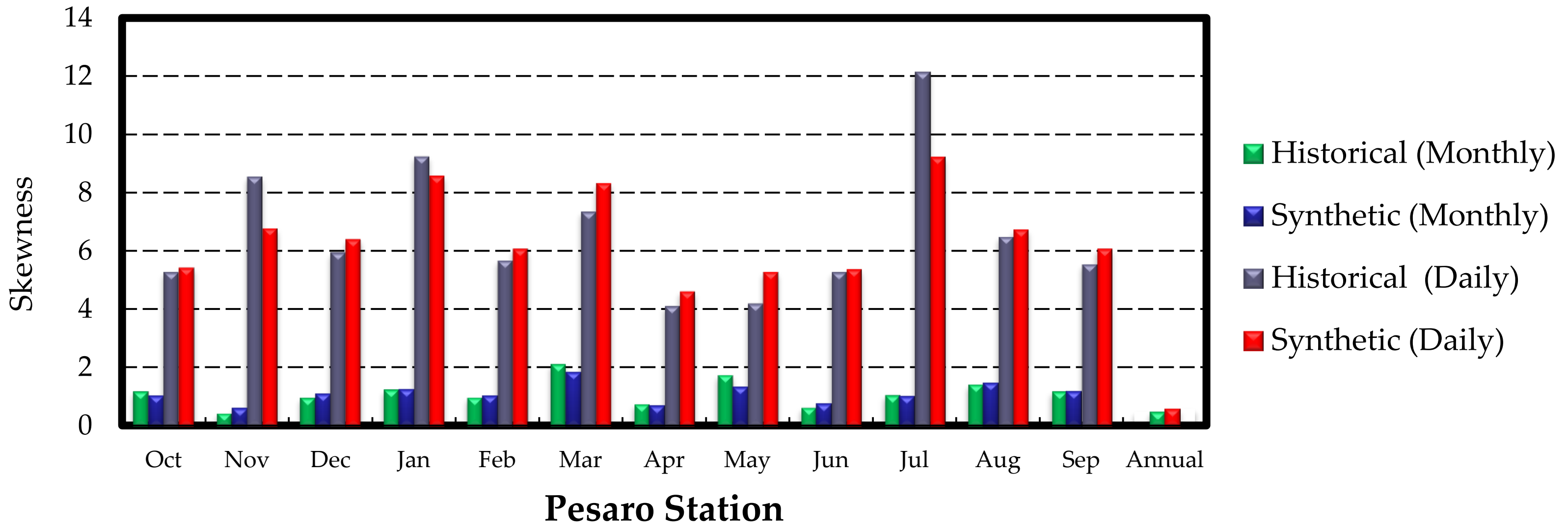
11. Precipitation (2)











Here, the preservation of marginal statistics is illustrated

12. Conclusions

- ❑ The Castalia software can be used for multivariate stochastic simulation of hydrometeorological processes such as sunshine duration, wind speed and precipitation on annual, monthly and daily scale,
- ❑ The software system preserves satisfactorily the marginal and the joint second order statistics.
- ❑ High values of the Hurst coefficient were observed for all processes. It is important that the Castalia software preserves the Hurst coefficient too.
- ❑ Considering that Castalia software was initially developed for the simulation of rainfall and runoff, it is confirmed that this program can further conduct stochastic simulations of a wide spectrum of hydrometeorological variables on these three time scales.
- ❑ There is room for improving Castalia software to explicitly incorporate the use of nonlinear transformations of variables, to avoid pre-processing and post-processing of the generated time series, as happened in the generation of sunshine duration.

References

- Dialynas, Y., A computer system for the multivariate stochastic disaggregation of monthly into daily hydrological time series, Diploma thesis, 337 pages, Department of Water Resources and Environmental Engineering – National Technical University of Athens, Athens, March 2011.
- Dialynas, Y., S. Kozanis, and D. Koutsoyiannis, A computer system for the stochastic disaggregation of monthly into daily hydrological time series as part of a three–level multivariate scheme, European Geosciences Union General Assembly 2011, Geophysical Research Abstracts, Vol. 13, Vienna, EGU2011 290, European Geosciences Union, 2011.
- Koutsoyiannis, D., and A. Efstratiadis, A stochastic hydrology framework for the management of multiple reservoir systems, Geophysical Research Abstracts, Vol. 3, European Geophysical Society, 2001.
- Koutsoyiannis, D., A generalized mathematical framework for stochastic simulation and forecast of hydrologic time series, Wat. Resour. Res., 36(6), 1519–1533, 2000.