Stochastic analysis and simulation of hydrometeorological processes associated with wind and solar energy

Georgios Tsekouras and Demetris Koutsoyiannis

Department of Water Resources and Environmental Engineering, Faculty of Civil Engineering, National Technical University of Athens, Heroon Polytechneiou 5, GR-157 80 Zographou, Greece(georgtsek@yahoo.gr)

Abstract

The current model for energy production, based on the intense use of fossil fuels, is both unsustainable and environmentally harmful and consequently, a shift is needed in the direction of integrating the renewable energy sources into the energy balance. However, these energy sources are unpredictable and uncontrollable as they strongly depend on time varying and uncertain hydrometeorological variables such as wind speed, sunshine duration and solar radiation. To study the design and management of renewable energy systems we investigate both the properties of marginal distributions and the dependence properties of these natural processes, including possible long-term persistence by estimating and analyzing the Hurst coefficient. To this aim we use time series of wind speed and sunshine duration retrieved from European databases of daily records. We also study a stochastic simulation framework for both wind and solar systems using the software system Castalia, which performs multivariate and multi-time-scale stochastic simulation, in order to conduct simultaneous generation of synthetic time series of wind speed and sunshine duration, on yearly, monthly and daily scale.

Keywords: wind speed, sunshine duration, long term-persistence, Hurst coefficient, marginal distributions, multivariate stochastic simulation

1. Introduction

The drawbacks of conventional energy sources including their negative environmental impacts emphasize the need to integrate renewable energy sources, like wind and solar, into the energy balance. However these energy sources, unlike fossil fuels on which the past model was based on, are unpredictable, at both short and long run, and uncontrollable as they are associated with relevant hydrometeorological processes which are characterized by high variability and uncertainty [1]. As a result, the estimation of renewable energy sources requires analyzing hydrometeorological data, especially wind speed, sunshine duration and solar radiation time series.

Such hydrometeorological time series are typically modelled as stationary stochastic processes. In this case their probability distribution does not change in time. It is thus necessary, in order to characterize uncertainty, to assume a theoretical distribution function that fits properly the data. Normal distribution is suitable for most hydrometeorological processes at an annual or higher time scale due to the Central Limit Theorem. However, variables of lower time scales, such as monthly or daily, are characterized by skewness and thus, non-normal distributions are more appropriate for their representation. In particular, the sunshine duration is represented as a random variable bounded from both below and above, and thus a theoretical distribution with these properties should be chosen for its representation.

Various studies have been performed in order to examine theoretical distribution functions fitting properly wind speed and relative sunshine duration data. The Weibull and Gamma distributions have been found to be appropriate to represent wind speed data while the Beta distribution has been assumed to be suitable for the relative sunshine duration. Carta *et al*. [2] and Zhou *et al*. [3] proposed that both the Weibull and the Gamma distribution are suitable for hourly wind speed representation in the Canary Islands and in North Dakota respectively. Garcia *et al*. [4] and Yilmaz and Celik [5] suggested that the Weibull distribution fits satisfactorily hourly wind speed data in Spain and Turkey. Darbandi *et al*. [6] noted that the Gamma distribution fits well the maximum annual wind speed data in Iran. Concerning the relative sunshine duration, Sulaiman *et al*. [7] and Bashahu and Nsabimana [8] examined the fit of the Beta distribution to daily data in Malaysia and Africa and concluded that it fitted satisfactorily the data in several cases. Chia and Hutchinson [9] examined the possibility of fitting the Beta distribution to daily relative cloud duration time series in Australia.

In addition to the marginal distributional properties, the fact that climate is not static but exhibits fluctuations at all time scales should also be taken into consideration for the effective exploitation of renewable energy sources. In a stochastic framework, these fluctuations are quantified through the dependence of the processes in time [10]. While in several studies these processes have been regarded as independent identically distributed random variables, particularly at coarse time scales such as monthly or annual, other studies have verified the presence of dependence, sometimes long-range dependence, which has been also termed as long-term persistence (LTP), self-similar behaviour, or the Hurst-Kolmogorov dynamics. This long-range dependence expresses the tendency of similar physical events to cluster in time and its quantification is expressed through the Hurst coefficient, the mathematics of which is discussed later. The Hurst-Kolmogorov dynamics has been identified by several researchers in various environmental variables [11]. The British engineer Hurst [12] was the first to notice this behaviour while studying Nile's runoff and, earlier, the Russian mathematician Kolmogorov [13] had proposed a mathematical model consistent with this behaviour when studying turbulence. Later, this behaviour was also observed in other variables such as river runoff [14,15,16,17,18,19,20], temperature [21,22,23,24,25,17,26,27], climatic indices such as North Atlantic oscillation [28] and decadal Pacific oscillation [29] and atmospheric pressure fields [30].

So far, the concept of LTP has been incorporated in the modelling of runoff time series applied to reservoir sizing design studies, where it has been found that neglect of LTP usually leads to undersizing of the reservoir storage [31-34]. However, the Hurst-Kolmogorov behaviour has been found to be omnipresent in the most natural processes [11]. Consequently, it will strongly affect the design variables of any engineering project related to the exploitation of natural phenomena, like a large scale hybrid renewable energy system comprising wind and solar energy [35], as it expresses a general natural variability, manifested by the aforementioned fluctuations in multiple time scales [36,10]. Such large scale renewable energy projects aim not only at maximizing the profit but also the reliability in satisfying energy demand or energy target via storage of the excess energy in each time step of the project's operation, usually by means of pumped storage hydropower facilities. Therefore, neglect of this clustering of natural events at the design phase may lead to undersizing of the facilities for energy storage or overestimation of the reliability of produced energy.

A prerequisite for the investigation of Hurst-Kolmogorov dynamics is the analysis of time series having long time period records so that their fluctuations on multiple time scales can be detected. For this reason a sufficient time length (which is generally regarded 100 years or more and no less than 70-80 years) is required. However, hydrometeorological data records, relevant to wind and sun, started in the beginning of the previous century or later and as a consequence time series of such a length are not often available.

Once the marginal distribution and the dependence structure of the wind speed and sunshine duration processes are identified, another relevant issue is to construct a simulation model that can simulate such processes for arbitrarily long times, respecting the identified stochastic properties of each process, as well as the cross-correlation between the two processes. This issue is not trivial as typical stochastic generators may not be able to generate processes with non-normal distributions and with autocorrelations departing from a Markov process.

This study aims to perform a comprehensive analysis of existing data with sufficient length of observation, concerning wind speed and sunshine duration, retrieved from measuring stations all over Europe, in order to investigate the properties of their marginal distributions and mainly to detect the possible presence of LTP by estimating and analyzing the Hurst coefficient values of both processes.

At the same time, it also aims to construct a general stochastic simulation scheme, able to reproduce the statistical characteristics of the natural series including their long term variations. By this, the unpredictable fluctuations characterizing the natural resources (wind and solar radiation) can be taken into account effectively in the design and management of renewable energy systems that comprise both wind and solar systems, and possibly hydropower systems with pumped storage, which provide means for energy storage. The framework is particularly useful to study the reliability in fulfilling the energy demand or the energy production target. This is attempted by using the multivariate stochastic simulation system, Castalia, for the simultaneous generation of long synthetic time series, on the concept of the steady-state simulation, of both variables, being statistically equivalent to the historical ones at all time scales including the large scale, annual and over-annual, in which the Hurst−Kolmogorov behaviour dominates.

2. Description of data

Wind speed and sunshine duration records used in the analysis were retrieved from measuring stations whose data are available online. The criteria for station selection were:

a) A minimum acceptable record length of 70 years so as the Hurst coefficient to be estimated with some accuracy.

b) The existence of metadata related to wind speed data and especially information about the measurement height above ground.

c) A maximum number of three homogeneous periods, in each of which the measurement height above ground is constant (stations with too frequent changes may not be reliable).

However, records of this time length are rare worldwide. In particular, in some continents (Asia, Africa) there is lack of data series of this length. Even in Europe and North America most of the free access records refer to a period of 50 years or less. Furthermore, there is total lack of long solar radiation records. However, this variable can be indirectly estimated by sunshine duration [37].

After an extensive search, wind speed and sunshine duration records fulfilling these criteria were retrieved exclusively from European databases, the KNMI Climate Explorer, the European Climate Assessment & Data (ECA&D) and the Deutscher Wetterdienst databases. A total number of 20 wind speed records and 21 sunshine duration records were found. In addition, wind speed time series from Dublin Airport and Valentia stations were found from the website of the Irish Meteorological Service in chart form and were digitized using the Engauge software. The data retrieved consist of daily records except for the ones of the Irish stations whose records are annual averages. In Tables 1 and 2 stations details are shown.

As known, wind speed at a certain site increases with height and thus, the analysis of wind speed time series requires a single measurement height above ground during the time period of operation. However, the collected daily time series do not refer to a single height above ground due to changes of the observation height at certain time periods in measuring stations. Based on the assumption of process stationarity, this problem was overcome by modifying the data of all the time periods except for the last period of records for each station so as to refer to a unified altitude (data homogenization). Specifically the following procedure was applied:

a) The daily time series were grouped to time periods where the measurement height above ground is constant.

b) The average value of each homogeneous period was estimated.

c) The daily values of each period were multiplied with the ratio of the last period average value to each period average value.

Thus, all records refer to the last measurement height above ground. Neglect of this modification would lead to incorrect analysis, especially Hurst coefficient overestimation, because, as exemplified in Figure 1, the annual raw time series at De Bilt station present a sudden shift owing to the change of measurement height above ground. It is noted though that the procedure followed may result in negative bias for the Hurst coefficient, because the assumption that time averages of different periods are equal, which lies behind step (c), is not true in general and may result in too stable time series.

In addition to wind speed gauge data, the potential of using Reanalysis data for further investigation of LTP in other sites was also examined. Data were obtained from the NCEP/NCAR database in grid mode and their reliability was checked by comparing their records with the station-based ones. This was achieved by selecting the same longitude and latitude with this of measuring stations. As can be seen in Appendix A, Reanalysis data are unreliable as their annual average values are higher than the station-based ones even though they refer to the same or a lower altitude and consequently they are not used any further.

3. Analysis

3.1. Properties of marginal distributions

While wind speed is a non-negative variable not bounded from above, sunshine duration has a domain bounded from both below and above and its upper bound varies in different sites. Thus, the relative sunshine duration is used instead, which varies in the interval [0,1] as it is the ratio of the sunshine duration, *n*, to the astronomical day length, *N*. The latter depends on the latitude and time of the year.

In this study, it was verified that both the Weibull and the Gamma distributions are suitable for wind speed data (as summarized in the Introduction), mainly using graphical depictions like that in Figure 2. Wind speed on daily scale is characterized by positive skewness as shown in Figure 2, representing the frequency histogram of the time series at the Eelde station. The probability density functions of the Gamma and the Weibull distributions are nearly indistinguishable from each other and fit satisfactorily the data. Thus, both distribution functions can be regarded as appropriate for the representation of wind speed.

Considering the fact that the Gamma distribution is used in several software systems (including the software Castalia used in this study as described below) for generation of synthetic time series, it is useful to also associate the distribution of the relative sunshine duration with the Gamma distribution by performing a nonlinear transformation so that the transformed variable takes values in the interval $[0,+\infty)$. Specifically, denoting the relative sunshine duration variable as *X,* the logarithmic transformation $Y= -\ln(1 - X)$ is defined, whose values belong indeed to $[0,+\infty)$, the value $X = 0$ corresponds to $Y = 0$ and X approaching to 1 corresponds to Y approaching to $+\infty$. As shown in Figure 3, the Gamma distribution fits satisfactorily the empirical distribution of the transformed variable *Y* for the relative sunshine duration data of the Eelde station.

Thus, assuming that Y has Gamma distribution with shape parameter κ and scale parameter λ, it is proved (see Appendix B) that the probability density function of the variable X is:

$$\boldsymbol{f_X(x) = (1 - x)^{\lambda-1}\frac{\lambda^{\kappa}}{\Gamma(\kappa)}[-\ln(1 - x)]^{\kappa-1}} \tag{1}$$

The probability density functions of both the theoretical distribution (1) and the Beta distribution are fitted to the frequency histogram of the historical time series of the relative sunshine duration as shown in Figure 3. These two distributions are almost indistinguishable from each other and they fit very well the historical data.

3.2. Long-term persistence

3.2.1. The presence of the LTP in wind speed and sunshine duration

The basic condition for the investigation of LTP is the elimination of seasonality so that the Hurst coefficient estimation be reliable and not affected by periodic fluctuations. As a result, it is preferable to analyze annual time series. Bakker and Bart [38] analyzed annual geostrophic wind speed, deriving from sea level pressure, in Northwestern Europe, and identified LTP using the maximum likelihood method for the simultaneous estimation of standard deviation and the Hurst coefficient, according to Tyralis and Koutsoyiannis [39].

However, in most relevant studies, time series of lower time scales were used, especially daily and hourly ones. Haslett and Raftery [40] proposed the existence of long term persistence in daily wind speed time series, which were firstly deseasonalized and then adjusted to an ARFIMA model. Bouette *et al.* [41], based on the former study, adjusted a GARMA model using the periodogram and Whittle methods and analyzing both seasonality and long term persistence. The results of the methods indicated that seasonality is the dominant phenomenon in these time series.

Furthermore, several researchers tried to investigate LTP using the so-called Detrended Fluctuation Analysis [42]. Feng *et al.* [43] and Kavasseri and Nagarajan [44] used Multi-Fractal Detrended Fluctuation Analysis to analyze daily and hourly wind speed time series in China and North Dakota respectively, having previously deseasonalized the data. Feng *et al.* concluded that no specific behaviour can be determined while Kavasseri and Nagarajan noted that wind speed time series are multi-fractal processes, as suggested by the range of their multi-fractal spectrum. Kocak [45] noted, after analyzing hourly wind speed time series in Turkey using Detrended Fluctuation Analysis, that the scaling region is divided into two parts. In the first one, time series fluctuate randomly while in the second one they are characterized by long term correlations.

Additionally, Rehman and Siddiqi [46] estimated the Hurst coefficient in daily wind speed time series in Saudi Arabia using the wavelet method, but no exact result was extracted as these processes presented either persistence or anti-persistence. Finally Benth and Saltyre [47] claimed that daily wind speed data in Lithuania can be modelled using autoregression models of third and fourth order which indicates short term memory.

As far as sunshine duration is concerned, Liu *et al.* [48] noted intense self-similar behaviour in annual time series in China using the original Hurst's range analysis which led to remarkably high Hurst coefficient values. Tsekov [49] also used the Detrended Fluctuation Analysis in order to examine the existence of LTP in daily sunshine duration time series in Bulgaria and concluded that data do not indicate long term correlations. Finally, Harrouni and Guessoum [50] investigated LTP in both daily and annual global solar radiation time series in Panama and USA using the fractal dimension method and suggested that this process is characterized by anti-persistence.

3.2.2. Stochastic representation of LTP and estimation of the Hurst coefficient

Let X_i denote a hydrometeorological process with i=1,2,…., denoting discrete time (years). Furthermore, let its mean be μ :=E[X_i], its autocovariance γ_j :=cov[X$_\mathrm{i}$,X_{i+j}], its autocorrelation ρ_j :=corr[X_i.X_{i+j}] = γ_j/γ_0 (j=0,±1, ±2,…) and its standard deviation σ := $\gamma_0^{1/2}$. Let k be a positive integer representing time scale of aggregation for the X_i process. The mean aggregated stochastic process on that time scale is:

$$X_i^{(k)} := \frac{1}{k}\sum_{l=(i-1)k+1}^{ik} X_i \tag{2}$$

Hydrometeorological processes characterized by LTP can be represented by a stochastic process known as simple scaling stochastic (SSS) process or a Hurst-Kolmogorov (HK) process [17,51]. By definition of the HK process, the variance of the mean aggregated stochastic process is a power law of k with exponent $2H-2$, where H is the Hurst coefficient:

$$\mathrm{Var}[X_i^{(k)}] = k^{2H-2}\gamma_0, \tag{3}$$

The value of H varies from 0 to 1. A value in the interval (0,0.5) indicates anti-persistence, which denotes that an increase in the values of time series is followed by a subsequent decrease, a behaviour that is not normally observed in nature. A value 0.5 corresponds to a purely random process known as white noise. A value in the interval (0.5,1) indicates persistence and implies positive autocorrelation. This behaviour is visualized as multiyear fluctuations and apparent trends, as can be seen in Figure 4, which depicts a sunshine duration time series characterized by a high value of Hurst coefficient (H=0.9) and for comparison, a time series of a white noise with the same mean and standard deviation.

From equation (3) it is observed that if standard deviation (square root of variance) is plotted against time scale on a logarithmic graph, known as the climacogram, a straight line is formed with slope $H-1$. Wind speed and sunshine duration data support this behaviour, as an approximation, as can be seen in the climacograms of Figure 5. Based on this scaling property

Koutsoyiannis [17] proposed a method for the simultaneous estimation of standard deviation σ and the Hurst coefficient H using the least square error of sample standard deviation estimates (LSSD). The method was later examined by Tyralis and Koutsoyiannis [39] and Sheng *et al*. [52] and was found to perform very satisfactorily. In general, the method can be perceived as a fitting of a straight line in a climacogram like that of Figure 5 using linear regression of the mean aggregated standard deviation $\sigma^{(k)}$ on time scale k. The Hurst coefficient is estimated from the slope of the climacogram using sample standard deviation estimates on all time scales. The only difference in the LSDD method is that it explicitly considers the bias of the standard statistical estimator of standard deviation on all time scales.

3.2.3. Results of the LSSD method

The results derived from the LSSD method are summarized in Figure 6 in terms of frequency histograms of the Hurst coefficient values of the historical time series. Concerning both variables, markedly high Hurst coefficient values are estimated and it is remarkable that no value lies in the interval (0,0.5) i.e., no time series is characterized by anti-persistence. The majority of the values of both variables lie in the interval (0.6,0.9) and it is notable that nearly 15% of the values lie in the interval (0.9,1) which indicates very strong self-similar behaviour. In general, the results of the method imply that both wind speed and sunshine duration time series are characterized by Hurst-Kolmogorov behaviour.

3.2.4. Representative value of Hurst coefficient

In order to assume a representative value of the Hurst coefficient for the historical time series of both variables, the same number of synthetic time series is generated, each one of the same mean value, standard deviation and record length with the respective historical one, but to define the autocorrelation function a unique value of the Hurst coefficient is used, the same for all time series. For this generation the multiple time-scale fluctuation approach [16] is used. This method implies that a weighted sum of three exponential functions of time lag can approximate satisfactorily the autocorrelation function of a simple scaling process on the basic time scale. The generated process is the sum of three independent AR(1) processes.

In the end of the generation process the sample value of the Hurst coefficient is estimated for all synthetic time series using the LSSD method. The frequency histogram of these H values is then compared to the histogram of the historical H values. By trying consecutive initial theoretical values of H until a convergence between the histogram of historical and synthetic timeseries is achieved, it is concluded that the value $H = 0.84$ can be regarded as representative for both variables as can be seen in Figure 6.

4. Modelling

4.1. The software system Castalia

Castalia is a software system that was initially introduced for the simulation of hydrometeorological processes, like rainfall, runoff and evaporation. It performs multivariate stochastic simulation on annual, monthly and daily scale using the three-parameter Gamma distribution function in order to generate synthetic time series. The program preserves the essential marginal statistics, specifically mean value, standard deviation and skewness, as well as the joint second order statistics, namely auto- and cross-correlation, on these three time scales. To estimate these statistics, historical daily time series are used as input data and subsequently, historical monthly and annual time series are created by aggregating daily data inside the system. Furthermore, Castalia reproduces the LTP considering it as a special instance of a parametrically defined generalized autocovariance function [53].

The generation of the synthetic time series is performed in three time-aggregation levels. Initially, a theoretical autocovariance function is defined for each annual variable, which can include LTP. The dependence structure is reproduced through the Symmetric Moving Average (SMA) model, whose parameters are estimated using the annual marginal and joint second order statistics and the autocovariance function [53]. The SMA model is used for the generation of the synthetic annual time series. Following this, auxiliary monthly time series are generated using the Periodic Autoregressive (PAR(1)) model, whose parameters are estimated using the monthly marginal and joint second order statistics. A disaggregation process (called a coupling transformation) modifies the auxiliary monthly time series in order to become consistent with the annual ones [54]. The arising deviations in the statistics that must be preserved, which may be caused by the disaggregation process, are overcome through a repetitive Monte-Carlo process that accompanies the disaggregation process. By this, a statistically independent sequence of monthly variables is found which approximates the annual value. The synthetic daily time series are then generated in a similar manner by disaggregation of monthly values into daily.

4.2. Preservation of probability of zero values

Sunshine duration can take on a zero value with nonzero probability and thus the preservation of the historical probability of zero values is important. This is similar to rainfall, which is an intermittent process and the Castalia program, which was originally developed to simulate rainfall, can handle the intermittent behaviour on the daily time scale [55]. This can be achieved by appropriate choice of the parameters that are used for this aim into the system after the generation of daily auxiliary time series. These include:

a) The parameters λ_1 and λ_2 that are used to adjust the probabilities k_1 and k_2 of a Markov chain model. These probabilities are adjusted by the parameters λ_1 and λ_2, respectively, of the historical probability of zero values of each variable. Specifically, k_1 expresses the probability of a zero value occurring in the current time step if there is also a zero value in the previous time step while k_2 expresses the probability of a zero value occurring in the current time step if there is a non-zero value in the previous time step.

b) A parameter k_3 which expresses the probability of the values of all variables in the current time step being zero if one of them is zero.

c) The parameters of a round-off rule and specifically a threshold l_0 and a percentage $\pi_{0,}$ where the rule implies that a percentage π_0 of values below this threshold are converted into zero values.

Note that all three procedures described above generate zero values and thus they all contribute to the final frequency of zero values.

4.3. Results

4.3.1. First application – Untransformed variables

In the first application of the program 1000-year synthetic time series are generated in 8 measuring stations having both wind speed and sunshine duration records (16 variables in total). The length of the simulation period (1000 years) was determined as explained in Appendix C. Although the domain of the sunshine duration is bounded from above and below and thus, the Gamma distribution is not appropriate to represent this variable, data are initially imported into the system without any transformation to test the program performance in a fast application without pre-processing of data. For this reason possible deviations concerning the skewness of the sunshine duration variable are expected.

The model parameters are calculated automatically by the program except for those controlling the preservation of the probability of zero values, which are determined by a trial and error procedure. For the latter, the following values were finally adopted π_0 = 0.9 for l_0=0.4; λ_1 = 0.3, λ_2 = 0.1; k_3= 0.

Application results are presented in Figures 7-10 for one of the stations (Eelde). It is observed that the statistical characteristics, the Hurst coefficient as well as the cross-correlation coefficient of both processes are preserved on annual and monthly scale. On daily scale wind speed statistics and cross-correlation coefficient are also satisfactorily preserved.

However, the skewness of the sunshine duration variable is overestimated. This divergence is caused, as expected, by the fact that the domain of the sunshine duration variable differs from the domain of the Gamma distribution, used for the generation of the time series. As a consequence, several synthetic sunshine duration values, being higher than the upper bound of the variable domain, are generated due to the fact that the domain of Gamma distribution is the $[0,+\infty)$. These values are obviously unrealistic and lead to skewness overestimation.

4.3.2. Second application – transformed sunshine duration

In order to overcome the overestimation of the skewness which was encountered in the first application, a second application is performed using the previous wind speed data and the proposed logarithmic transformation of sunshine duration data in each of the 8 stations. After the generation of the synthetic time series, the *Y* daily time series are manually subject to the reverse transformation so as the synthetic daily time series of the sunshine duration to arise. Then these time series are compared with the historical ones.

The model parameters controlling the preservation of the probability of zero values are also determined by a trial and error procedure. For the latter, the following values were finally adopted $\pi_0 = 0.9$ for $l_0 = 0$; $\lambda_1 = 0.3$, $\lambda_2 = 0.1$; $k_3 = 0$.

Both the Hurst coefficient and the marginal statistical characteristics of the wind speed process are preserved satisfactorily on all time scales as in the first application. Application results for the sunshine duration are presented in Figures 11-14 for the same station as before (Eelde). The Hurst coefficient and the statistical characteristics on annual and monthly scale are also generally preserved. The only deviations observed are related to monthly skewness, which is underestimated, and monthly autocorrelation, which is overestimated. However, these features are not of great importance as the design focuses on the daily scale. It is remarkable that the cross-correlation of the variables is preserved even though the logarithmic transformation is used in the simulation. On daily time scale, skewness is now satisfactorily reproduced and the only existing deviation refers to autocorrelation which is slightly overestimated. The cross-correlation of the variables, which is high as the time scale of the design is daily, as well as the probability of zero values are also preserved. The deviations mentioned above on monthly and daily time scales are caused due to the fact that the iterative Monte Carlo process and the adjusting-disaggregation procedures are performed inside the program, so they are applied to the logarithmic transformation of the variable instead of the actual one.

5. Conclusions

Simulation of wind speed and sunshine duration gains interest because these processes are associated with renewable energy production. Knowing the marginal distribution as well as the stochastic structure of these processes, including LTP, is important as the solar and wind energy strongly depend on them.

The results derived from the analysis using data from Europe indicate that the Gamma and the Weibull distributions are suitable for representing wind speed. For the relative sunshine duration a theoretical distribution function, which is a transformation of the Gamma distribution, is derived. This is practically indistinguishable from the Beta distribution and they both fit satisfactorily to relative sunshine duration data.

Both processes are found to be characterized by strong self-similar behaviour with values of the Hurst coefficient, H, (estimated using the LSSD method) markedly higher than the value H = 0.5, corresponding to white noise, for the majority of the time series. After implementing a Monte Carlo approach, a unique value H = 0.84 is found to be representative for both variables. However, it should be noted that, due to the small amount of samples, further investigation is required. Also, as the observations have started in their majority, all over the world, after 1940, the record lengths are not quite sufficient to support a safe estimation of the Hurst coefficient.

The detailed empirical analysis of the stochastic properties allows stochastic simulation of the two processes which is particularly useful for the rational design of a renewable energy system comprising wind and solar energy. Specifically, synthetic time series of wind speed and sunshine duration data are simultaneously generated on annual, monthly and daily scale using the software system Castalia.

Castalia performs this stochastic simulation satisfactorily as the marginal and joint second order statistics as well as the LTP of the historical data retrieved from measuring stations with common observations are well reproduced. An overestimation of the skewness of daily sunshine duration which arises when raw data are used is attributed to the bounded domain of the variable and it can be overcome by using the proposed logarithmic transformation. Some deviations which arise in the monthly skewness and autocorrelation when the transformed variable is used are not of great importance due to the fact that the time scale of the design is daily while the overestimation of the daily autocorrelation is slight. Thus, considering that Castalia was initially developed for the simulation of rainfall, runoff and evaporation, it is also confirmed that this program can conduct stochastic simulations of a wide spectrum of hydrometeorological variables on three important time scales, annual, monthly and daily.

Acknowledgements: We thank two anonymous reviewers for the constructive review comments on an earlier version of this manuscript as well as Andreas Efstratiadis, Stefanos Kozanis, Yannis Dialynas and Panagiotis Dimitriadis for their help during the preparation of the study.

References

[1] Koutsoyiannis D, Makropoulos C, Langousis A, Baki S, Efstratiadis A, Christofides A, et al. Climate, hydrology, energy, water: recognizing uncertainty and seeking sustainability. Hydrol Earth Syst Sci 2009;13:247-57.

[2] Carta JA, Ramirez P, Velasquez S. A review of wind speed probability distributions used in wind energy analysis: case studies in the Canary Islands. Renew Sust Energ Rev 2009;13:933-55.

[3] Zhou J, Erdem E, Li G, Shi J. Comprehensive evaluation of wind speed distribution models: a case study for North Dakota sites. Energy Convers Manag 2010;51:1449-58.

[4] Garcia A, Torres JL, Prieto E, de Francisco A. Fitting wind speed distributions: a case study. Sol Energy 1998;62:139-44.

[5] Yilmaz V, Çelik HE. A statistical approach to estimate the wind speed distribution: the case of Gelibolu region. Dogus¸ Üniversitesi Dergisi 2008;9:122-32.

[6] Darbandi A, Aalami MT, Asadi H. Comparison of four distributions for frequency analysis of wind speed. Env Nat Resour Res 2012;2.

[7] Sulaiman YM, Hlaing OWM, Wahab MA, Zakaria A. Application of beta distribution model to Malaysian sunshine data. Renew Energy 1999;18:573-9.

[8] Bashahu M, Nsabimana JC. Statistical analysis of sunshine duration measurements in Burundi using beta distributions. In: World Conference on Physics and Sustainable Development Durban Poster No102244, South Africa 2005.

[9] Chia E, Hutchinson MF. The beta distribution as a probability model for daily cloud duration. Agric For Meteorol 1991;56:195-208.

[10] Koutsoyiannis D. Hurst-Kolmogorov dynamics and uncertainty. J Am Water Resour Assoc 2011;47:481-95.

[11] Markonis Y, Koutsoyiannis D. Climatic variability over time scales spanning nine orders of magnitude: connecting Milankovitch cycles with Hurst-Kolmogorov dynamics. Surv Geophys 2013;34:181-207.

[12] Hurst HE. Long term storage capacities of reservoirs. Trans Am Soc Civ Eng 1951;16:776-808.

[13] Kolmogorov AN. Wienersche Spiralen und einige andere interessante Kurven in Hilbertschen Raum. Dokl Akad Nauk 1940;26:115-8.

[14] Eltahir EAB. El Niño and the natural variability in the flow of the Nile river. Water Resour Res 1996;32:131-7.

[15] Jiang T, Zhang Q, Blender R, Fraedrich K. Yangtze Delta floods and droughts of the last millennium: abrupt changes and long term memory. Theor Appl Climatol 2005;82:131-41.

[16] Koutsoyiannis D. The Hurst phenomenon and fractional Gaussian noise made easy. Hydrol Sci J 2002;47:573-95.

[17] Koutsoyiannis D. Climate change, the Hurst phenomenon, and hydrological statistics. Hydrol Sci J 2003;48:3-24.

[18] Radziejewski M, Kundzewicz ZW. Fractal analysis of flow of the river Warta. J Hydrol 1997;200:280-94.

[19] Sakalauskiene G. The Hurst phenomenon in hydrology. Environ Res Eng Manag 2003;3:16-20.

[20] Wang G, Jiang T, Chen G. Structure and long-term memory of discharge series in Yangtze River. Acta Geogr Sin 2006;61:47-56.

[21] Alvarez-Ramirez J, Alvarez J, Dagdug L, Rodriguez E, Carlos Echeverria J. Long term memory dynamics of continental and oceanic monthly temperatures in the recent 125 years. J Phys A 2008;387:3629-40.

[22] Bloomfield P. Trends in global temperature. Clim Change 1992;21:1-16.

[23] Fraedrich K, Blender R. Scaling of atmosphere and ocean temperature correlations in observations and climate models. Phys Rev Lett 2003;90:1-4.

[24] Fraedrich K, Blender R, Zhu X. Continuum climate variability: long-term memory, extremes, and predictability. Int J Mod Phys B 2009;23:5403-16.

[25] Koscielny-Bunde E, Bunde A, Havlin S, Roman HE, Goldreich Y, Schellnhuber HJ. Indication of a universal persistence law governing atmospheric variability. Phys Rev Lett 1998;81:729-32.

[26] Varotsos C, Kirk-Davidoff D. Long-memory processes in ozone and temperature variations at the region 60_ Se60_ N. Atmos Chem Phys 2006;6:4093-100.

[27] Yano J, Blender R, Zhang C, Fraedrich K. 1/f e noise and pulse-like events in the tropical atmospheric surface variabilities. Q J Roy Meteorol Soc 2004;130:1697-721.

[28] Stephenson DB, Paven P, Bojariu R. Is the North Atlantic oscillation a random walk? Int J Climatol 2000;20:1-18.

[29] Khaliq MN, Gachon P. Pacific decadal oscillation climate variability and temporal pattern of winter flows in Northwestern North America. J Hydrometeorol 2010;11:917-33.

[30] Giannoulis S, Ioannou C, Karantinos E, Malatesta L, Theodoropoulos G, Tsekouras G, et al. Long term properties of monthly atmospheric pressure fields, EGU General AssemblyIn Geophysical research abstracts, vol. 14; 2012. Vienna, 4680.

[31] Koutsoyiannis D., Reliability concepts in reservoir design. Water Encyclopedia, Vol. 4, Surface and Agricultural Water, edited by J. H. Lehr and J. Keeley, 259–265, Wiley, New York, 2005.

[32] Kottegoda NT. Stochastic water resources technology. London: Macmillan Press; 1980. p. 384.

[33] Klemes V, Srikanthan R, McMahon TA. Long-memory flow models in reservoir analysis: what is their practical value? Water Resour Res 1981;17:737-51.

[34] Lettenmaier DP, Burges SJ. Operational assessment of hydrologic models of long-term persistence. Water Resour Res 1977;13:113-24.

[35] Tsekouras G, Ioannou C, Efstratiadis A, Koutsoyiannis D. Stochastic analysis and simulation of hydrometeorological processes for optimizing hybrid renewable energy systems, EGU General Assembly In Geophysical research abstracts, vol. 15; 2013. Vienna.

[36] Koutsoyiannis D. Nonstationarity versus scaling in hydrology. J Hydrol 2006;324:239-54.

[37] Koutsoyiannis D, Xanthopoulos Th. Engineering hydrology (in Greek). 3rd ed. Athens: National Technical University of Athens; 1999. p. 418.

[38] Bakker AMR, Van den Hurk J J M Bart. Estimation of persistence and trends in geostrophic wind speed for the assessment of wind energy yields in Northwest Europe 4. Clim Dynam 2012;39:767-82.

[39] Tyralis H, Koutsoyiannis D. Simultaneous estimation of the parameters of the Hurst-Kolmogorov stochastic process. Stoch Env Res Risk A 2011;25:21-33.

[40] Haslett J, Raftery AE. Space modeling with long memory dependence: assessing Ireland's wind power resource. Appl Statv 1989;38:1-50.

[41] Bouette JC, Chassagneux JF, Sibai D, Terron R, Charpentier A. Wind in Ireland: long memory or seasonal effect? Stoch Env Res Risk A 2006;20:141-51.

[42] Peng C-K, Buldyrev SV, Havlin S, Simons M, Stanley HE, Goldberger AL. Mosaic organization of DNA nucleotides. Phys Rev E 1994;49:1685-9.

[43] Feng T, Fu Z, Deng X, Mao J. A brief description to different multi-fractal bevaviors of daily wind speed records over China. Phys Lett A 2009;373:4134-41.

[44] Kavasseri RG, Nagarajan R. A multifractal description of wind speed records. Chaos Soliton Fract 2005;24:165-73.

[45] Kocak K. Examination of persistence properties of wind speed records using detrended fluctuation analysis. Energy 2009;34:1980-5.

[46] Rehman S, Siddiqi AH. Wavelet based Hurst exponent and fractal dimensional analysis of Saudi climatic dynamics. Chaos Soliton Fract 2009;40:1081-90.

[47] Benth JS, Saltyte L. Spatial-temporal model for wind speed in Lithuania. J Appl Stat 2011;38:1151-68.

[48] Liu Liu, Xu ZX, Huang JX. Statial-temporal variation and abrupt changes for major climate variables in the Taihu Basin, China. Stoch Env Res Risk A 2011;26:777-91.

[49] Tsekov M. Detrended fluctuation analysis of weather records from local place in south Bulgaria. C R Acad Bulgare Sci 2003;56:8-25.

[50] Harrouni S, Guessoum A. Using fractal dimension to quantify long-range persistence in global solar radiation. Chaos Soliton Fract 2009;41:1520-30.

[51] Koutsoyiannis D. A random walk on water. Hydrol Earth Syst Sci 2010;14:585-601.

[52] Sheng H, Chen YQ, Qiu TS. Tracking performance and robustness analysis of Hurst estimators for multifractional processes. IET Signal Process 2012;6:213-26.

[53] Koutsoyiannis D. A generalized mathematical framework for stochastic simulation and forecast of hydrologic time series. Water Resour Res 2000;36:1519-33.

[54] Koutsoyiannis D. Coupling stochastic models of different time scales. Water Resour Res 2001;37:379-92.

[55] Koutsoyiannis D, Onof C, Wheater HS. Multivariate rainfall disaggregation at a fine timescale, Water Resour Res, 2002:39, doi:10.1029/2002WR001600

[56] Koutsoyiannis D. Stochastic hydrology (in Greek). 4th ed. Athens: National Technical University of Athens; 1997. p. 312.

[57] Koutsoyiannis D. On the quest for chaotic attractors in hydrological processes. Hydrolog Sci J 2006;51:1065-91.

Appendix A: Assessment of Reanalysis data

Figure A.1 compares station-based and reanalysis data for the location of the De Bild station at an annual basis. It can be seen that the two series do not correlate well. In addition the reanalysis data series has annual average values higher than the station-based ones even though the former refer to lower altitude than the latter.

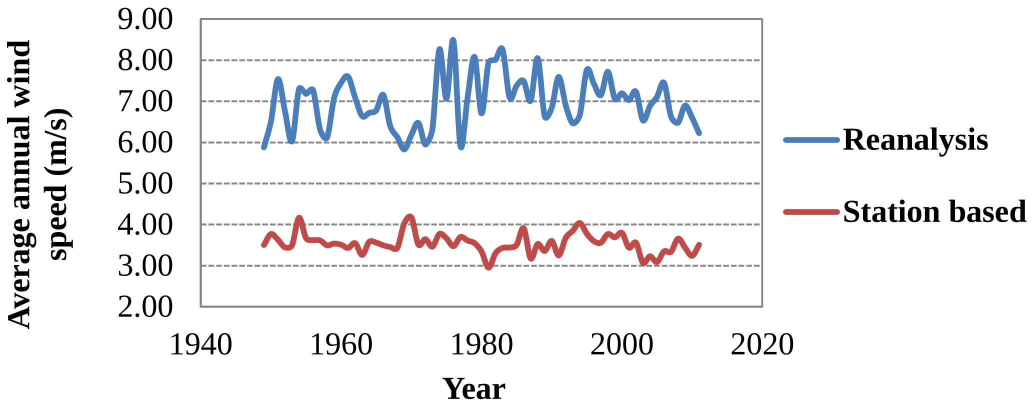


Figure A.1: Mean annual wind speed time series from station-based (20m above ground) and Reanalysis data (10m above ground) at De Bilt station.

Appendix B: Proof of equation (1)

Let X, Y, where 0< X <1 be random variables so that:

$$X = 1 - e^{-Y} = \mathrm{g}(Y) \tag{B.1}$$

$$Y \ = \ -\ln(1 - X) \ = \mathrm{g}^{-1}(X) \tag{B.2}$$

The event $\{X \leq x\}$ is identical to the event $\{Y \leq \mathrm{g}^{-1}(x)\}$. As a result, the distribution functions of Y and X are linked through the equation:

$$F_X(x) \ = \ P\{X \ \leq \ x\} \ = \ P\{Y \ \leq \ \mathrm{g}^{-1}(x)\} \ = F_Y\,(\mathrm{g}^{-1}(\mathrm{x})) \tag{B.3}$$

The variables are continuous, and the function g is differentiable with derivative equal to:

$$\mathrm{g}^{'}(Y) = \ (1 - \mathrm{e}^{-Y})' \ = \mathrm{e}^{-Y} \tag{B.4}$$

The probability density function of the variable X is a function of the density of the variable Y [56]:

$$f_X\,(x) = \frac{f_Y(\mathrm{g}^{-1}(x))}{|\mathrm{g}'\,(\mathrm{g}^{-1}(x))|} = \frac{1}{e^{-y}} f_Y\,[-\ln(1-x)] \tag{B.5}$$

As the variable Y for the continuous part of the distribution ($Y > 0$) is represented by the Gamma distribution function with shape parameter $\kappa > 0$ and scale parameter $\lambda > 0$, the probability density function of the variable X is:

$$f_X\,(x) = \frac{1}{e^{-y}} \frac{\lambda^{\kappa}}{\Gamma(\kappa)} [-\ln(1-x)]^{\kappa-1}\, e^{-\lambda[-\ln(1-x)]} \tag{B.6}$$

or

$$f_X\,(x)= \frac{1}{1-x} \frac{\lambda^{\kappa}}{\Gamma(\kappa)} [-\ln(1-x)]^{\kappa-1} (1-x)^{\lambda} \tag{B.7}$$

and finally

$$f_X\,(x) =(1-x)^{\lambda-1} \frac{\lambda^{\kappa}}{\Gamma(\kappa)} [-\ln(1-x)]^{\kappa-1} \tag{B.8}$$

Appendix C: Determination of the length of simulation period

In the estimation, through stochastic simulation, of a probability p, from a sample with relatively large size N of independent identically distributed random variables, it is known that the length of the confidence interval of the estimate is [31,57]:

$$2\, z_{(1+\gamma)/2}\, [p\,(1-p)\,/\,N]^{0.5} = 2\,c\,p \tag{1}$$

where γ is the confidence coefficient, z_a is the a-quantile of the standard normal distribution, and c is the acceptable relative error. Solving for N, the minimum sample size $N_{\min}$ that is required for estimating the probability p is:

$$N_{\min}= (z^2_{(1+\gamma)/2}\,/\,c^2)\,(1/p-1) \tag{2}$$

For a daily step of design, a confidence coefficient $\gamma = 0.95$, an acceptable relative error c = 10% and probability of failure p = 1‰, the required sample size (number of days of simulation) needed to obtain an accurate estimate of p on a daily basis is $N_{\min} \approx 384000$ (days), which is rounded off to a simulation period of 1000 years. It is noted though that, as the hydrometeorological processes are not independent in time, the required period should be greater than 1000 years.

Tables

Station	Country	Longitude(^{o})	Latitude (^{o})	Measurement height above ground (m)	Altitude (m)	Source	Period of record
Vlissingen	Netherlands	3.60 E	51.45N	15	8	KNMI	1909-2011
Regenburg	Germany	12.10 E	49.00 N	15	377	ECA&D	1946-2011
Potsdam	Germany	13.10 E	52.40 N	37.7	81	ECA&D	1911-2011
Maastricht	Netherlands	5.78 E	50.92 N	10	114	KNMI	1906-2011
Karlsruhe	Germany	8.40 E	49.00 N	47.4	113	ECA&D	1945-2011
Hohenpeissenberg	Germany	11.01 E	47.48 N	15	980	ECA&D	1939-2011
Hof	Germany	11.90 E	50.30 N	12	624	ECA&D	1946-2011
Giessen Wettenberg	Germany	8.38 E	50.36 N	10	203	ECA&D	1939-2011
Eelde	Netherlands	6.58 E	53.13 N	10	3.5	KNMI	1904-2011
Dresden	Germany	13.80 E	51.10 N	10	220	KNMI	1941-2011
De Kooy	Netherlands	4.78 E	52.92 N	10	0.5	KNMI	1908-2011
De Bilt	Netherlands	5.18 E	52.10 N	20	1.9	KNMI	1904-2011
Tarifa	Spain	5.35 W	36.00 N	10	32	ECA&D	1945-2012
Bjoernoeya	Norway	19.0 E	74.50 N	10	16	ECA&D	1920-2011
San Sebastian	Spain	2.54 W	43.18 N	10	251	ECA&D	1933-2011
Dublin Airport	Ireland	6.30 W	53.42 N	-	85	MET EIREANN	1944-2010
Valentia	Ireland	10.24 W	51.94 N	-	9	MET EIREANN	1940-2010
Fokstua	Norway	9.17 E	62.07 N	10	975	ECA&D	1923-2011
Karasjok	Norway	25.52 E	69.47 N	10	133	ECA&D	1939-2011
Sula	Norway	8.45 E	63.85 N	10	6	ECA&D	1938-2011

Table 1: Stations with wind speed measurements

Station	Country	Longitude(°)	Latitude (°)	Altitude (m)	Source	Period of record
Alicante	Spain	0.60 W	38.30 N	82	ECA&D	1939-2011
Vlissingen	Netherlands	3.60 E	51.45N	8	KNMI	1907-2011
Lindenberg	Germany	14.10 E	52.20 N	112	ECA&D	1907-2011
Potsdam	Germany	13.10 E	52.40 N	81	ECA&D	1901-2011
Maastricht	Netherlands	5.78 E	50.92 N	114	KNMI	1906-2011
Karlsruhe	Germany	8.40 E	49.00 N	113	ECA&D	1936-2011
Hohenpeissenberg	Germany	11.01 E	47.48 N	980	ECA&D	1937-2011
Zugspitze	Germany	10.98 E	47.42 N	2960	ECA&D	1901-2011
Zuerichfluntern	Switzerland	8.57 E	47.38 N	556	ECA&D	1901-2011
Eelde	Netherlands	6.58 E	53.13 N	3.5	KNMI	1906-2011
Geneve	Switzerland	6.10 E	46.20 N	238	ECA&D	1901-2011
De Kooy	Netherlands	4.78 E	52.92 N	0.5	KNMI	1909-2011
De Bilt	Netherlands	5.18 E	52.10 N	1.9	KNMI	1901-2011
Granada	Spain	3.59 W	37.17 N	680	ECA&D	1942-2010
Aachen	Germany	6.10 E	50.80 N	264	ECA&D	1935-2010
Basel	Switzerland	7.61 E	47.56 N	265	ECA&D	1901-2011
Lugano	Switzerland	8.97 E	46.00 N	276	ECA&D	1901-2011
Madrid	Spain	3.65 W	40.40 N	687	ECA&D	1920-2011
Saentis	Switzerland	9.40 E	47.20 N	2500	ECA&D	1901-2011
Valencia	Spain	0.38 W	39.48 N	35	ECA&D	1938-2011
Zagreb	Croatia	16.03 E	45.82 N	123	ECA&D	1900-2011

Table 2: Stations with sunshine duration measurements

Figures

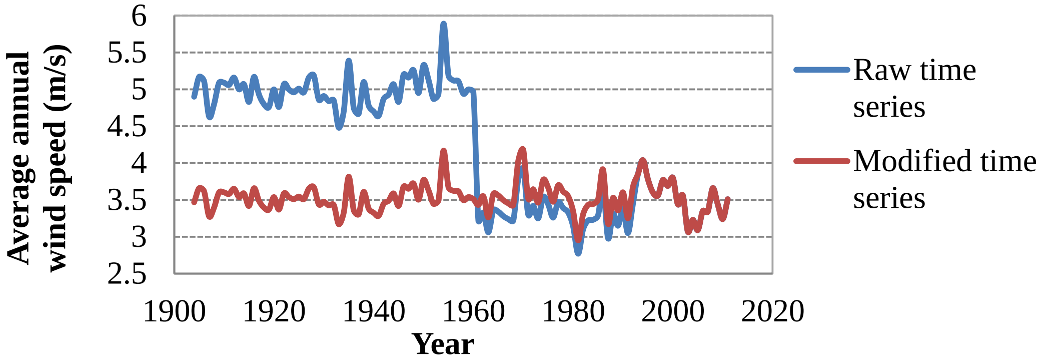


Figure 1: Annual averages of wind speed time series at De Bilt station before and after modification; three different periods were identified: 1904-1960 (37 m above ground), 1961-1992 (10 m above ground), 1993-2011 (20 m above ground).

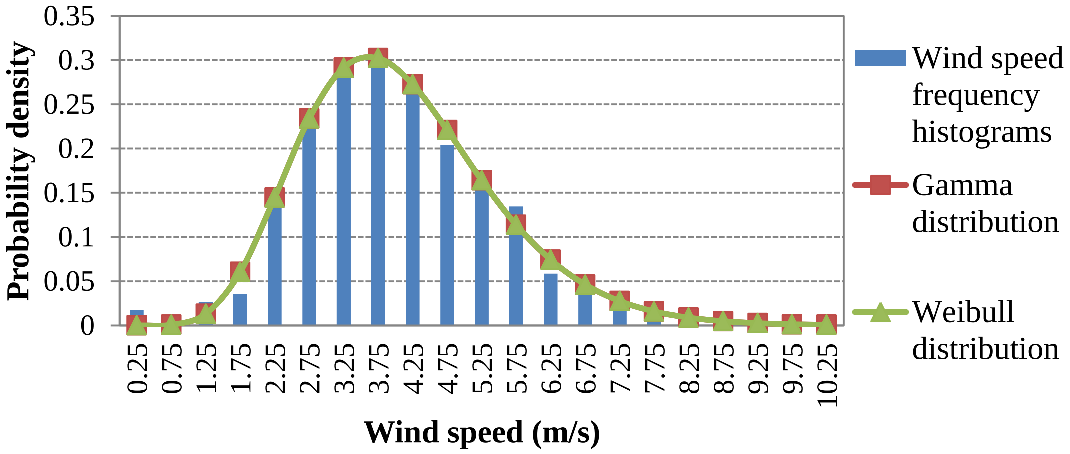


Figure 2: Frequency histograms and probability density functions of historical daily wind speed time series at Eelde station in June.

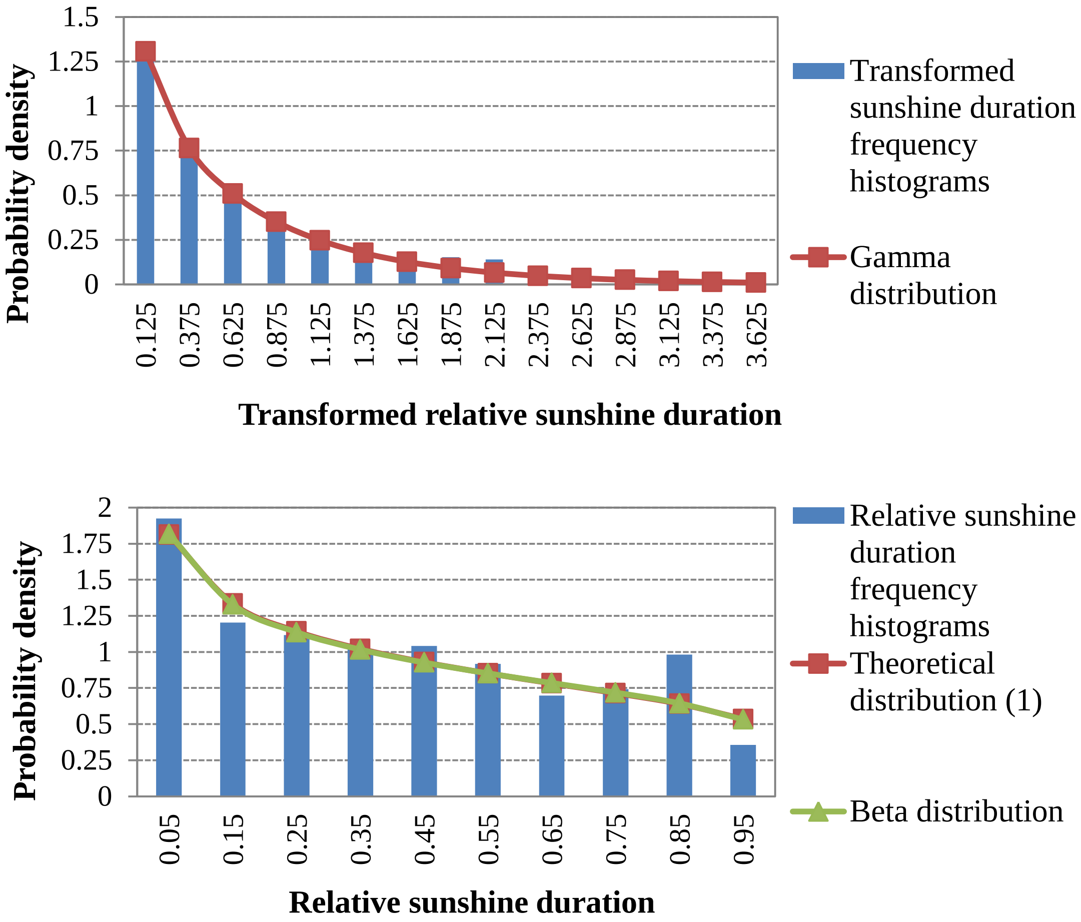
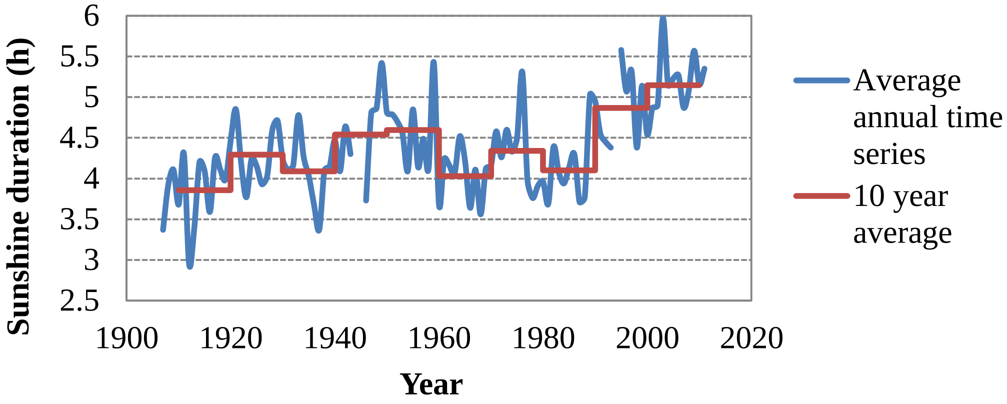


Figure 3: Frequency histograms and probability density functions of the logarithmic transformation of historical daily sunshine duration time series (above) and the relative sunshine duration time series (below) at Eelde station in June.



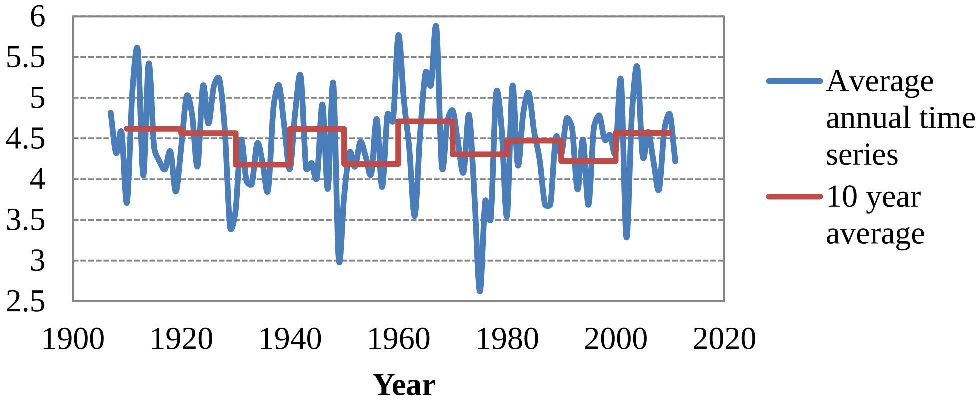


Figure 4: Annual and 10-year averages of sunshine duration time series at Vlissingen station (above) and a series of white noise with the same mean and standard deviation (below).

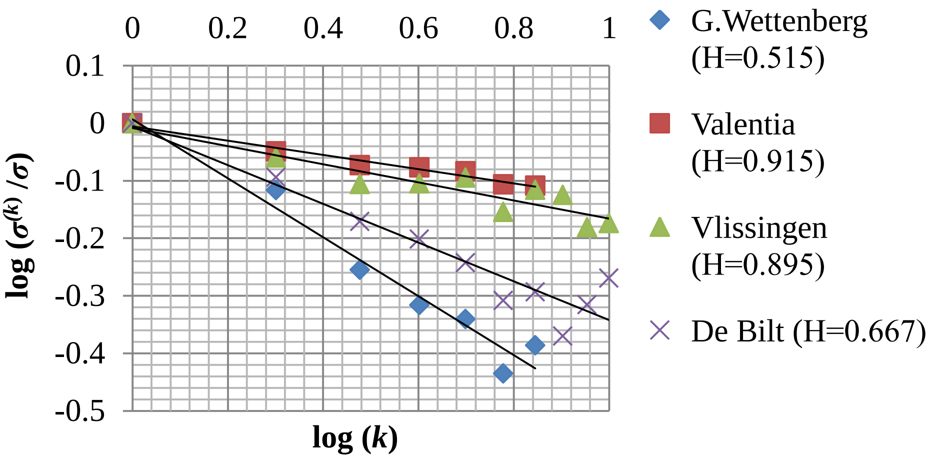
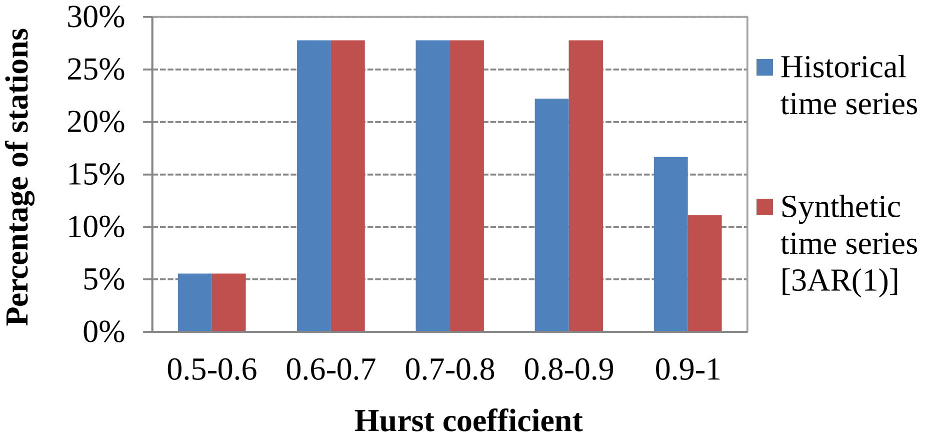


Figure 5: Climacograms of wind speed (Valentia, G.Wettenberg) and sunshine duration (Vlissingen, De Bilt) time series.



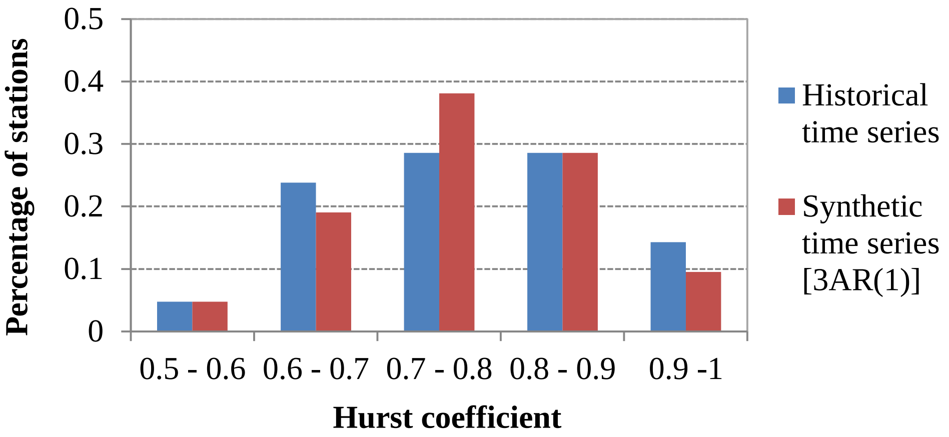


Figure 6: Comparison of Hurst coefficient frequency histograms of historical and synthetic, with a theoretical *H* = 0.84, time series for wind speed (above) and sunshine duration (below).

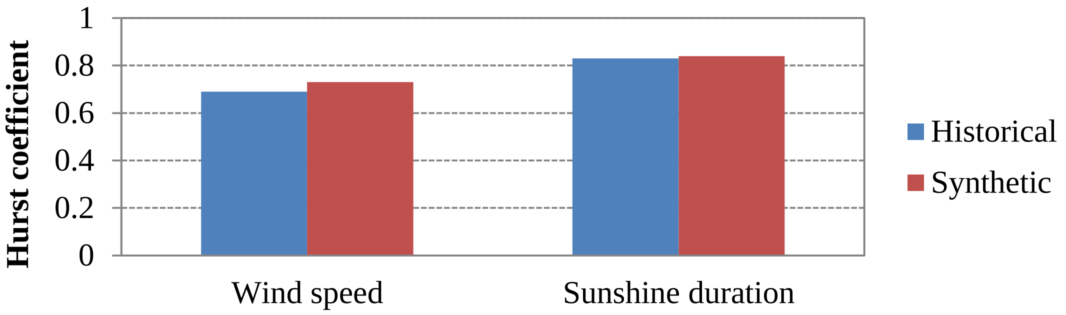


Figure 7: Comparison of Hurst coefficient of historical and synthetic time series at the Eelde station; first application.

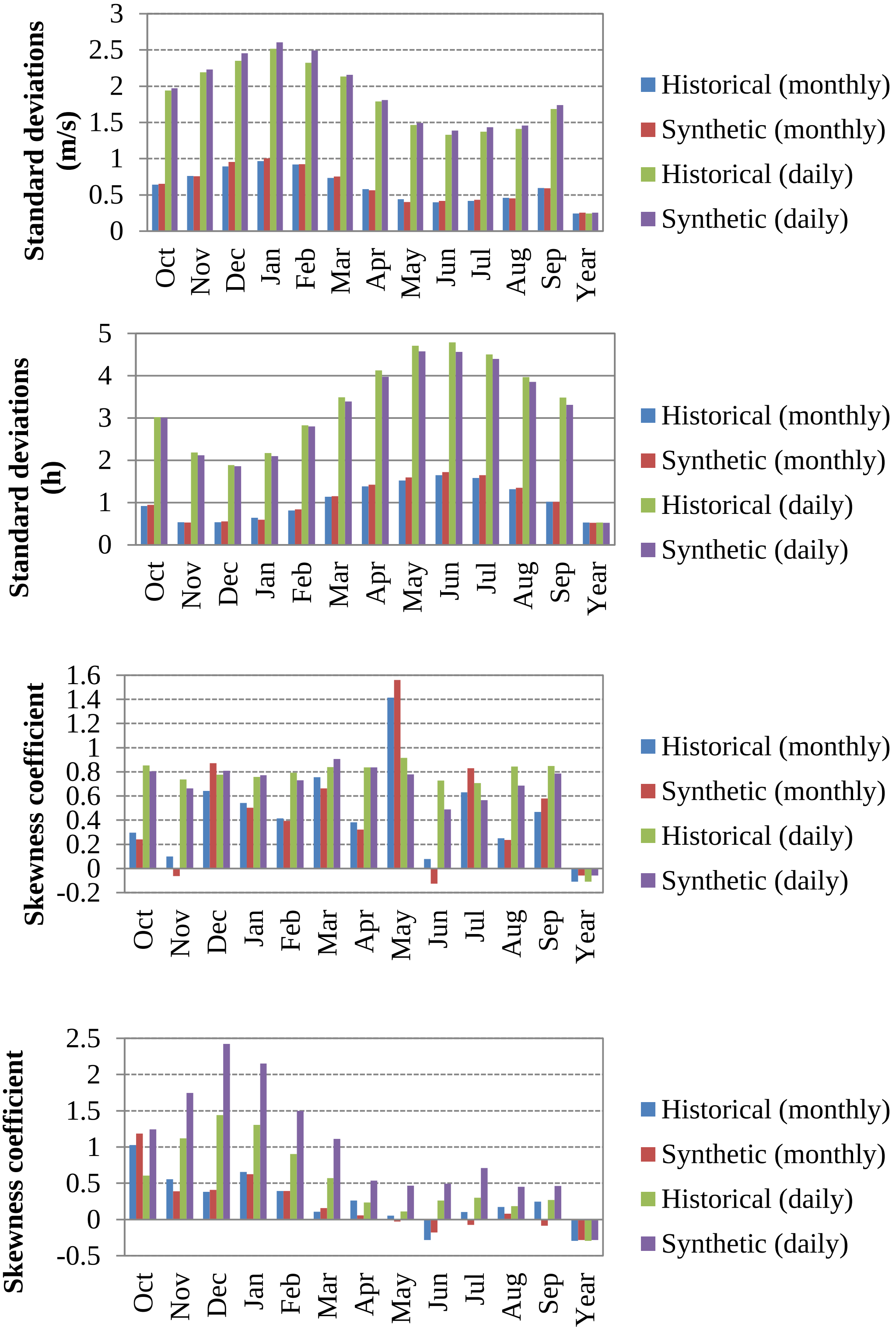


Figure 8: Comparison of annual, monthly and daily standard deviations and skewness coefficients of historical and synthetic wind speed and sunshine duration time series at the Eelde station; first application.

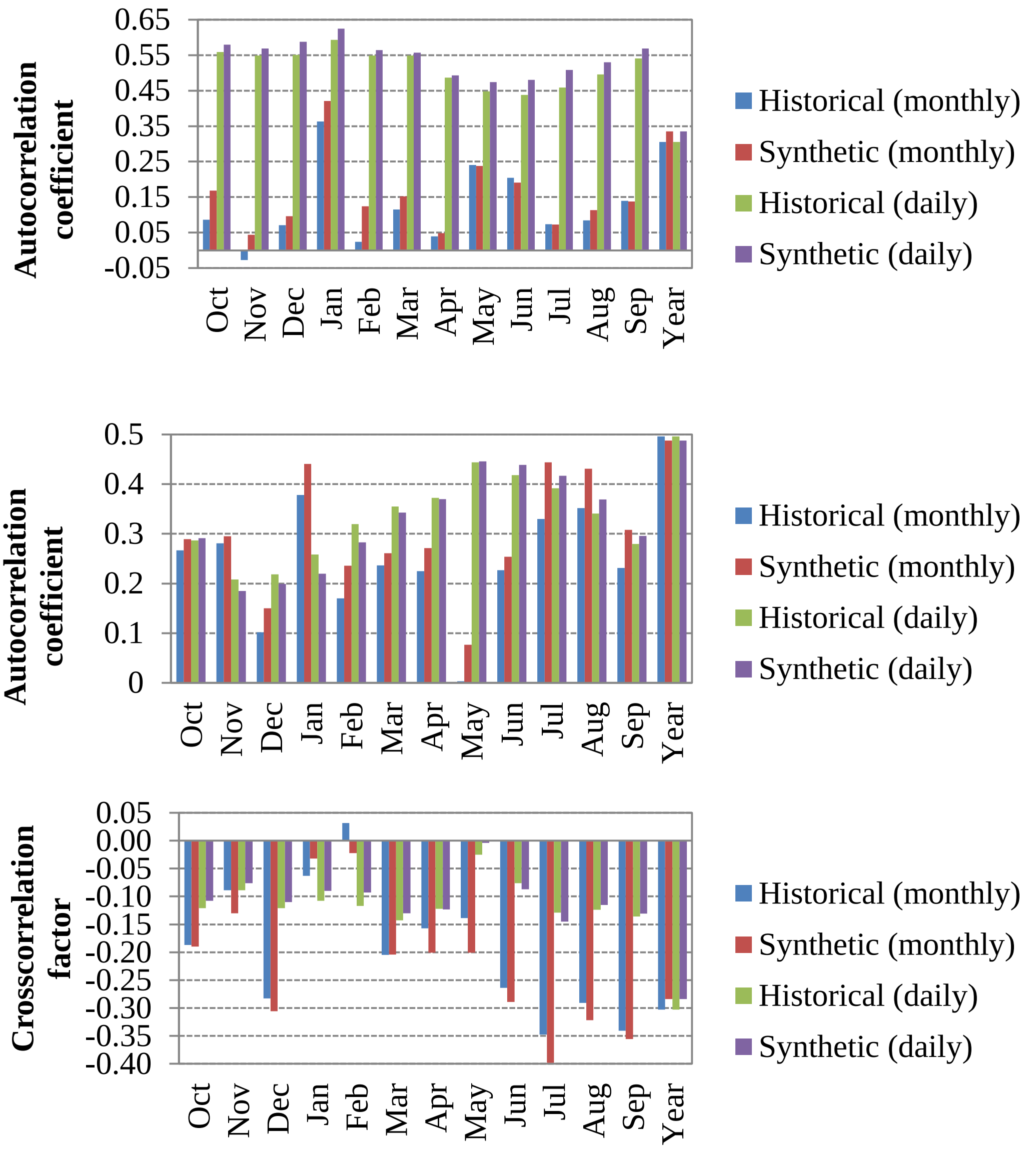


Figure 9: Comparison of annual, monthly and daily auto- and cross-correlation coefficient of historical and synthetic wind speed time series at the Eelde station; first application.

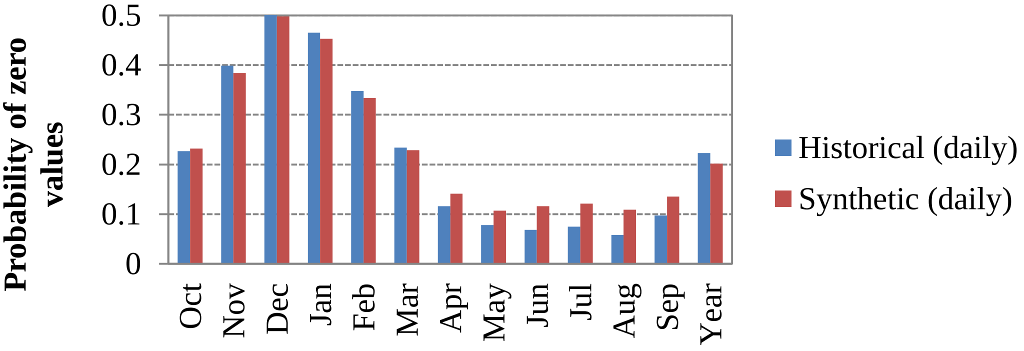


Figure 10: Comparison of the probability of zero sunshine duration of the historical and synthetic time series at the Eelde station; first application.

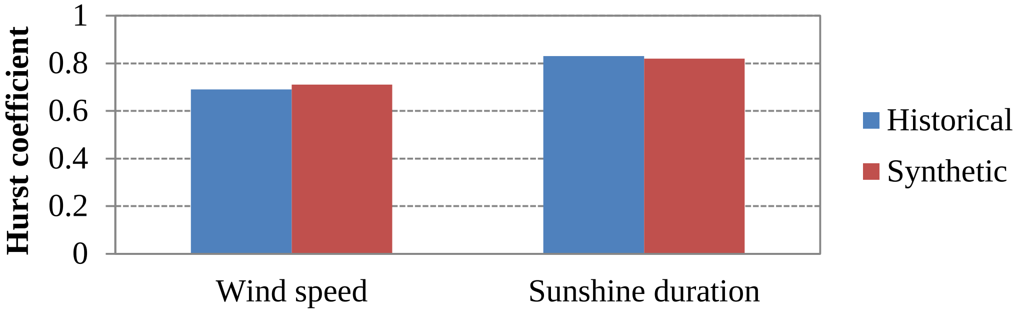


Figure 11: Comparison of Hurst coefficient of historical and synthetic time series at the Eelde station; second application.

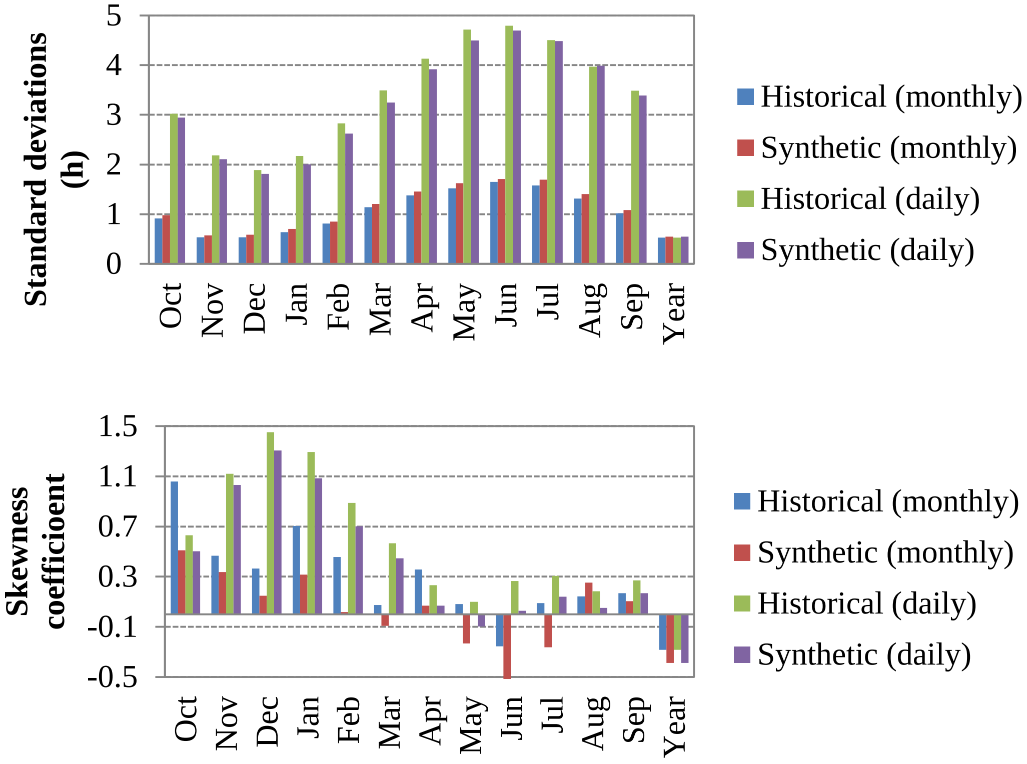
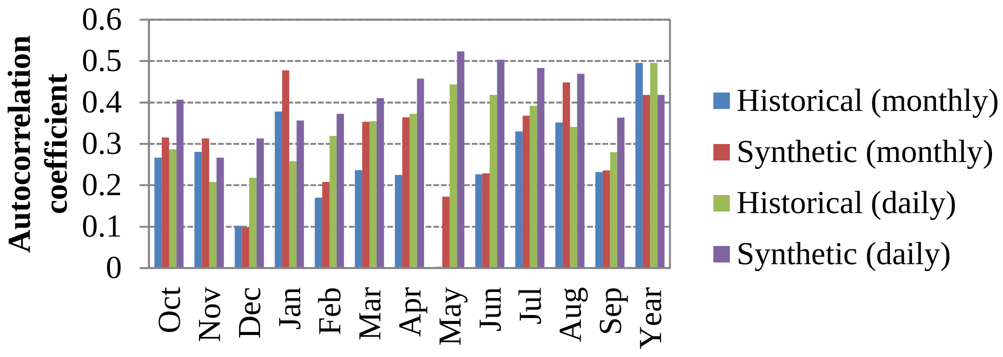


Figure 12: Comparison of annual, monthly and daily standard deviations and skewness coefficients of historical and synthetic sunshine duration time series at the Eelde station; second application.



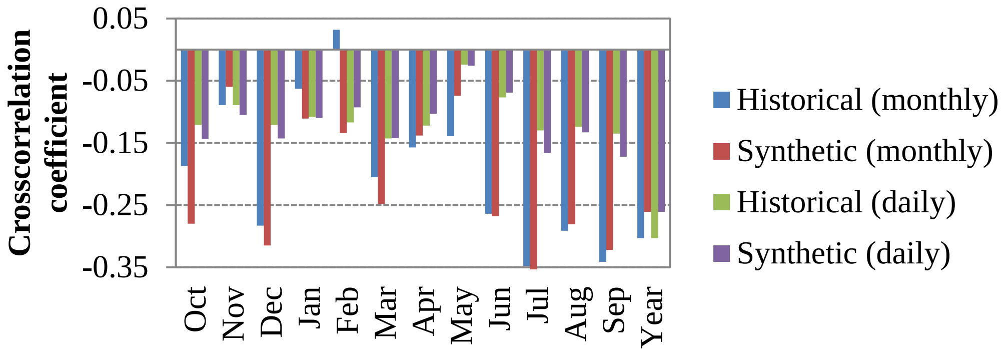


Figure 13: Comparison of annual, monthly and daily auto- and cross-correlation coefficient of historical and synthetic sunshine duration time series at the Eelde station; second application.

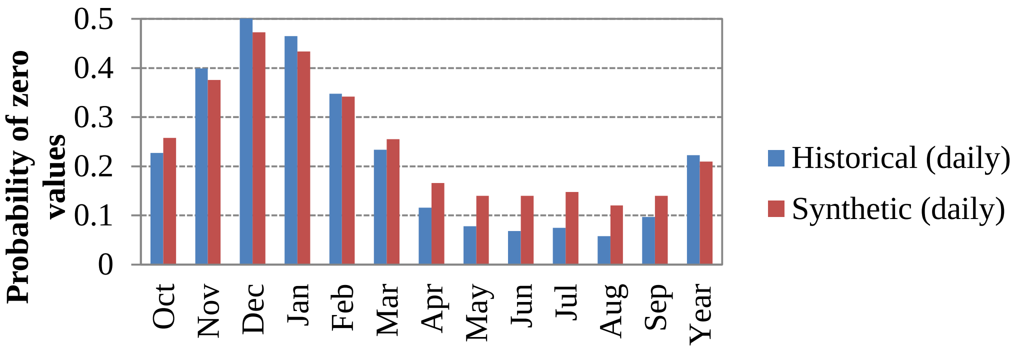


Figure 14: Comparison of the probability of zero sunshine duration of the historical and synthetic time series at the Eelde station; second application.