European Geosciences Union
General Assembly 2013

Vienna, Austria,
07 - 12 April 2013

Session HS7.5/NP8.4 :
Hydroclimatic stochastics

Stochastic analysis and simulation of hydrometeorological processes for optimizing hybrid renewable energy systems

Georgios Tsekouras, Christos Ioannou, Andreas Efstratiadis
and Demetris Koutsoyiannis

Department of Water Resources & Environmental Engineering,
National Technical University of Athens, Greece

Presentation available online: http://itia.ntua.gr/1326/

1. Introduction and motivation

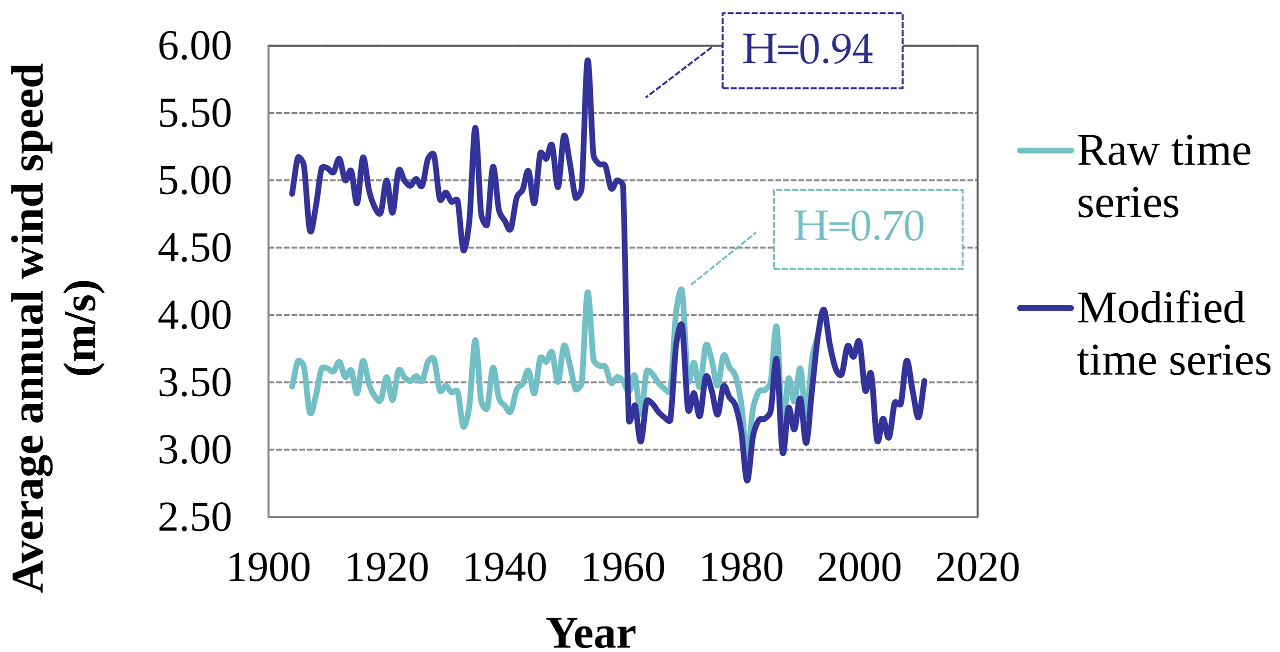
- The drawbacks of **conventional energy sources**, including their negative environmental impacts, emphasize the need to integrate renewable energy sources into energy balance.
- Given that renewables strongly depend on **time varying** and **unpredictable** hydro-meteorological processes, at multiple time scales, it is essential to establish a consistent **stochastic simulation** framework to deal with uncertainty (Koutsoyiannis *et al.*, 2009).
- Yet, the vision of a future scene in which renewable sources dominate will be feasible only if the former are combined with technologies for energy storage, thus formulating **hybrid renewable energy systems** (HRES; Koutsoyiannis and Efstratiadis, 2012).
- In this vein, we investigate the design of a hypothetical hybrid renewable energy system, comprising **wind turbines**, **solar panels**, and a **pumped-storage reservoir** system.
- We investigate the stochastic properties of the two key driving processes, i.e. **wind speed** and **sunshine duration**, on the basis of daily data retrieved from a European database.
- We particularly examine the **long-term persistence** (LTP) of the two variables, seeking representative values of the Hurst coefficient (cf. Koutsoyiannis, 2002, 2003).
- In the simulations we use **synthetic meteorological data** that are generated through the Castalia software, which performs a thee-level (i.e. annual, monthly, daily) multivariate stochastic simulation scheme; we also generate synthetic data of daily energy demand.
- Using these time series as inputs, we estimate key design variables of the system by optimizing the performance expressed in terms of **cost** and **reliability**.

The research of this paper is mainly based on the diploma thesis of the first two authors, elaborated at the National Technical University of Athens (Tsekouras, 2012; Ioannou, 2012).

2. Historical meteorological data

- To investigate the stochastic properties of the two meteorological processes of interest, i.e. wind speed (*W*) and sunshine duration (*S*), we analysed daily records from three European databases (KNMI Climate Explorer, European Climate Assessment & Data, and Deutscher Wetterdienst).
- We retrieved 21 and 20 records of *W* and *S*, respectively, on the basis of three criteria:
  - minimum record length of 70 years (in order to properly account for long-term persistence);
  - existence of wind speed metadata (height above ground of anemometer, for wind observations);
  - up to three changes of anemometer height (stations with too many changes may not be reliable).
- Although we didn't find long enough records of solar radiation, we indirectly estimated this variable, on the basis of sunshine duration.
- Wind speed data were adjusted, in order to refer to the most recent observation height; in this context, the values of each homogenous period were multiplied by the ratio of the last period's average value to each period's average.

The usage of non-homogenous wind speed data may provide inconsistent statistical estimations; particularly, the Hurst coefficient *H* may be significantly overestimated. As shown in the figure, the raw time series at De Bilt station, Netherlands, exhibits a non-realistic value of $H = 0.94$, due to the presence of a sudden shift, indicating a change of observation height. The modified (adjusted to recent mean) data provide an *H* value of 0.70.



3. Investigation of theoretical distributions

- Since both wind speed and sunshine duration are **highly-skewed** variables at the daily time scale, non-normal distributions should be applied for their statistical representation.
- The Gamma and Weibull distributions are almost identical and both suitable to describe the daily wind speed, which is characterized by positive skewness.
- The sunshine duration is bounded, with finite probability at $X = 0$; by employing the logarithmic transformation

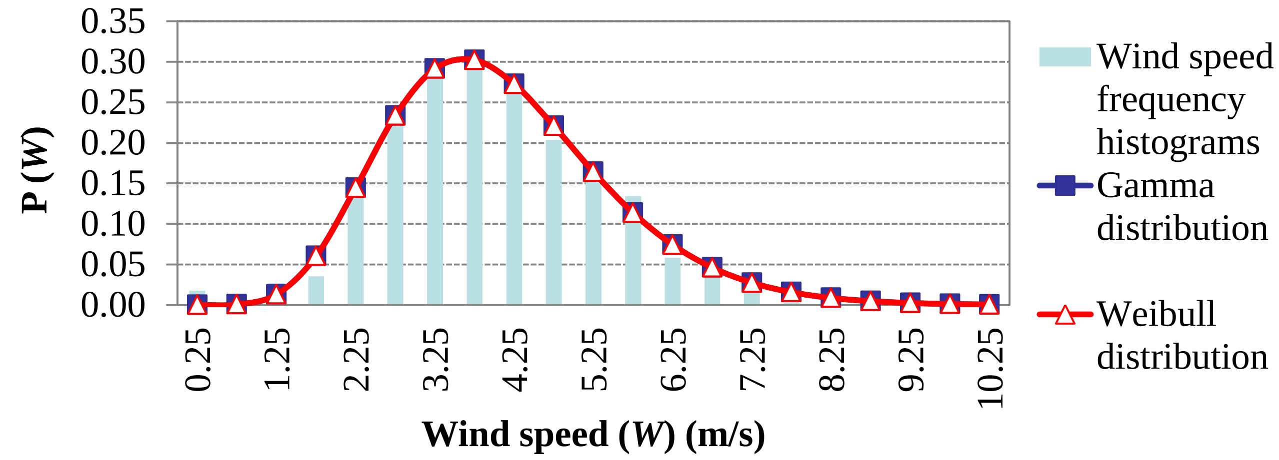
$$Y = -\ln(1 - X)$$

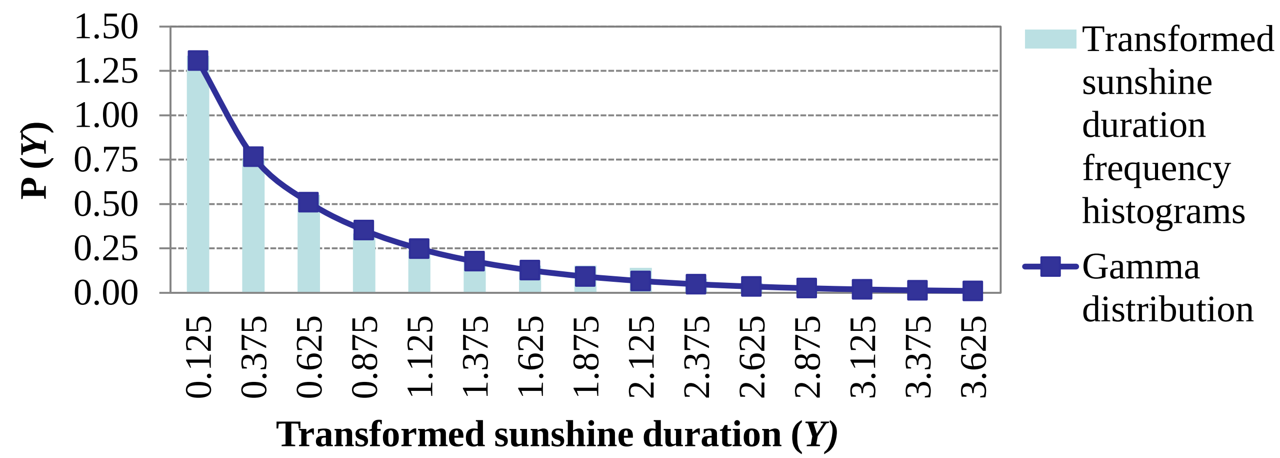
the transformed variable Y is defined in $[0, +\infty)$, while $X = 0$ corresponds to $Y = 0$.
- If Y is Gamma distributed, with shape and scale parameters κ and λ, it is proved that the pdf of X is:

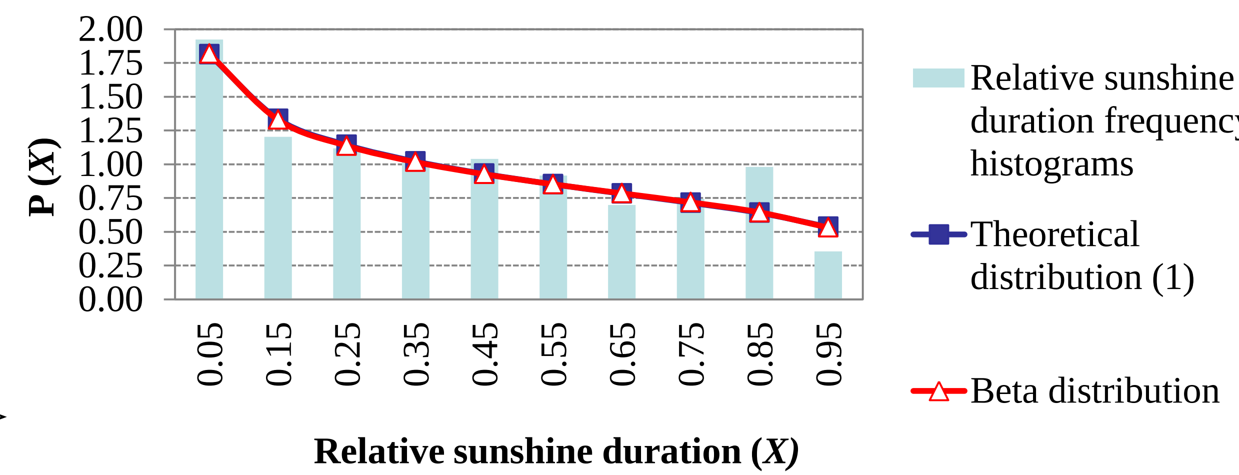
$$f_X(x) = (1 - x)^{\lambda - 1} [-\ln(1 - x)]^{\kappa - 1}$$

- As shown from the histograms, both the theoretical pdf (transformation of Gamma) and the beta distributions fit well the sunshine duration data.

Histograms for June data at Eelde station

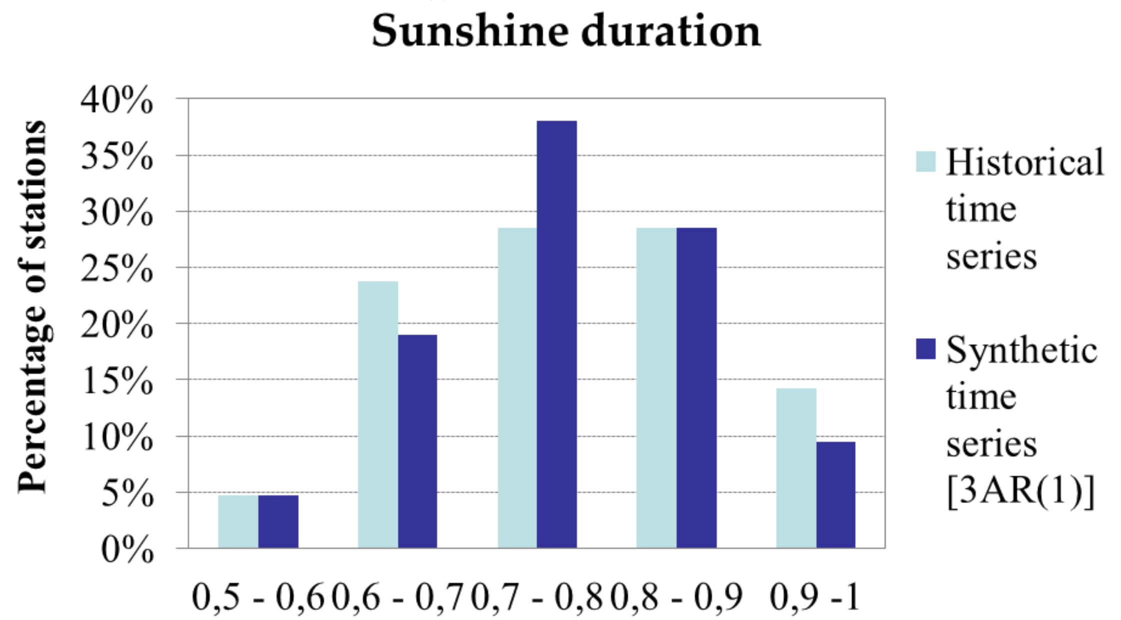


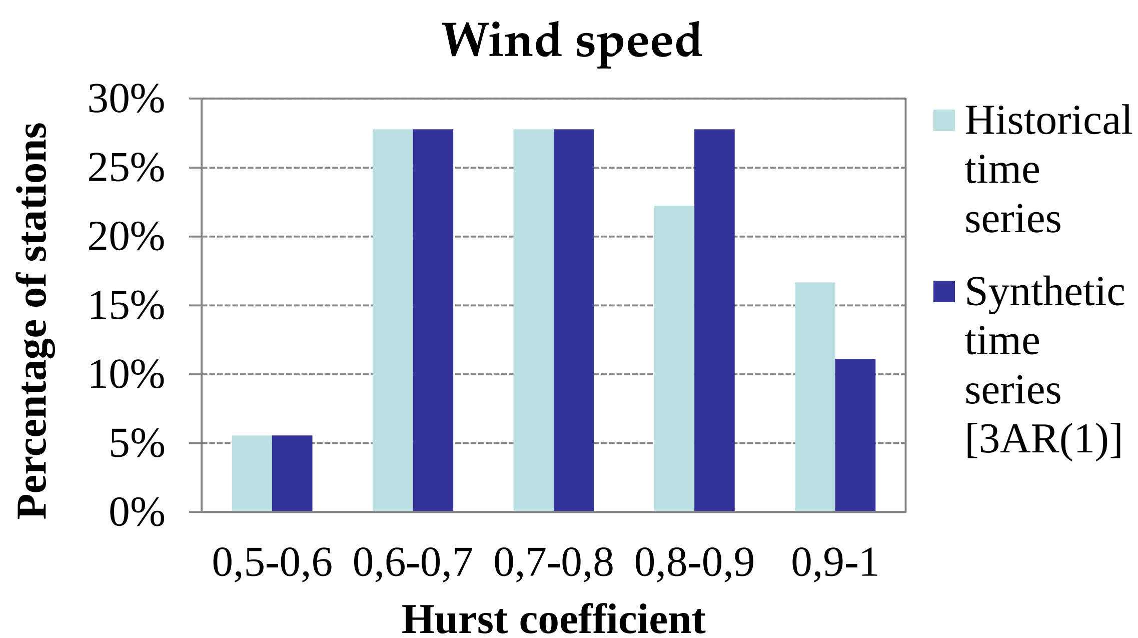




4. Long-term persistence – Hurst coefficient

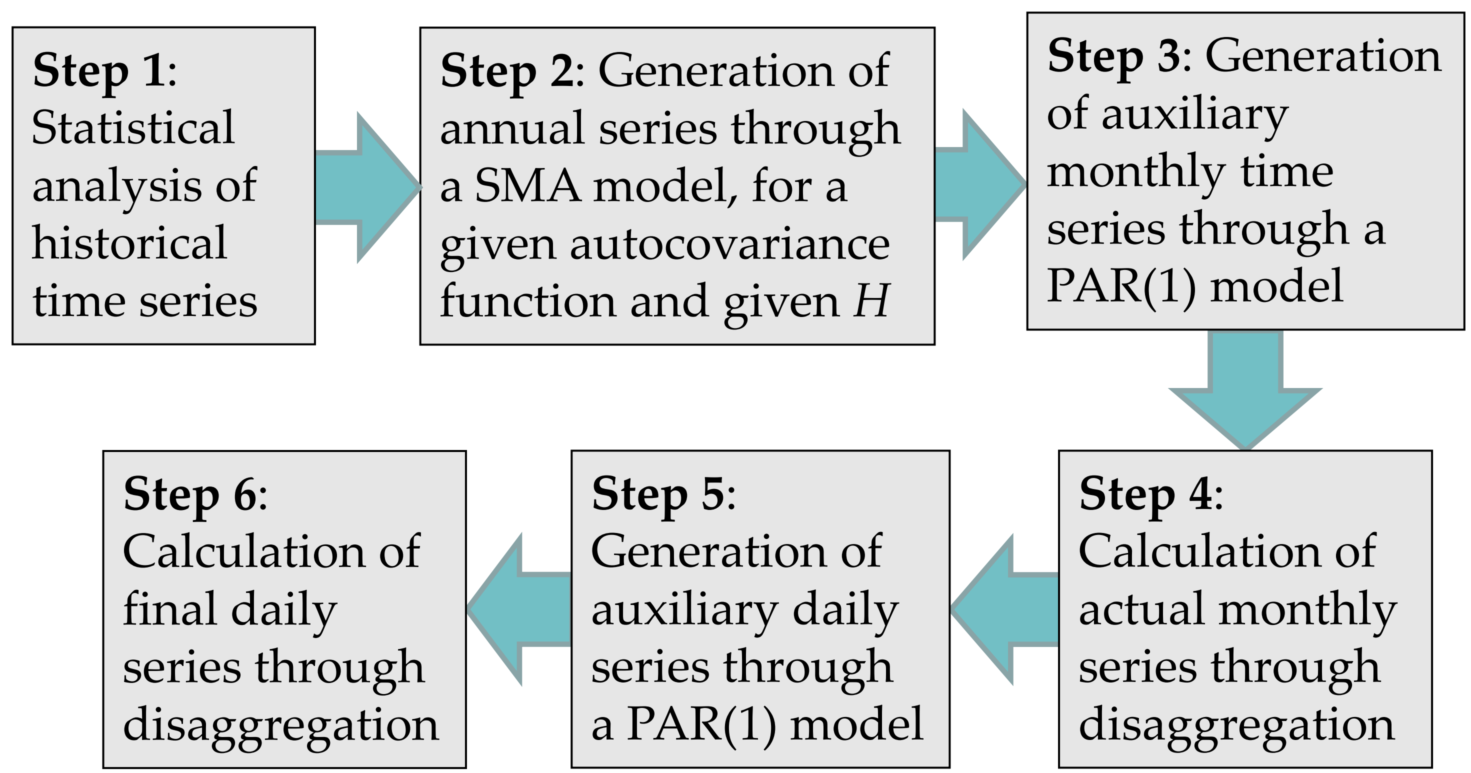
- We investigate the existence of Long-Term Persistence (LTP) on annual data, by estimating the Hurst coefficient *H* via the LSSD method (Koutsoyiannis, 2003).
- For both variables, most of *H* values range from 0.60 to 0.90, while ~15% of them exceed 0.90, which indicates an exceptionally strong self-similar behaviour. Note that in all records $H > 0.50$, and thus none of the time series indicates anti-persistence.
- To find a **representative value of *H***, we generated synthetic samples for all stations, with unique *H* and length, mean and variance same as in the observed records.
- In simulations we used the same *H* value for all stations, through which we produced synthetic data by the multiple time-scale fluctuation method (using 3 AR(1) processes; Koutsoyiannis, 2002).
- We employed a Monte Carlo approach, by examining different values of *H*, until the frequency histogram of the simulated *H* values becomes as close as possible to the histogram of the observed ones.
- On the basis of the above trial-and-error procedure, we concluded that, for both processes, the most representative value of the Hurst coefficient is $H = 0.84$.





5. The software system Castalia

- Castalia is an open-access software that is used for the generation of synthetic daily time series of the two meteorological processes that are inputs to the hypothetical HRES.
- The program employs a **multivariate stochastic simulation** at the daily, monthly and annual time scales, for which it preserves the marginal (mean, standard deviation, skewness) and joint second order statistics (i.e. auto- and cross-correlations).
- It also reproduces the long-term persistence (LTP) at the annual and over-annual scales, the periodicity at the monthly scale, and the intermittency at the daily scale.
- 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).



Flowchart of computational procedures of Castalia

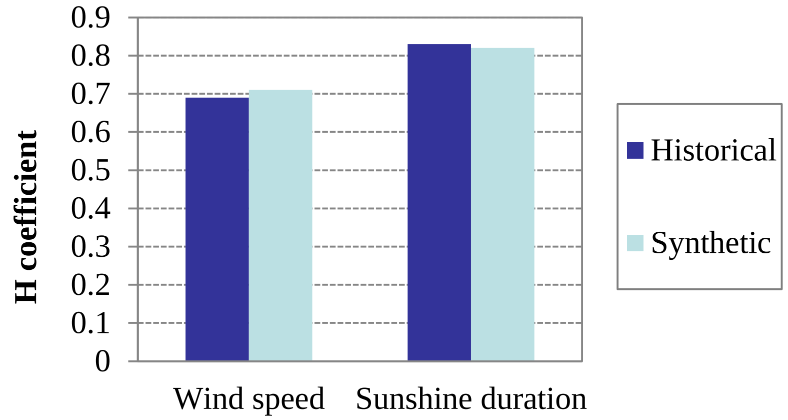
6. Reproduction of intermittent processes within Castalia

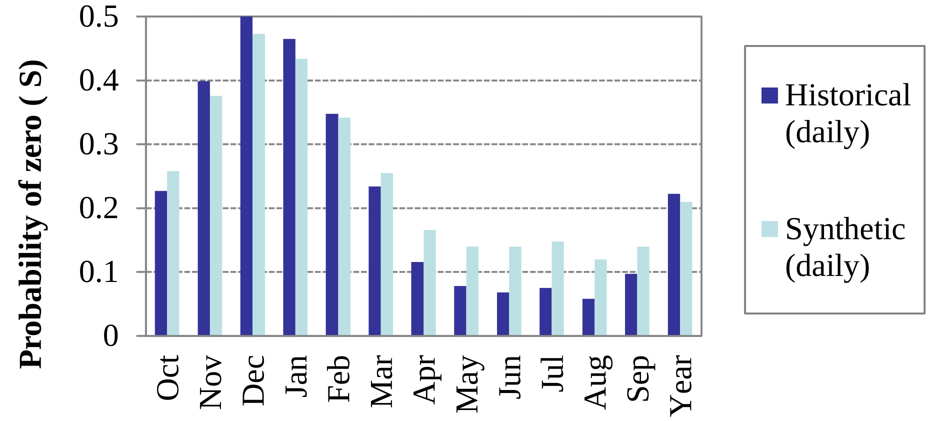
- At the daily time scale, some hydrometeorological processes, among which the sunshine duration, are intermittent as they can take on zero values with finite probability.
- The statistical peculiarities of daily processes, such as the very high values of the coefficients of variability and skewness, are closely linked to intermittency.
- In order to reproduce the intermittent behaviour, it is essential to preserve the **probability of zero values** of the observed time series.
- Castalia handles this problem by introducing the following parameters, which can be also determined by the user (Dialynas *et al.*, 2011):
  - the probabilities k_1 and k_2 of a Markov chain model, which 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;
  - the probability 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;
  - the parameters π_0 and l_0 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 the above procedures allow for generating zero values, and they are all contributing to the frequency of zero values in the final synthetic sample.

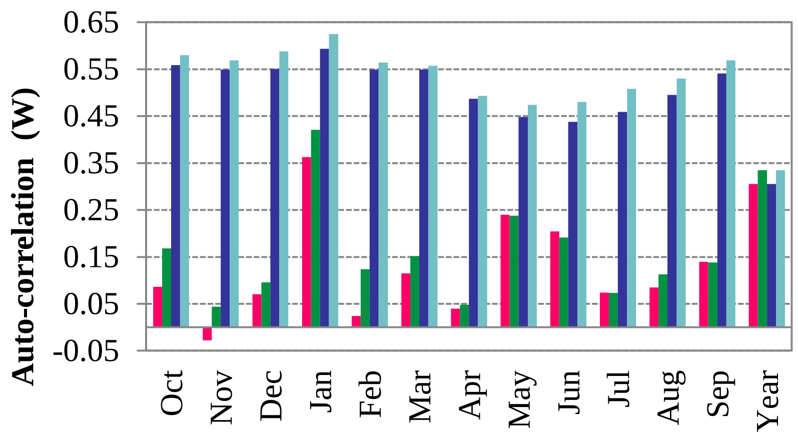
Castalia has been developed by the research team ITIA in the National Technical University of Athens, and it is freely available at http://itia.ntua.gr/en/softinfo/2/

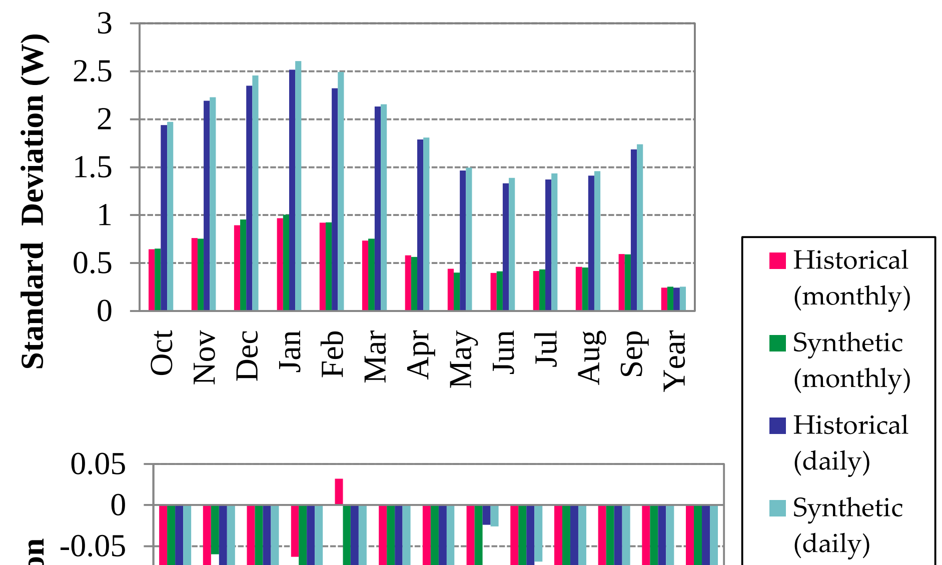
7. Test example

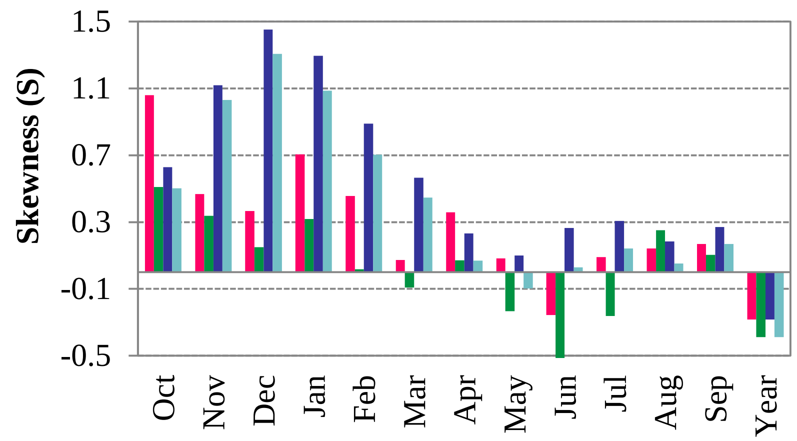
- In order to evaluate the performance of Castalia, we generated 1000 years of daily data in eight stations, where both wind speed and sunshine duration records are available; to the latter we employed the proposed logarithmic transformation, while the final synthetic data of daily sunshine duration were obtained by applying the reverse transformation.
- In the charts some of the key statistical characteristics of the wind speed and sunshine duration data at Eelde station are compared.
- In general, the synthetic time series preserve with satisfactory accuracy the key statistical characteristics of the observed data at all time scales, apart from monthly skewness, which is yet not so critical in daily simulations.

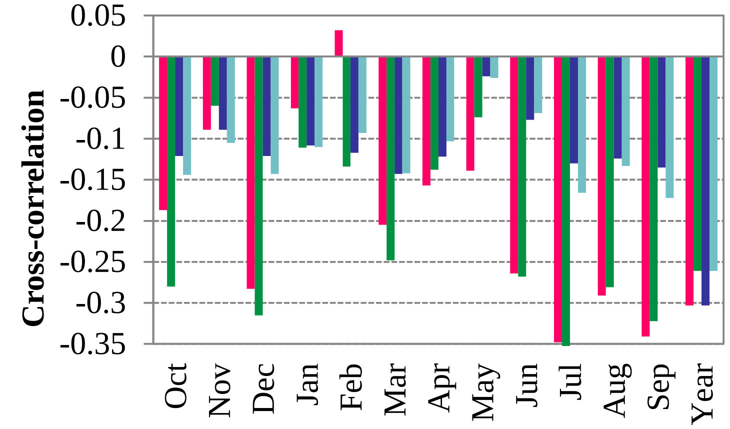






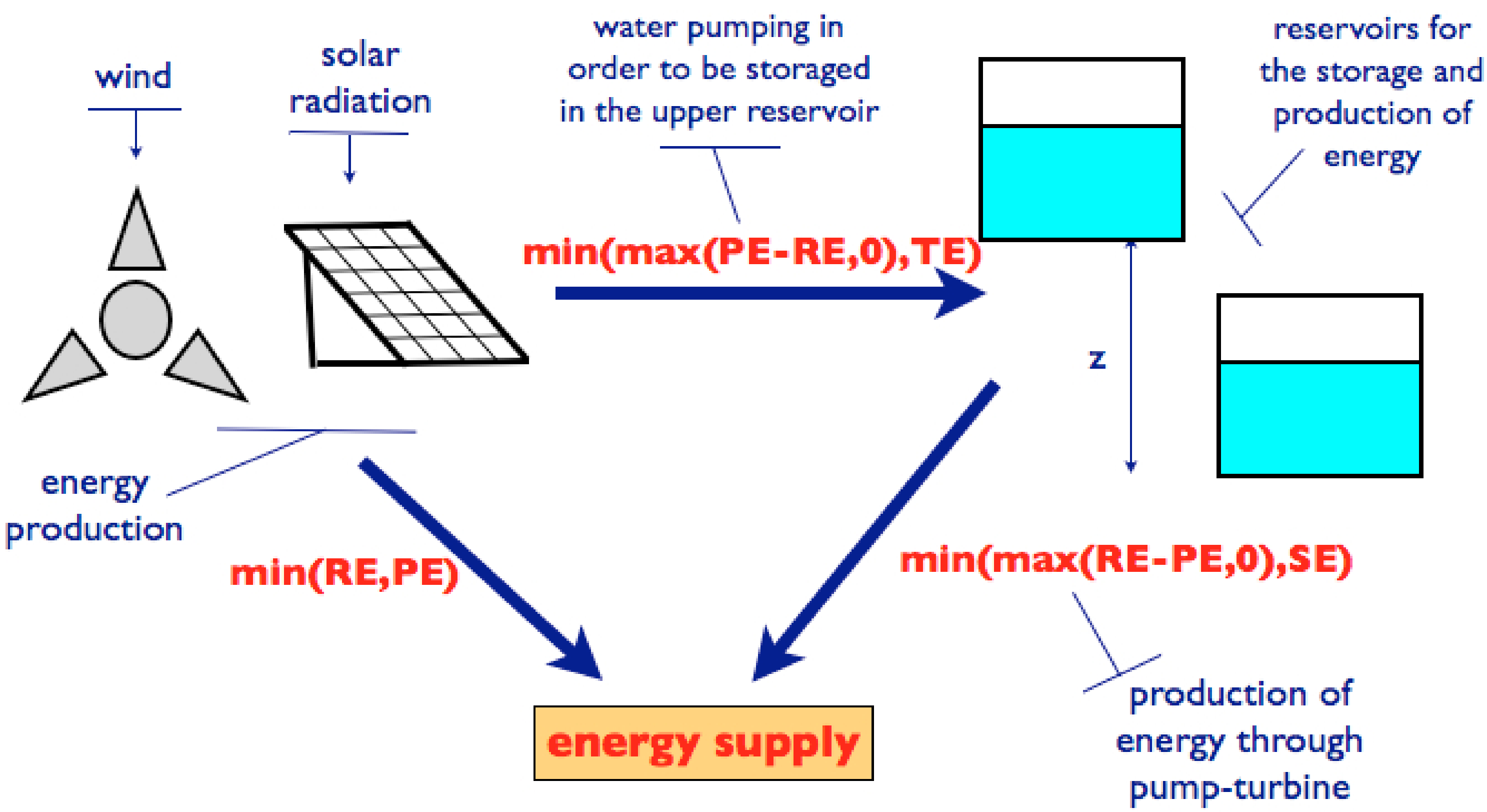






8. Proof-of-concept through a hypothetical case study

- We study a hypothetical, autonomous hybrid renewable energy system, located in Greece and comprising a number of wind turbines and solar panels, as well as two interconnected reservoirs, with pumped-storage facility.
- Input variables are wind speed, sunshine duration and energy demand time series.
- To represent the daily operation of the system for a long-term horizon we follow a stochastic simulation approach, using the Castalia software to generate synthetic time series of wind speed and sunshine duration.
- We set as design variables to optimize the number of wind turbines (n_w), the number of solar panels (n_s), and the capacity of the two equally-sized reservoirs (k).
- Performance criteria are the minimization of the project cost, subject to satisfying the energy demand with very high reliability.



Symbols

***PE**: energy production*

***RE**: energy demand*

***SE**: available hydroelectric energy in the upper reservoir*

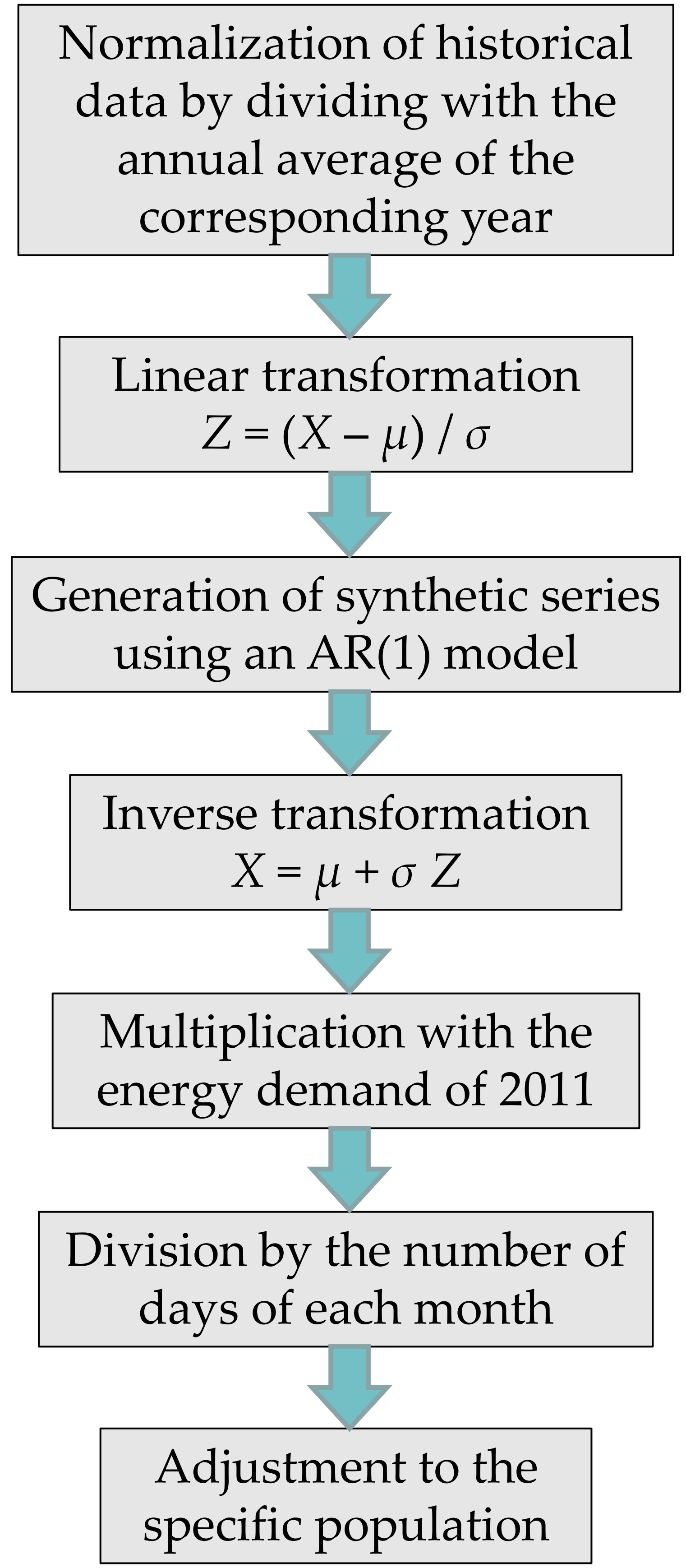
***TE**: theoretical energy for pumping all water stored in the lower reservoir*

9. Generation of synthetic meteorological data

- The objective is to minimize the cost of the hypothetical HRES project, which should operate with very high reliability.
- In this respect, we accept a very low failure probability (i.e. probability of not fulfilling the energy demand), specified as one day per five years, on the average.
- To estimate the aforementioned probability with satisfactory accuracy, we assume a simulation length of 500 years (approximately 182 500 days).
- Meteorological inputs are synthetic time series of daily wind speed and sunshine duration of 500 years length, which were generated by Castalia.
- Historical data for daily wind speed and sunshine duration are obtained from the automatic meteorological station of Agios Kosmas, South Attica (run by the Hydrological Observatory of Athens).
- The historical records extend over a seven year period (Feb. 2005 to June 2012), which is not sufficient for estimating all statistical characteristics. For this reason, some unreasonable values were corrected manually.
- In the theoretical autocovariance function we introduced the representative value of the Hurst coefficient that was found from the analysis of long-term European data ($H = 0.84$).
- The specific value of H which was well-reproduced in the synthetic time series (0.85 for wind speed and 0.83 for sunshine duration).
- In order to assess the influence of LTP, we also generated an alternative set of synthetic data, with significantly lower Hurst coefficients (0.64 for wind speed and 0.61 for sunshine duration).

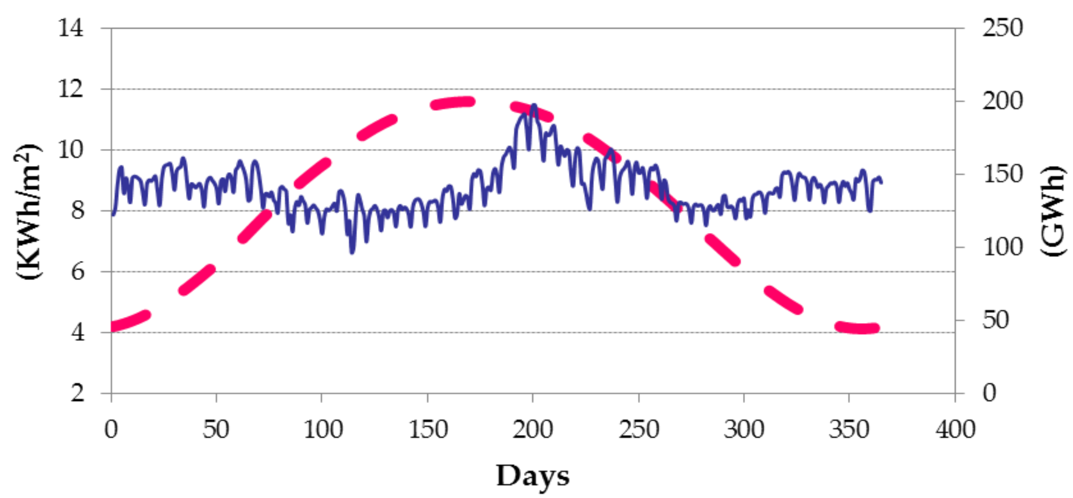
10. Generation of energy demand data

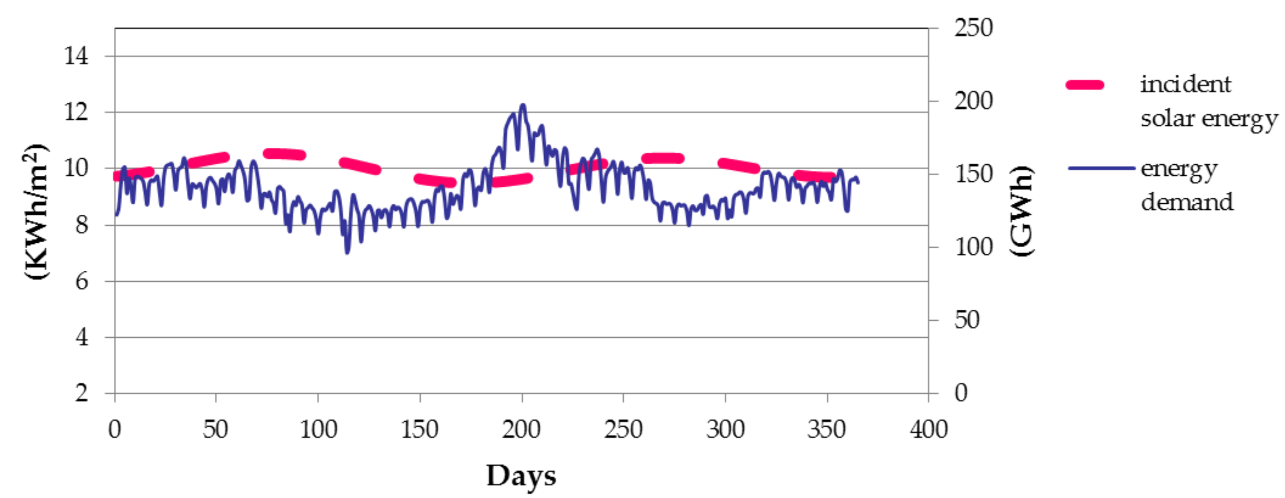
- We assumed a hypothetical area of 100 000 inhabitants, for which we generated synthetic time series of energy demand for the 500-year simulation period.
- The time step of analysis was monthly, since the available daily energy demand data were not found reliable enough.
- We retrieved monthly energy consumption data over the continental Greece from 2004 to 2011, provided by the Independent Power Transmission Operator of Greece.
- In order to remove over-annual trends and seasonality, historical data were normalized, first by dividing with the corresponding annual average value and next by employing the linear transformation $Z = (X - \mu) / \sigma$, where μ and σ are the monthly mean and standard deviation of standardized data.
- For the generation of synthetic normalized data we used an AR(1) model; next we applied the inverse transformation to obtain the monthly values of energy demand, taking as standard the annual energy demand of year 2011.
- Daily data were calculated by dividing by the number of days of each month (this assumption, albeit unrealistic, was deemed consistent with the purpose of the study).
- The final estimates were adjusted to the specific population size (100 000).



11. Calculation of energy production/consumption

- For the pumped-storage system, we assumed two equally-sized reservoirs, which is control variable of the optimization problem. The reservoir system was assumed to be closed, without inflows and outflows.
- The hydroelectric energy production and consumption (due to pumping) depend on the hydraulic head (difference of reservoir levels), the discharge and the efficiency ($n = 0.85$).
- The other two system components are a hypothetical photovoltaic panel of 240 W per unit for 1 000 W/m^2 irradiance and a commercial wind turbine (7.5 MW; type E-126 by Enercon).
- For the given system, the solar energy production is function of the panel angle and the incoming solar radiation, while the wind production is function of the power curve of the turbine and wind velocity data.
- The solar radiation is function of latitude and daily sunshine duration data. Given that different panel angles match to different temporal distributions of solar energy, which in turn affects the size of reservoirs, we investigated alternative angle values.

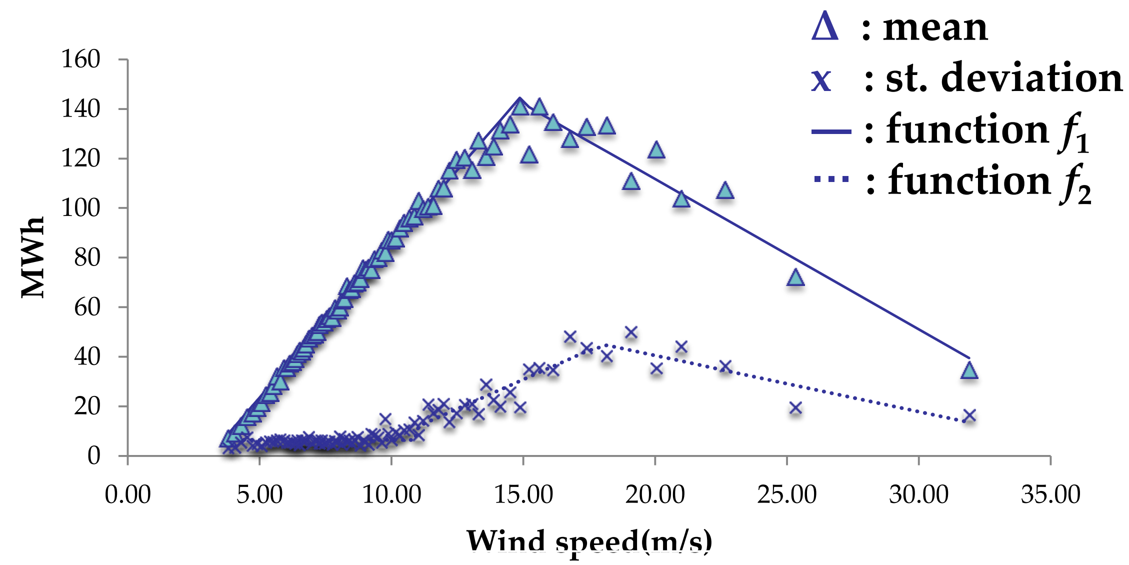


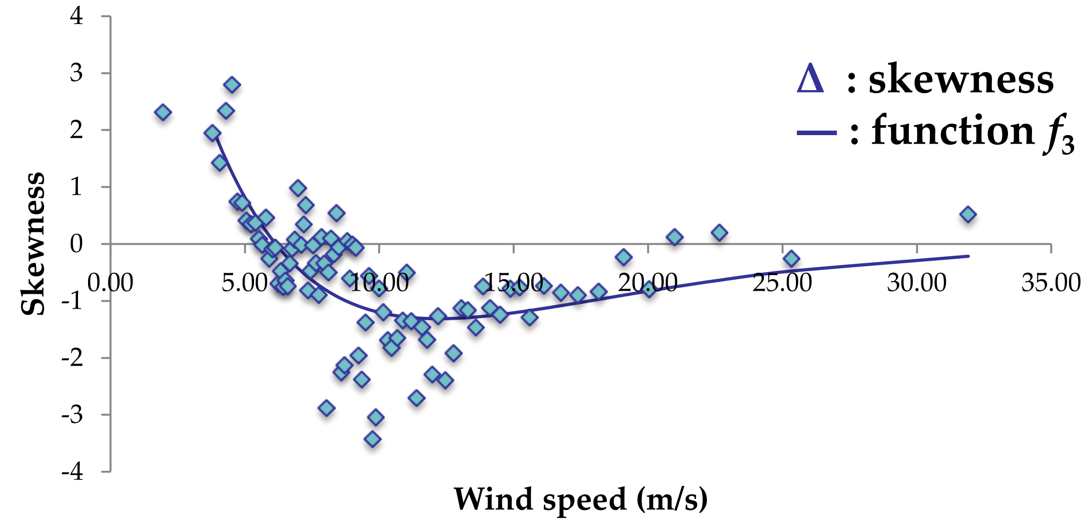


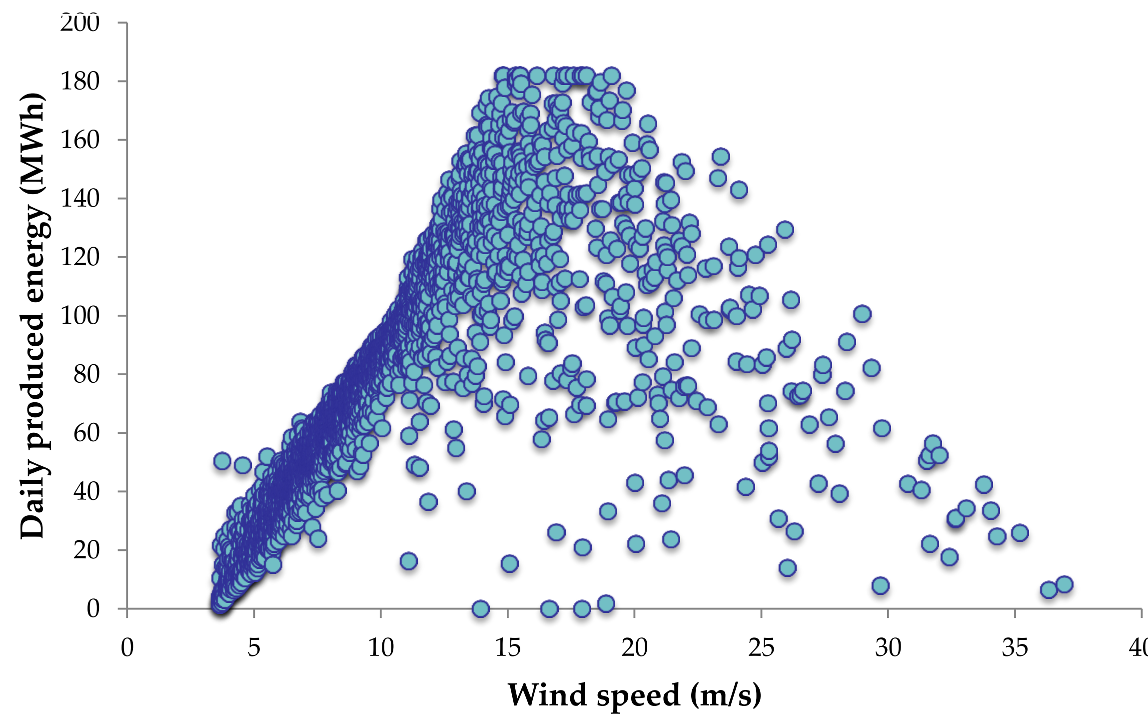
Energy demand of Greece vs. incident solar energy at Agios Kosmas for panel angles 0º (left) and 38º (right)

12. Accounting for hourly distribution of wind energy

- Using daily averaged wind data, we neglect the hourly fluctuations of wind speed which result to a significant variability of the produced energy.
- Thus, we employed a stochastic approach, in which we reproduced the statistical characteristics of hourly energy, on the basis of hourly wind speed at Agios Kosmas station.
- The daily energy calculated from hourly data was classified into 87 classes of daily wind speed W, and for each class we calculated the mean μ, standard deviation σ, and skewness, ξ.
- For each statistical characteristic we fitted an empirical function to the corresponding sample of 87 values, thus $\mu = f_1(W)$, $\sigma = f_2(W)$, and $\xi = f_3(W)$.
- Given the synthetic daily wind speed data, we estimated the daily production through a Gamma distribution, with random probability and parameter values depending on W.







13. Optimization model and results

- For each set of synthetic data (i.e. with high and low H values), we formulated a nonlinear optimization problem, where the objective is the minimization of the annual cost of the system, for the specified failure probability (i.e. 100 days out of 500 simulated years).
- The total cost depends on the number of wind turbines n_w, the number of solar panels n_s, and the reservoir capacity k, which are the design variables of the optimization problem.
- This comprises initial investment as well as maintenance costs for the mechanical equipment (turbines, panels, pumps); the economic time life of reservoirs was taken 50 years, while for the rest of components the time life was assumed 25 years.
- Optimization was carried out using evolutionary algorithms.
- For both scenarios, the use of solar panels provided suboptimal solutions and thus the optimal design value was $n_s = 0$ (because they are much less efficient than wind turbines); the optimal number of wind turbines remained the same in both scenarios ($n_w = 23$).
- The size of the pumped-storage system, in terms of reservoir capacity, is strongly affected by the long-term persistence of the driving meteorological processes; yet, the differences in the total cost are rather small.

	Hurst values 0.83-0.85	**Hurst values 0.61-0.64**
Number of wind turbines, n_w	23	23
Number of solar panels, n_s	0	0
Capacity of reservoirs, k (hm^3)	250	195
Annual cost (million €)	33.6	32.8

14. Conclusions & Summary

- The suitability of Gamma and Weibull distributions for the representation of wind speed is confirmed. Both the proposed distribution function and the Beta distribution fit satisfactorily to the relative sunshine duration data.
- Both processes are characterized by LTP. A unique value $H = 0.84$ is found to be representative for both variables.
- Due to the relatively small available amount of samples globally, having sufficient record length, further investigation of the stochastic structure of both processes is recommended.
- The software system Castalia performs multivariate stochastic simulation of both processes on annual, monthly and daily scale satisfactorily, preserving the marginal, the joint second order statistics and the LTP.
- The proposed HRES transforms the two natural processes, which are characterized by uncertainty and unpredictability, into regular energy outflows that satisfy the energy demand at multiple time scales.
- The presence of LTP should not be neglected in the design of a HRES as this may lead to the incorrect estimation of design variables.
- Further investigation is required in order to detect the influence of the design variables, such as type of wind turbines and solar panels, and the simulation time scale of meteorological variables.

References

- 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, *EGU General Assembly 2011, Geophysical Research Abstracts, Vol. 13*, Vienna, EGU2011-290, European Geosciences Union, 2011.
- Ioannou, C., *Stochastic simulation and optimization of a hybrid renewable energy* system, Diploma thesis, Department of Water Resources & Environmental Engineering – National Technical University of Athens, October 2012 (in Greek; http://itia.ntua.gr/en/docinfo/1301/).
- Koutsoyiannis, D., and A. Efstratiadis, The necessity for large-scale hybrid renewable energy systems, Hydrology and Society, *EGU Leonardo Topical Conference Series on the Hydrological Cycle 2012*, Torino, 2012.
- Koutsoyiannis, D., A generalized mathematical framework for stochastic simulation and forecast of hydrologic time series, *Wat. Resour. Res.*, 36(6), 1519–1533, 2000.
- Koutsoyiannis D, Climate change, the Hurst phenomenon, and hydrological statistics, *Hydrol. Sci. J.*, 48(1), 3-24, 2003.
- Koutsoyiannis, D., C. Makropoulos, A. Langousis, S. Baki, A. Efstratiadis, A. Christofides, G. Karavokiros, and N. Mamassis, Climate, hydrology, energy, water: recognizing uncertainty and seeking sustainability, *Hydrol. Earth Sys. Sci.*, 13, 247–257, 2009.
- Koutsoyiannis, D., Coupling stochastic models of different time scales, *Wat. Resour. Res.*, 37(2), 379–392, 2001.
- Koutsoyiannis, D., Optimal decomposition of covariance matrices for multivariate stochastic models in hydrology, *Wat. Resour. Res.*, 35(4), 1219–1229, 1999.
- Koutsoyiannis D., The Hurst phenomenon and fractional Gaussian noise made easy, *Hydrol. Sci. J.*, 47(4), 573-595, 2002.
- Tsekouras, G., *Stochastic analysis and simulation of hydrometeorological processes associated with wind and solar energy*, Diploma thesis, Department of Water Resources & Environmental Engineering – National Technical University of Athens, October 2012 (in Greek; http://itia.ntua.gr/1300/).

Databases

- Deutscher Wetterdienst (http://www.dwd.de)
- European Climate Assessment & Dataset (http://eca.knmi.nl/)
- Hydrological Observatory of Athens (http://hoa.ntua.gr/)
- Independent Power Transmission Operator of Greece (http://www.admie.gr/nc/en/home/)
- KNMI Climate Explorer (http://climexp.knmi.nl/)