NATIONAL TECHNICAL UNIVERSITY OF ATHENS
SCHOOL OF CIVIL ENGINEERING
DEPARTMENT OF WATER RESOURCES & ENVIRONMENTAL ENGINEERING

INTERDEPARTMENTAL PROGRAM OF POSTGRADUATE STUDIES
WATER RESOURCES SCIENCE & TECHNOLOGY

Application of a generic simulation method for demanding stochastic processes to hydrometeorological variables for the quantification of renewable energy resources

MSc Thesis

Panagiotis Andreas Valakos
Thesis Supervisor: Emeritus Professor Demetris Koutsoyiannis

Acknowledgments

First of all, I would like to thank Emeritus Professor Demetris Koutsoyiannis, supervisor of this thesis, for the assignment of this project and the opportunity he gave me to work under his supervision. I would also like to thank Panayiotis Dimitriadis, for his continuous contribution as well as the warm support he provided throughout the preparation of this project.
Finally, I want to thank my parents and Theo for their continuous support and patience over the years. This degree would not have been possible without the support of Theo.

Περίληψη

Οι ανανεώσιμες πηγές ενέργειας προσφέρουν λύση σε περιπτώσεις νησιών τα οποία προσπαθούν να είναι ενεργειακά αυτόνομα, ωστόσο η στοχαστική φύση των καιρικών φαινομένων δημιουργεί προκλήσεις για την αξιόπιστη ποσοτικοποίηση των ανανεώσιμων πηγών ενέργειας. Η εργασία εστιάζεται στην περίπτωση του νησιού της Αστυπάλαιας το οποίο βρίσκεται στο Αιγαίο πέλαγος ενώ αναδεικνύεται η σημασία της στοχαστικής προσομοίωσης για την ποσοτικοποίηση των ανανεώσιμων πηγών ενέργειας. Η Αστυπάλαια, όπως και πολλά νησιά, είναι εξαρτημένα από τα εισαγόμενα ορυκτά καύσιμα. Η παρούσα εργασία πραγματεύεται αυτά τα ζητήματα εφαρμόζοντας ένα μοντέλο στοχαστικής προσομοίωσης που αναπτύχθηκε από (Koutsoyiannis, 2021, 2000) και επεκτάθηκε από (Dimitriadis and Koutsoyiannis, 2018) σε διαφορετικά είδη δεδομένων, όπως ταχύτητα ανέμου, ηλιακή ακτινοβολία και ύψος κύματος τα οποία αντιστοιχούν στο νησί της Αστυπάλαιας. Στην εργασία χρησιμοποιήθηκε το μοντέλο Συμμετρικού Κινούμενου Μέσου σε συνδυασμό με ανέλιξη Hurst-Kolmogorov που περιγράφεται από το κλιμακόγραμμα της, ενώ η προσέγγιση αυτή μπορεί να αναπαράγει τη βραχυπρόθεσμη και τη μακροπρόθεσμη εμμονή των δεδομένων. Το μοντέλο μπορεί να διατηρεί τις στατιστικές ροπές της διαδικασίας και όπως αποδείχθηκε μέσω της διατήρησης του συντελεστή κύρτωσης να προσομοιώνει τις διακοπτόμενες διεργασίες. Τα αποτελέσματα αυτά μπορούν να βοηθήσουν στην διαδικασία λήψης αποφάσεων και να καθοδηγήσουν στη βέλτιστη χρήση ανανεώσιμων πηγών ενέργειας.

Abstract

Renewable energy offers solution for island communities who try to be energy independent, yet the stochastic nature of weather phenomena creates challenges for reliable energy generation. The significance of stochastic simulation in renewable energy planning for islands is highlighted, focusing in the case study of Astypalea island in the Aegean Sea. Astypalea, like many islands, faces energy insecurities due to its reliance on imported fossil fuels. This thesis addresses this issue implementing a stochastic simulation model developed by (Koutsoyiannis, 2021, 2000) and (Dimitriadis and Koutsoyiannis, 2018) in various types of datasets, such as, wind, solar irradiance, wave height and period corresponding to the island of Astypalea. In this thesis the Symmetric Moving Average scheme was used for the generation of the simulated values combined with Hurst-Kolmogorov process described by its climacogram. The extended Symmetric Moving Average Model as will be introduced can preserve high order moments of a process and as proved by preserving the coefficient of kurtosis can capture the intermittency of hydrometeorological processes. With this methodology we can expect insights for optimizing renewable energy systems in island contexts, ensuring sustainable and reliable energy generation.

Table of Contents

List of Tables

List of Figures

1. Introduction

In a period marked by increasing environmental concerns and the urgent need for sustainable energy solutions, the role of renewable energy resources has become paramount. However for island communities achieving energy independence presents some challenges due to their geographical isolation, limited resources and limited land cover. In this context, stochastic simulation of hydrometeorological data emerges as a tool for assessing and optimizing renewable energy systems, creating a path for autonomous and resilient energy solutions. Renewable energy resources, such as wind, solar and wave power hold particular promise for island communities due to their abundance and sustainability. Unlike finite fossil fuel reserves, renewable resources are inexhaustible and can be harnessed locally, reducing the need for costly imports and strengthening local economies.

Moreover, advancements in renewable energy technologies, coupled with declining costs, have made it increasingly feasible for islands to pursue autonomous energy solutions tailored to their needs and conditions. Situated in the southern Aegean Sea, the island of Astypalea is a representation of the challenges and opportunities for energy independence. With a population of approximately 1300 inhabitants spread across an area of 97 km^2, Astypalea relies on diesel generators for electricity generation resulting in high energy costs. However the islands location and favorable climatic conditions make it an ideal candidate for renewable energy deployment, offering opportunities for wind, solar and wave energy harvesting. With the theory of stochastic simulation we try to quantify the uncertainty associated with the renewable energy resources to ensure reliability and efficiency. Stochastic simulation involves generating synthetic time series of meteorological variables based on statistical models derived from historical data. It also provides valuable insights into the temporal and spatial variability of energy generation supporting effective grid management.

We implement a stochastic simulation model developed by (Koutsoyiannis, 2021, 2000) and extended by (Dimitriadis and Koutsoyiannis, 2018) for three different dataset types, wind speed, solar irradiance, wave height and period of Astypalea's island. The aim of this project is to introduce and apply a methodology (Dimitriadis and Koutsoyiannis, 2018; Koutsoyiannis, 2000) that can preserve any arbitrary second order structure, as well as the high order moments of a process. With that said we can produce synthetic datasets from historic time series, having the same dependence structure and approximately having the same first four central moments, meaning, we approximate the distribution of the data to a desired degree which is sufficient for our applications.

These processes are governed by Hurst-Kolmogorov dynamics (Koutsoyiannis, 2016) and are described by their climacogram, which is, the variance of the averaged process versus its scale. The climacogram of an HK process is described by:

$$\gamma(k) = \frac{\gamma(D)}{\kappa^{2-2H}}$$

Where $\gamma(k)$ is the variance of the process at time scale k, $\gamma(D)$ is the variance at the unit time scale, H is the Hurst parameter and $\kappa = k/D$ is the discrete time scale. For the generation of the synthetic time series we use the Symmetric Moving Average scheme developed by (Koutsoyiannis, 2000) described by:

$$X_i = \sum_{j=-l}^{l} a_{|j|} V_{i+j}$$

where a_j are coefficients which can by calculated analytically and v_i is white noise error terms. (Koutsoyiannis, 2000) showed that the SMA scheme can preserve the first three central moments, while (Dimitriadis and Koutsoyiannis, 2018) extended this method to the fourth central moment and found that it was sufficient for various distributions used in geophysical processes.
In chapter 2 we present the theoretical background needed for the study of stochastic hydrology e.g. probability theory, random variables, moments of distribution, stochastic processes, statistics of stochastic processes, time series methods and spectral analysis of time series.
In chapter 3 we present the theory of the methodology which was used, including the Symmetric Moving Average scheme, the Hurst-Kolmogorov process, the climacogram, its advantages and its variations and the extension of the Symmetric Moving Average scheme to the fourth moment.
In chapter 4 we talk about the case study, the island of Astypalea, we present how various types of renewable energy resources can be harvested and we present the climacograms and the statistical characteristics of the historical and synthetic time series.
In chapter 5 we point all the conclusions from the statistical comparison of the historical with the simulated time series.

2. Theoretical Background

2.1 Hurst Phenomenon

In the 1950s while H.E. Hurst was trying to solve an engineering problem concerning the design capacity of reservoirs (Hurst, 1951) (e.g. the storage required on a stream to give a minimum discharge), using statistics, probability theory and long period time series of natural events, observed a statistical property that describes the behavior of time series, that is, the tendency in natural phenomena, to occur clustering of high or low values which was later called Hurst Phenomenon. Hurst found that in natural events, groups of high or low values occur more often than random events, while investigating a time series of discharge of the Nile river and noticed that there are discharge "stretches" when there is a flood event or a drought period. This work was based on range analysis and various time series, like river flows, precipitation, lake levels, temperatures, growth of tree rings and sunspot numbers. Given a discharge time series and by calculating the accumulated mean and the accumulated departures from the mean, we can find the adjusted range R_n which is the highest and lowest of the accumulated totals. This is also the size of a reservoir, required to maintain the maximum possible steady discharge. Hurst's goal was to establish a relationship between R_n, variability S_n and the number of observations n assuming that the observations are normally distributed. Hurst proved the relationship $\frac{R_n}{S_n} \sim n^H$ where R_n is the adjusted range, S_n is the standard deviation, n is the size of the time series, H is the Hurst coefficient and R_n/S_n is the rescaled range. He proceeded using probability theory and a coin tossing experiment and derived a theoretical relationship of a process with normal distribution with equation:

$$\frac{R_n}{\sigma} = 1.25n^{0.5} \tag{2.1}$$

After establishing Eq. (2.1), he focused on the available data and investigated the range on these time series. The longest period used was 2000 years, while there was not a period of less than 30 years. Hurst plotted each data group separately in a log-log plot and showed that $\log(R/\sigma)$ is a linear function of $\log N$ and is expressed as:

$$\log\frac{R}{\sigma} = K\log\frac{N}{2} \; or \; K = \frac{\log(R/\sigma)}{\log(N/2)} \tag{2.2}$$

Where K is the slope of the different lines, which varied from 0.69 to 0.80 and the mean was 0.729. Hurst suggested for the case of a long series (several hundred years) the best practice was to take the mean value of K while for shorter series take the mean value of K from all available material. The general equation representing all the different lines is:

$$\frac{R_n}{S_n} = \left(\frac{N}{2}\right)^K \tag{2.3}$$

It is worth noting that in the case of natural events where $K > 0.5$ the term R_n/S_n increases faster as the data size increases, rather than random events where $K = 0.5$.

The first mathematical description of this phenomenon first appeared in the work of (Mandelbrot and Van Ness, 1968) where Mandelbrot and Van Ness introduced Fractional Brownian Motion with equation:

$$\begin{aligned} B_H(t) - B_H(0) \\ = \frac{1}{\Gamma(H+0.5)}\left\{\int_{-\infty}^{0}[(t-u)^{H-0.5} - (-u)^{H-0.5}]dB(u) \right. \\ \left. + \int_0^t (t-u)^{H-0.5} dB(u)\right\} \end{aligned} \tag{2.4}$$

Where $B_H(0)$ is the starting value, H is the Hurst parameter and $dB(u)$ is white noise. It's a gaussian process, self-similar, non stationary process whose increments are stationary $X(t) = B_H(t+s) - B_H(s)$. The variance of the fBm is defined as:

$$E[B_H(t+T,\omega) - B_H(t,\omega)]^2 = Var[(B_H(T) - B_H(0)] = V_H T^{2H} \tag{2.5}$$

Where V_H is the variance of the unit increments. However the process $B_H(t)$ is not differentiable and that led them to introduce Fractional Gaussian Noise by "smoothing" B_H (making it differentiable everywhere, hence continuous) with equation:

$$B_H(t,\delta) = \frac{1}{\delta}\int_t^{t+\delta} B_H(s)ds \tag{2.6}$$

The autocovariance of the process is:

$$C_H(\tau,\delta) = \frac{1}{2}V_H\delta^{2H-2}\left[\left(\frac{|\tau|}{\delta}+1\right)^{2H} - 2\left|\frac{\tau}{\delta}\right|^{2H} + \left|\frac{|\tau|}{\delta}-1\right|^{2H}\right] \tag{2.7}$$

With finite variance:

$$C_H(0,\delta) = V_H\delta^{2H-2} \tag{2.8}$$

For $0.5 < H < 1$ (the range of values of hydrologic time series), the covariance function does not converge which means:

$$\int_0^\infty C_H(\tau,\delta)\, d\tau = \infty \tag{2.9}$$

In a series of computational experiments (Mandelbrot and Wallis, 1969a, 1969b, 1969c), discrete fGn was introduced by using integer values of t:

$$\Delta B_H(t) = [B_H(t) - B_H(t-1)] \tag{2.10}$$

The variance of the discrete process is:

$$var[B_H(t+n) - B_H(t)] = var[B_H(n) - B_H(0) = n^{2H} V_H \tag{2.11}$$

And the variance of the sample mean is of the form:

$$var[X] = \sigma^2 n^{2H-2} \tag{2.12}$$

Lastly the autocovariance of the discrete fGn is:

$$C(\tau, H) = \frac{1}{2}[|\tau+1|^{2H} - 2|\tau|^{2H} + |\tau-1|^{2H} \tag{2.13}$$

(Koutsoyiannis, 2002) analyzed the basic properties of Hurst's phenomenon and created three simple algorithms that could reproduce this effect. First he derived the equations characterizing the multiple scale properties of a typical stochastic process and found that a markovian process on aggregated time scale becomes more complicated than its basic scale and tends to white noise as the scale gets larger. He then showed that for any aggregated time scale fractional gaussian noise has an autocorrelation function:

$$\rho_j^{(k)} = \rho_j \left(\frac{1}{2}\right)[(j+1)^{2H} + (j-1)^{2H}] - j^{2H} \tag{2.14}$$

where k is the aggregated time scale and for large j:

$$\rho_j^{(k)} = \rho_j = H(2H-1)j^{2H-2} \tag{2.15}$$

This shows us that the autocovariance function of fGn in aggregated time scales is independent of the scale k and its statistical properties compared with the aggregated markovian process are more effective, much simpler hence more parsimonious. Koutsoyiannis produced a multiple time scale fluctuation approach, a disaggregation approach and a symmetric moving average approach, whereas , every approach could reproduce the desired effect. In a further paper (Koutsoyiannis, 2003) states that the classic hydrological statistics are not sufficient with the varying climate (following a simple scaling law), which is generally admitted, because the classical statistical estimators are based on samples consisting of independent and identically distributed variables which is not the case with hydrologic time series. Koutsoyiannis showed that the sample mean is an unbiased estimator but the variance is considerably higher than that of the classical statistics which results in differences in other statistics too. He also found that the quantiles of a distribution function are different than those of the classic statistics and when the Hurst coefficient and the standard deviation are unknown the confidence intervals are wider. Lastly he investigated the

estimation of the autocovariance and autocorrelation function and showed that the classical estimates are lower than the true values and after a small numbers of lags diminish, which is not the case for the simple scaling signal estimates which respect the long term persistence of the process and are identical with the theoretical ones.
(Koutsoyiannis, 2006) discusses that the most common way to model a hydrological process is by non-stationary models and that long-term trends are deterministic components of the time series. He states that the best practice is to assume stationarity with a parallel scaling behavior which reproduces climatic trends. This approach also indicates that the uncertainty increases because of the large scale fluctuations. This method though can be difficult to adapt in cases where there is change in a physical system (e.g. change of land use) and in those cases should be combined with a deterministic hydrological model.

2.2 Probability Concepts

2.2.1 Random Variables and Distribution Functions

In probability theory and statistics, a random variable $\underline{x}$ is a function that maps outcomes of a random process (such as the result of a coin flip) to numerical values. There are two types of random variables, discrete and continuous. A discrete random variable is a random variable that can take on only a countable number of distinct values. The probability of each outcome is represented by a probability mass function (PMF), which gives the probability of each possible value of the discrete random variable. A continuous random variable can take on any value within a specified interval or range on the real number line, meaning that their possible values form a continuous range of numbers. The probability is represented by a probability density function (PDF), which gives the relative likelihood for the random variable to take on a value within a given interval. By that meaning, the PDF is used to calculate probabilities for ranges of values rather than specific values, as with discrete random variables.

In the case of discrete random variables, the cumulative distribution function (CDF), denoted by $F_X(x)$ is a function, which satisfies

$$F_X(x) = P[X \leq x] \tag{2.16}$$

with properties of being a monotone and nondecreasing function and that $\lim_{x\to-\infty} F_X(x) = 0$ and $\lim_{x\to+\infty} F_X(x) = 1$. The probability mass function is denoted by $f_X(x)$ and defined by

$$f_X = P[x = x_j] \text{ for } x = xj, j = 1,2, \dots, n \tag{2.17}$$

with properties that satisfy $f(x_j) > 0$ for j=1,2,...,n and $\sum f(x_j) = 1$.

In continuous random variables the cumulative distribution function,

denoted by $F_X(x)$, satisfies

$$F_X(x) = \int_{-\infty}^{x} f(u)du \tag{2.18}$$

while the probability density function is obtained by differentiating the CDF, that is

$$f_X(x) = \frac{dF_X(x)}{dx} \tag{2.19}$$

with properties $f_X(x) > 0$ and $\int_{-\infty}^{\infty} f(x)dx = 1$.

2.2.2 Expectations, Moments and Cumulants

A very important tool in problems of probability theory and statistics because of its ability to summarize specific properties of a random variable or a distribution function is the expected value $E[X]$. Regardless of the type of the r.v. or the distribution, the expected value is an average of the values that a r.v. takes, while each value is weighted by a probability, or else it's a measure of the center of gravity of the density function of the r.v. Values that are more probable receive more weight and vice versa.

The expected value of a discrete random variable with probability mass function $P_X(x_i) = P[X = x_i]$ for i=1,...,n is defined by:

$$E[X] = \sum_{i=1}^{n} x_i P_X(x_i) \tag{2.20}$$

while, in the case of a continuous random variable with probability density function $f_X(x)$ the expected value is defined by:

$$E[X] = \int_{-\infty}^{+\infty} x f_x(x)dx \tag{2.21}$$

The moments or raw moments of a random variable are the expectations of the powers of the random variable, which describe certain quantitative measures related to the shape, the scale and the location of the distribution of the corresponding random variable.

If X is a random variable the r^{th} moment of X or the raw moment of X is:

$$\mu'_r = E[X^r] \tag{2.22}$$

The mean of the r.v. is the first raw moment:

$$\mu_1' = E[X] \tag{2.23}$$

The r^{th} central moment or moment about the mean of a random variable X is:

$$\mu_r = E[(X - \mu_X)^r] \tag{2.24}$$

One of the most frequent used central moments, is the central moment of order 2 or μ_2, which gives a measure of dispersion of the pdf of the random variable, it is called variance and is defined as:

$$\mu_2 = E[(X - \mu_X)^2] = \sigma^2 = var[X] \tag{2.25}$$

Finally, in applied statistics the 3^{rd} and 4^{rth} central moments are used, which are part of high order moments, used to describe the skewness and the kurtosis respectively. Skewness is a measure of the asymmetry of the distribution (value of zero means that the pdf is symmetric) and kurtosis describes the "tailedness" of the pdf around the expected value.

The cumulants of a random variable X are used to describe the same measures as classical moments but in a different way. According to (Johnson et al., 2005), cumulants are useful because of their additivity properties, meaning that when we deal with the sum of multiple independent variables, the cumulants can be calculated by adding the sums of the individual cumulants. The logarithm of the moment generating function of X is defined to be the cumulant generating function of X:

$$K(t) = lnE[e^{tx}] \tag{2.26}$$

The r^{th} cumulant of X denoted by κ_r is defined by the power series expansion of the cumulant generating function:

$$K(t) = \sum_{r=1}^{\infty} \kappa_r \frac{t^r}{r!} \tag{2.27}$$

The following formulas (Smith, 1995) connect the raw moments μ_r' to the cumulants:

$$\mu_r' = \sum_{i=0}^{r-1} \binom{r-1}{i} \kappa_{r-1} \mu_r' \qquad \kappa^r = \mu_r' - \sum_{i=0}^{r-1} \binom{r-1}{i} \kappa_{r-1} \mu_r'$$

If we acquire a random sample of size n of X_i random variables from a cumulative distribution function, then the order statistics(David and Nagaraja, 2003) of the random samples X_i, is defined to be the Y_i for i=1,…,n where Yi are the rearranged Xi in order of increasing magnitude. L-moments, introduced by (Hosking, 1990), is also an alternative way of describing the statistical properties of data using linear combinations of order statistics.

$$\lambda_1 = E[X] \tag{2.28}$$

$$\lambda_2 = \frac{1}{2}E\left[X_{(1:2)} - X_{(2:2)}\right] \tag{2.29}$$

$$\lambda_3 = \frac{1}{3}E\left[X_{(1:3)} - 2X_{(2:3)} + X_{(3:3)}\right] \tag{2.30}$$

$$\lambda_4 = \frac{1}{4}E\left[X_{(1:4)} - 3X_{(2:4)} + 3X_{(3:4)} - X_{(4:4)}\right] \tag{2.31}$$

2.3 Stochastic Processes

A stochastic process is a collection of random variables, typically indexed by time or space, which describe how the values of these random variables change over time. Stochastic processes provide a way to study, analyze and make predictions about processes that evolve in an uncertain manner(Papoulis and Pillai, 2009). The random variables $\{X(t), t \in T\}$ are defined on a common probability space (Ω, F, P), where Ω is the sample space, which is the set of all possible outcomes, F is a sigma algebra (σ-algebra) on Ω and P is a probability measure on (Ω, F), assigning probabilities to events. The index set T is interpreted as time, but it can be any index set, typically a subset of the real numbers. For each t in T, $X(t)$ is a random variable defined on (Ω, F) that assigns a real number to each outcome in Ω. The collection of random variables $\{X(t), t \in T\}$ is considered a stochastic process because it describes how the system's state evolves over the index set T in a probabilistic manner(Karlin and Taylor, 1975).

Like random variables, stochastic processes have two classifications based on the nature of the index set. Discrete stochastic processes, where the index set T is a countable set, often consisting of integers and continuous stochastic processes, where the index set T is a subset of the real numbers or an uncountable set.

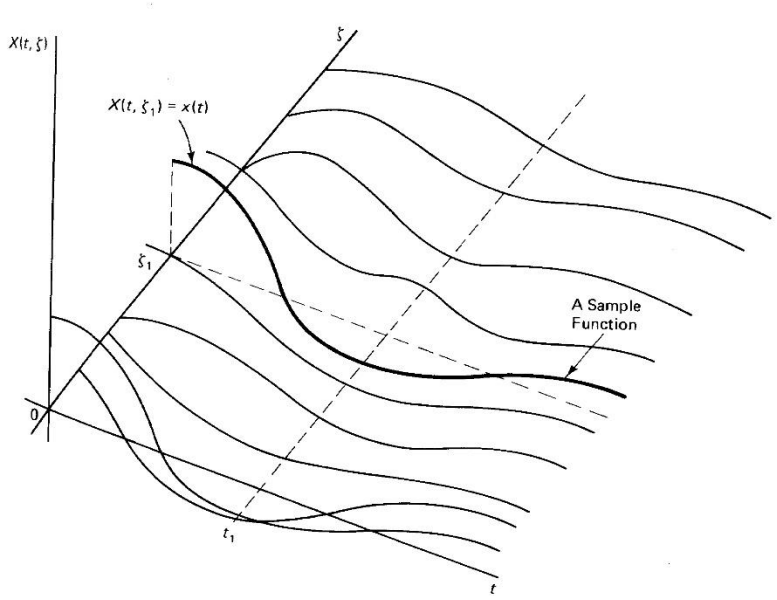

Figure 2.1: Ensemble of sample realizations of a stochastic process. Source: (Stark et al., 2012)

2.3.1 Statistics of Stochastic Processes

In this chapter we will present some statistical techniques and tools, which are used in the stochastic processes to model, analyze and make predictions with specific characteristics, who play an important role in understanding random phenomena.
A random variable $X(t)$ for a specific t has a distribution function:

$$F(x,t) = P\{x(t) \leq x\} \tag{2.32}$$

which is the first order distribution function of the process. Its derivative with respect to x:

$$f(x,t) = \frac{\partial F(x,t)}{\partial x} \tag{2.33}$$

is the first order density function of the process.
The second order distribution of the process $X(t)$ is the joint distribution:

$$F(x_1, x_2; t_1, t_2) = P\{X(t_1) \leq x_1, X(t_2) \leq x_2\} \tag{2.34}$$

of the random variables $X(t_1)$ and $X(t_2)$ and the equivalent density function is:

$$f(x_1, x_2; t_1, t_2) = \frac{\partial^2 F(x_1, x_2; t_1, t_2)}{\partial x_1 \partial x_2} \tag{2.35}$$

The complete distribution of a stochastic process refers to the entire probability distribution that describes the behavior of the process over its entire index set, which means it provides a comprehensive view of the random variables' behavior at all time points. The nth order distribution of the process $X(t)$ is the nth joint distribution:

$$F(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n) = P\{X(t_1) \leq x_1, X(t_2) \leq x_2, \dots, X(t_n) \leq x_n\} \tag{2.36}$$

and the equivalent density function is:

$$f(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n) = \frac{\partial^n F(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n)}{\partial x_1 \partial x_2, \dots, \partial x_n} \tag{2.37}$$

2.3.2 Second-Order Properties of Stochastic Processes

The second-order properties of a stochastic process refer to the statistical characteristics that take account the moments of the random variables in the process, which are important for understanding the process's variability and correlation between its values at different points in time. The mean $\mu(t)$ of $X(t)$ is the expected value of the random variable $X(t)$:

$$\mu(t) = E[X(t)] = \int_{-\infty}^{+\infty} x f(x, t) dx \tag{2.38}$$

The variance of the process is:

$$\gamma_0(t) = var[X(t)] = \int_{-\infty}^{+\infty} (x - \mu(t))^2 f(x, t) dx \tag{2.39}$$

The autocovariance $c(t, h)$ of $X(t)$ is the covariance of the random variables $X(t)$ and $X(t + h)$:

$$c(t, h) = cov[(X(t), X(t + h)] = E[(X(t) - \mu(t))(X(t + h) - \mu(t + h))] \tag{2.40}$$

The autocorrelation $r(t, h)$ of $X(t)$ is the correlation coefficient of the product

$X(t)X(t+h)$:

$$r(t,h) = corr[X(t), X(t+h)] = \frac{c(t,h)}{\sqrt{\gamma_0(t)\gamma_0(t+h)}} \tag{2.41}$$

The autocovariance function measures the covariance between values of a time series at different time points and it provides information about the strength and direction of the linear relationship between observations. Autocorrelation can take values between $1 < r(t,h) < 1$. While autocovariance measures the linear dependence between values at different time points, autocorrelation standardizes this measure providing a unitless values that is easier to interpret in terms of strength and direction of the relationship between the time series and its lagged version. Autocorrelation is essentially the normalized version of autocovariance(Papoulis and Pillai, 2009).

2.3.3 Stationarity and Ergodicity

In this subsection we will introduce two important statistical concepts used in the field of stochastic processes and time series analysis, stationarity and ergodicity. A process is considered stationary if its statistical properties do not change over time. There are two main types of stationarity, strict-sense stationarity and wide-sense stationarity(Papoulis and Pillai, 2009). A process is strictly stationary if the distribution function remains the same over time or else the joint probability distribution of any set points within the process is invariant to shifts in time. From the above definition we conclude that:

$$f(x_1, \dots, x_n; t_1, \dots, t_n) = f(x_1, \dots, x_n; t_1 + c, \dots, t_n + c) \tag{2.42}$$

for any c. It follows that $f(x;t) = f(x;t+c)$ for any c and so the first order density of $x(t)$ is independent of t, $f(x;t) = f(x)$. Wide-sense stationarity is a less restrictive form of stationarity that is also commonly used. A process is called wide-sense stationary of the mean of the process remains constant over time, that is $E[x(t)] = \mu$ and if its autocovariance function depends only on time differences $\tau = h$, $cov[x(t), x(t+h)] = c(h)$.
A non-stationary process is characterized by changes in its statistical properties over time. This means that the mean, variance or other important characteristics of the process are not constant across different time periods. Non-stationary behavior often manifests as seasonality or other patterns that evolve over time(Koutsoyiannis, 2021). An example of a non-stationary process is a random walk, where each value in the series is determined by adding a random step to the previous value.

In stochastic processes ergodicity refers to the idea that the long-term statistical behavior of a system is representative of its average behavior over

time. Moreover ergodicity simplifies the analysis of stochastic processes by allowing statistical properties estimated from a long sample path to be representative of the ensemble behavior. In a more mathematical manner, a stochastic process is said to be ergodic if the time average of any function of the process converges to the expected value of that function. For continuous time process the equation that follows the above statement is:

$$\lim_{T\to\infty}\frac{1}{T}\int_{o}^{T} g(x(t))dt = E[g\big(x(t)\big)] \tag{2.43}$$

and for discrete time process:

$$\lim_{T\to\infty}\frac{1}{T}\sum_{\tau=0}^{T} g(x_\tau) = E[g(x(t))] \tag{2.44}$$

2.4 Time Series Analysis

Time series analysis refers to statistical methods and models used to analyze, model and forecast time series. In this chapter we will refer to four basic time series models that have been used in hydrology, Autoregressive Models of order p, AR(p), Moving-Average Models of order q, MA(q), Autoregressive-Moving-Average Models, ARMA(p,q) and the Autoregressive Integrated Moving-Average Model (ARIMA). These models are based on the idea of linear filter model, which relates a stochastic process with a series of independent white noise terms a_t (Box et al., 2008). Mathematically this translates to:

$$x_t = a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \cdots + = \sum_{j=0}^{\infty}\psi_j a_{t-j} = \psi(B)a_t \tag{2.45}$$

where: a_i is the white noise error terms
$\psi_1, \psi_2 \ldots$ are the weights of the a_i parameters
$\psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \cdots$ is the transfer function of the filter.
B is a backward shift operator
x_t is the random variable normalized by subtracting the mean μ

Shift operator is a tool which shifts the process, m time units into the past or the future, either if we have a backward or forward shift operator respectively. In a mathematical manner the backward shift operator denoted by B states that $B^m Z_t = Z_{t-m}$, where for $m = 1$, $BZ_t = Z_{t-1}$. Following the same logic, the

forward shift operator is $F^m Z_t = Z_{t+m}$. Lastly the backwards difference operator is defined as $\nabla Z_t = Z_t - Z_{t-1} = (1-B)Z_t$. The white noise process is formed by a series of uncorrelated random variables with zero mean $E[a_t] = 0$ and $var[a_t] = \sigma_a^2$. That is why the autocovariance function is of the form:

$$\gamma_k = E[a_t a_{t+k}] = \begin{cases} \sigma_a^2, & k = 0 \\ 0, & k \neq 0 \end{cases} \tag{2.46}$$

and the autocorrelation function is $\rho_k = 1\ for\ k = 0$ and $\rho_k = 0\ for\ k \neq 0$. Subject to some prerequisites (Koopmans, 1995), we can relate the variable x_t to past x_{t-j} variables and an error term:

$$x_t = \pi_1 x_{t-1} + \pi_2 x_{t-2} + \cdots + a_t = \sum_{j=1}^{\infty} \pi_j x_{t-j} + a_t \tag{2.47}$$

The transfer functions $\psi(B)$ and $\pi(B)$ are related by the equation $\psi(B)\pi(B) = 1$.
In order for the process to be stationary the sequence of the weights ψ must converge $\sum_{j=0}^{\infty} |\psi_j| < \infty$, otherwise the filtered is not stable therefore the process is non stationary. If the π weights in Eq. (2.47) follow the same rules as the ψ weights, that is, $\sum_{j=0}^{\infty} |\pi_j| < \infty$, we say that the process is invertible(Brockwell and Davis, 2010). Alternatively, invertibility refers to a property of a model, which allows to alter its current form, to express the relationship between the current value of the time series and its past values.

2.4.1 Autoregressive Models, AR(p)

In this family of models called Autoregressive Models of order p, AR(p), the equation that describes these models assumes that the current value of the process is expressed as a linear ensemble of previous values of the process and a random part. The relationship between the current observation and past observations weakens as you go further back in time, or else, the influence of distant past observations diminishes and recent observations have a stronger impact on the forecast. This is known as memory decay and for that reason these models are often used for short-term forecasting where recent observations are more relevant than those from a long time ago. The equation governing these characteristics is:

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + a_t \tag{2.48}$$

and it's called autoregressive process of order p.

The autocovariance function of AR(p) can be obtained by multiplying the whole equation by x_{t-k} and taking the expected values:

$$x_t\, x_{t-k} = x_{t-\mathrm{k}}\left(\phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + \ a_t\right) \Rightarrow$$

$$E[x_t\, x_{t-k}] = E[x_{t-k}\left(\phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + \ a_t\right) \Rightarrow$$

$$E[x_t\, x_{t-k}] = E[x_{t-k}\phi_1 x_{t-1} + x_{t-k}\phi_2 x_{t-2} + \cdots + x_{t-k}\phi_p x_{t-p} + x_{t-k}a_t] \Rightarrow$$

$$E[x_t\, x_{t-k}] = E[x_{t-k}\phi_1 x_{t-1}] + E[x_{t-k}\phi_2 x_{t-2}] + \cdots + \mathrm{E}[x_{t-k}\phi_p x_{t-p}] + E[x_{t-\mathrm{k}}a_t]$$

There is no relationship between variables x_{t-k}, a_t and so $E[x_{t-\mathrm{k}}a_t] = 0$. The autocovariance of AR(p) is:

$$\gamma_k = \phi_1\gamma_{k-1} + \phi_2\gamma_{k-2} + \cdots + \phi_p\gamma_{k-p} \tag{2.49}$$

The autocorrelation function can easily be derived by dividing the process variance c_0, which provides:

$$\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2} + \cdots + \phi_p\rho_{k-p} \tag{2.50}$$

Lastly the variance of the AR(p) process can be derived by the Eq. (2.49) for $k = 0$ and dividing by σ_x^2

$$\gamma_o = \phi_1\gamma_1 + \phi_2\gamma_2 + \cdots + \phi_p\gamma_p + \sigma_\alpha^2 \Rightarrow$$

$$\sigma_x^2 = \frac{\sigma_\alpha^2}{1 - \phi_1\rho_1 + \phi_2\rho_2 + \cdots + \phi_p\rho_p} \tag{2.51}$$

Different forms of the autoregressive models have been used in stochastic hydrology, one of the most common is the AR(1) model. The equation of the AR(1) model is:

$$x_t\ = \phi_1 x_{t-1} + \ a_t \tag{2.52}$$

The autocovariance function of the AR(1) is:

$$\gamma_k = \phi^k\gamma_{k-1} \tag{2.53}$$

the autocorrelation function is:

$$\rho_k = \rho_1^k \tag{2.54}$$

and the process variance is:

$$\sigma_x^2 = \frac{\sigma_\alpha^2}{1 - \phi_1 \rho_1} = \frac{\sigma_\alpha^2}{1 - \rho_1^2} \tag{2.55}$$

Equation (2.52) can be written equivalently as $(1 - \phi_1 B)x_t = a_t = \psi^{-1}(B)x_t = \sum_{j=0}^{\infty} \phi_1^j a_{t-j}$. This proves that the autoregressive models, given that the order of the model is finite, equals to an infinite moving average. For the process AR(1) to be stationary, parameter ϕ_1 should be bounded so that $|\phi_1| < 1$. This proves that the root of $(1 - \phi_1 B)$ which is ϕ_1^{-1}, must lie outside of the unit circle. This can be generalized for an AR(p) process and the roots of the polynomial should also lie outside of the unit circle.

2.4.2 Moving Average Models, MA(q)

The Moving Average Model MA(q) is a family of time series models that represent the relationship between an observation and a linear combination of the current white noise error and the past, q, white noise error terms. The parameter q, represents the order of the MA model and indicates the number of past white noise error terms considered in the model, meaning it determines how far back in time the model looks to capture dependencies. Like the AR(p) process, for stationarity, the roots of the polynomial must lie outside the unit circle. The equation of the Moving Average Model of order q, MA(q) is:

$$x_t = a_t - \theta_1 \alpha_{t-1} - \cdots - \theta_p \alpha_{t-q} \tag{2.56}$$

Using the same approach as the AR(p) models, to obtain the autocovariance function of the model $\gamma_k = E[x_t x_{t-k}]$

$$\gamma_k = E[(a_t - \theta_1 \alpha_{t-1} - \cdots - \theta_p \alpha_{t-q})(a_{t-k} - \theta_1 \alpha_{t-k-1} - \cdots - \theta_p \alpha_{t-k-q})] \tag{2.57}$$

The autocovariance for $k > 0$ is:

$$\gamma_k = (-\theta_k + \theta_1\theta_{k+1} + \cdots + \theta_{q-k}\theta_q)\sigma_a^2 \tag{2.58}$$

The process variance for $k = 0$ becomes:

$$\gamma_0 = (1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2)\sigma_a^2 \tag{2.59}$$

Dividing the autocovariance function by γ_0 we get the autocorrelation function:

$$\rho_k = \frac{-\theta_k + \theta_1\theta_{k+1} + \cdots + \theta_{q-k}\theta_q}{1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2} \tag{2.60}$$

One used Moving Average model is the first order moving average denoted by MA(1) and the equation is:

$$x_t = a_t - \theta_1\alpha_{t-1} \tag{2.61}$$

The variance of the process is:

$$\gamma_0 = (1 + \theta_1^2)\sigma_a^2 \tag{2.62}$$

while the autocorrelation function is of the form:

$$\rho_k = \frac{-\theta_1}{1 + \theta_1^2} \; for\; k = 1 \; and \; 0 \; for \; k \; \geq 2 \tag{2.63}$$

2.4.3 Autoregressive-Moving-Average Models, ARMA(p,q)

Autoregressive Moving-Average Models, ARMA(p,q), is a family of models, that combines the previous aforementioned two models, Autoregressive and Moving Average. By allowing the noise term in the AR(p) model to consist of a moving average of independent and identically distributed random variables we obtain the ARMA(p,q) model. The autoregressive part

captures the linear relationship between an observation and its past values, while the moving-average part models the relationship between an observation and past white noise error terms. The stationarity conditions are the same as with the AR(p) and MA(q) models, that is, the roots of the two polynomials must lie outside the unit circle. The general form of an ARMA(p,q) model is given by the equation:

$$x_t = \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q} \tag{2.64}$$

where: x_t is the value of the time series at time t

$\phi_1, \dots, \phi_p$ are the autoregressive coefficients, p is the order of the AR part

$\theta_1, \dots, \theta_q$ are the moving average coefficients, q is the order of the MA part

a_t is the white noise error term at time t

Equation (2.64) can also take the form:

$$\left(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p\right) x_t = \left(1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_p B^p\right) a_t \Rightarrow$$

$$\phi(B) x_t = \theta(B) a_t \tag{2.65}$$

The autocovariance function is found the same way as the previous models, by multiplying and taking the expected value of the corresponding equation.

$$\gamma_k = E[x_t x_{t-k}] = E[x_{t-k}\left(\phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}\right)] \Rightarrow$$

$$\gamma_k = \phi_1 \gamma_{k-1} + \cdots + \phi_\rho \gamma_{k-\rho} + \gamma_{xa}(k) - \theta_1 \gamma_{xa}(k-1) - \cdots - \theta_q \gamma_{xa}(k-q) \tag{2.66}$$

In the above equation $\gamma_{xa}(k) = E[x_{t-k} a_t]$ and $\gamma_{xa}(k-q) = E[x_{t-k} a_{t-q}]$. We can see from Eq. (2.66) that x_{t-k} depends only on white noise error terms up to time $t-k$ and so for $k \geq 0$, $\gamma_{xa}(k) = 0$ but for $k \leq 0$, γ_{xa} has a value. With that said for $k \leq q$ the autocovariance function is Eq. (2.66) meaning the autocovariance of the ARMA(p,q) process depends on coefficients of the autoregressive as much as the coefficients of the moving average part and at the variance σ_a^2. On the other hand, for $k > 0$, the autocovariance function is:

$$\gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} + \cdots + \phi_\rho \gamma_{k-\rho} \tag{2.67}$$

and the autocorrelation function:

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \cdots + \phi_\rho \rho_{k-\rho} \tag{2.68}$$

reducing to exactly the same autocovariance and autocorrelation of AR(p) process. Finally the variance of an ARMA(p,q) process is defined by Eq. (2.67) for $k = 0$.

A model of the ARMA(p,q) family used in hydrology is the ARMA(1,1) (Salas et al., 1982; Salas and Obeysekera, 1982; Tao and Delleur, 1976) with equation:

$$x_t = \phi_1 x_{t-1} + a_t - \theta_1 a_{t-1} \tag{2.69}$$

The ARMA(1,1) process is stationary if $-1 < \phi_1 < 1$ and invertible if $-1 < \theta_1 < 1$.
The autocovariance function is:

$$\gamma_0 = \frac{1 + \theta_1^2 - 2\phi_1\theta_1}{1 - \phi_1^2}\sigma_a^2 \tag{2.70}$$

$$\gamma_1 = \frac{(1 - \phi_1\theta_1)(\phi_1 - \theta_1)}{1 - \phi_1^2}\sigma_a^2 \tag{2.71}$$

$$\gamma_k = \phi_1 \gamma_{k-1} \quad for\ k \geq 2 \tag{2.72}$$

The autocorrelation function is:

$$\rho_1 = \frac{(1 - \phi_1\theta_1)(\phi_1 - \theta_1)}{1 - \theta_1^2 - 2\phi_1\theta_1}\sigma_a^2 \tag{2.73}$$

$$\rho_k = \phi_1 \rho_{k-1} \quad for\ k \geq 2 \tag{2.74}$$

2.4.4 Autoregressive Integrated Moving-Average Model, ARIMA

The ARIMA(p,d,q) is a widely used model which consists of three main components, the autoregressive, the integrated and the moving average part. The autoregressive component represent the linear regression of the current value of the seiries on its past values, the moving average component represents the linear regression of the current value on past white noise terms, and I(integrated) component involves differencing the series to make it stationary. It is an extension of the ARMA model which can handle seasonal nonstationaries in the data. The model is denoted as ARIMA(p,d,q) where:

p: order of the autoregressive component
d: order of differentiation of the original data
q: order of the moving average component

If we have datasets of scale less than a year then we will probably observe seasonal patterns which involve repeating cycles over fixed time intervals (e.g. monthly). We may also observe systematic increase or decrease in the mean of the time series over time. These cases can be handled by differencing the dataset until the dataset becomes stationary. The difference between observations is expressed as $\nabla x_t = x_t - x_{t-1}$(∇^d where d is the dth difference of the series). If the order of differencing is zero we get ARIMA(p,0,q)=ARMA(p,q), ARIMA(p,0,0)=AR(p) and ARIMA(0,0,q)=MA(q). We have seen that for an ARMA process if the roots of the polynomial lie outside of the unit circle the process is stationary. In contrary if the roots lie inside the unit circle the generated values follow an exponential curve. We will deal with the case that the roots lie on the unit circle. The mathematical form of ARIMA model is

$$\varphi(B)x_t = \phi(B)\nabla^d x_t = \theta_0 + \theta(B)a_t \tag{2.75}$$

where: $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$
$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_p B^p$

The operator $\varphi(B) = \phi(B)\nabla^d$ is the generalized nonstationary autoregressive operator, $\phi(B)$ is the autoregressive operator which is assumed to be stationary so that the roots of $\phi(B)$ lie outside the unit circle and $\theta(B)$ is the moving average operator with roots $\theta(B) = 0$ outside of the unit circle so that it is invertible. The ARIMA(1,1,1) process which incorporates the autoregressive part, differencing and moving average is of the form $\nabla x_t - \phi_1 \nabla x_{t-1} = a_t - \theta_1 a_{t-1} \Rightarrow (1 - \phi_1 B)\nabla x_t = (1 - \theta_1 B)a_t$

2.5 Spectral Analysis of Time Series

2.5.1 Fourier Transform

Fourier transform is a mathematical technique that transforms a function of time or space into a function of frequency. It decomposes a complex signal into its constituent frequencies, which is helpful in understanding and manipulating signals in various applications. It expresses a function as a continuous integral of trigonometric or exponential functions(Osgood, 2019). The imaginary unit is a component in expressing both sine and cosine functions in a unified way through Eule's formula ($e^{i\theta} = \cos(\theta) + i\sin(\theta)$ and its reverse $\cos(\theta) = \frac{1}{2}(e^{i\theta} + e^{-i\theta})$, $\sin(\theta) = \frac{1}{2i}(e^{i\theta} - e^{-i\theta})$). The fourier transform of a non-periodic function $f(t)$ is defined as:

$$F(\omega) = \int_{-\infty}^{\infty} f(t)e^{-i2\pi\omega t}dt \tag{2.76}$$

Transforming now $F(\omega)$ into $f(t)$ is the inverse fourier transform and is defined as:

$$f(t) = \int_{-\infty}^{\infty} F(\omega)e^{i2\pi\omega t}dt \tag{2.77}$$

According to (Körner, 2004) there in not only one definition for the fourier transform and the different variations are summarized in the following equation:

$$F(\omega) = \frac{1}{A}\int_{-\infty}^{\infty} e^{iB\omega t}dt \tag{2.78}$$

Where $A = \sqrt{2\pi}, B = \pm 1$
$A = 1, B = \pm 2\pi$
$A = 1, B = \pm 1$

The fourier transform exists for certain classes of functions. The conditions for the existence of the fourier transform depend on the properties of the function being transformed. The function must be integrable over its entire domain and should be finite meaning $\int_{-\infty}^{\infty} f(t)dt < \infty$ and the function $f(t)$ should be piecewise continuous. This means that it can have a finite number of discontinuities (Bracewell, 2000). For the case that function $f(t)$ is real-valued and even, meaning $f(t) = f(-t)$ then $F(\omega)$ is also an even function and fourier transform simplifies to:

$$F(\omega) = \int_{-\infty}^{\infty} f(t)\cos(2\pi\omega t)\,dt = 2\int_{0}^{\infty} f(t)\cos(2\pi\omega t)\,dt \tag{2.79}$$

and the inverse fourier transform is:

$$f(t) = \int_{-\infty}^{\infty} F(\omega)\cos(2\pi\omega t)\,d\omega = 2\int_{0}^{\infty} F(\omega)\cos(2\pi\omega t)\,d\omega \tag{2.80}$$

For the case that function $f(t)$ is periodic with period 1, fourier transform for any non integer value of ω, is set to zero. This is a case where ω takes only k integer values, t is defined at (e.g. $[-1/2, 1/2]$) the fourier transform is:

$$F_k = \int_{-1/2}^{1/2} f(t)e^{-i2\pi kt}dt \tag{2.81}$$

the inverse transform becomes a summation:

$$f(t) = \sum_{k=-\infty}^{\infty} F_k e^{i2\pi kt} \tag{2.82}$$

If $f(t)$ is an even function for the case of $f(t)$ being a periodic function of period 1, the transformation simplifies to the Fourier cosine transform:

$$F_k = \int_{-1/2}^{1/2} f(t)\cos(2\pi kt)\,dt = 2\int_{0}^{1/2} f(t)\cos(2\pi kt)\,dt \tag{2.83}$$

and the inverse transform:

$$f(t) = \sum_{k=-\infty}^{\infty} F_k\cos(2\pi kt) = F_0 + 2\sum_{k=1}^{\infty}\cos(2\pi kt) \tag{2.84}$$

2.5.2 Power Spectrum

Generally, the power spectrum of a signal represents its frequency content. It provides insights into the distribution of power across different frequencies in the signal and it is useful for understanding the periodicities and dominant frequencies present in a time series.(Bloomfield, 2000) The power spectrum is derived from the Fourier cosine transform of the autocovariance function. For a stochastic process $\underline{x}_\tau$ in discrete time, with autocovariance function $c_\eta = Cov[\underline{x}_\tau, \underline{x}_{\tau+\eta}]$, the inverse finite fourier transform of the autocovariance function is the power spectrum $s(\omega)$, $\omega \in [0, 1/2]$. Autocovariance is an even function meaning $c_\eta = c_{-\eta}$

$$s(\omega) = 2\sum_{\eta=-\infty}^{\infty} c_\eta \cos(2\pi\eta\omega) = 2c_0 + 4\sum_{\eta=1}^{\infty} c_\eta \cos(2\pi\eta\omega) \tag{2.85}$$

And the inverse formula is:

$$c_\eta = \int_0^{1/2} s(\omega)\cos\,(2\pi\eta\omega)d\omega \tag{2.86}$$

In the case where a time series is used to compute the power spectrum, according to classical estimates for the autocovariance the equation is:

$$\widehat{c_\eta} = \frac{1}{n}\sum_{\tau=0}^{n-\eta}(x_\tau - \overline{x})(x_{\tau+\eta} - \overline{x}) \tag{2.87}$$

And this proves that the power spectrum $s(\omega)$ is equivalent with the periodogram for each discrete frequency $\omega = k/n$ with k being a positive integer $\leq n/2$. Respectively we can use the autocorrelation ρ_k rather than the autocovariance c_η and the corresponding function is:

$$p(\omega) = 2\left[1 + \sum_{\eta=1}^{\infty}\rho_k\cos\,(2\pi\eta\omega)\right] \tag{2.88}$$

is called the spectral density function and has the property:

$$\int_0^{1/2} p(\omega)d\omega = 1 \tag{2.89}$$

2.6 Sample Statistics

Sample statistics are numerical measures that provide insights into the characteristics of a sample of data. A statistic is a function of observable random variables, which does not contain any unknown parameters. These statistics summarize and describe the main features of a dataset such as the mean, variability and distribution. Let $X_1, X_2, \ldots, X_n$ be a random sample. The rth sample moment about 0 denoted by M_r' is:

$$M_r' = \frac{1}{n}\sum_{i=1}^{n} X_i^r \tag{2.90}$$

If $r = 1$ we get the sample mean denoted by $\overline{\mathrm{X}}$, that is:

$$\overline{\mathrm{X}} = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{2.91}$$

The rth sample moment about X denoted by M_r is:

$$M_r = \frac{1}{n}\sum_{i=1}^{n} (X_i - \overline{\mathrm{X}})^r \tag{2.92}$$

The sample variance from a random sample $X_1, X_2, \ldots, X_n$ is defined to be:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n} (X_i - \overline{\mathrm{X}})^2 \tag{2.93}$$

The first sample moment is the sample mean, which is a function of the random variables $X_1, X_2, \ldots, X_n$. If the random sample from a density $f(\cdot)$ has a mean μ and finite variance σ^2 then:

$$E[\overline{\mathrm{X}}] = \mu_{\overline{\mathrm{X}}} = \mu \tag{2.94}$$

and

$$var[\overline{\mathrm{X}}] = \sigma_{\overline{\mathrm{X}}}^2 = \frac{1}{n}\sigma^2 \tag{2.95}$$

Equation (2.94) tells us that on average $\overline{\mathrm{X}}$ the distribution of $\overline{\mathrm{X}}$ is centered about μ. Equation (2.95) says that the spread of the values of $\overline{\mathrm{X}}$ about μ is proportional with the sample size that is for small sample sizes the variance is larger than of that of a bigger sample size (e.g. the variance of a sample size of 50 is two times bigger than that of a sample size of 100). The larger the sample size, $\overline{\mathrm{X}}$ tends to be more concentrated around μ.

For the autocovariance estimation there is a downward bias (Salas, 1980) and the typical estimator of the lag η autocovariance is:

$$c_\eta = \frac{1}{n}\sum_{\tau=1}^{n-\eta}(x_\tau - \underline{\hat{\mu}})(x_{\tau+\eta} - \underline{\hat{\mu}}) \tag{2.96}$$

All the above classical statistical estimates have been used in hydrologic time series as tools to approximate various statistical results through available data time series or to approximate various parameters of stochastic models. The above estimates however, are used in cases where the samples come from independent and identically distributed random variables, which is not the case for hydrologic processes which have a physical underlying mechanism. We will now focus on (Koutsoyiannis, 2003) paper, who revisited these terms and under a simple scaling stochastic process or simple scaling signal reconstructed the statistical estimates as to respect the Hurst effect. The hydrometeorological processes exhibit scale invariant properties at any scale greater than annual with equation(Koutsoyiannis, 2002):

$$\left(Z_i^{(k)} - k\mu\right) = \left(\frac{k}{l}\right)^H \left(Z_j^{(l)} - l\mu\right) \tag{2.97}$$

The variance of the aggregated process for $i = j = l = 1$ is:

$$\gamma_0^{(k)} = var\left[Z_i^{(k)}\right] = k^{2H}\gamma_0 \tag{2.98}$$

Using equation (2.93) and observing $\overline{\mathrm{X}} = Z_1^{(n)}/n$ we get:

$$var[\overline{\mathrm{X}}] = \frac{\sigma^2}{n^{2-2H}} \tag{2.99}$$

In the case of $H = 0.5$ we get the same equation as classical statistics, but in hydrologic time series where $H > 0.5$ and for fixed H the variance is dramatically higher which results which result in difference in other statistics too. The variance becomes:

$$\tilde{S}^2 = \frac{n-1}{n-n^{2H-1}}S^2 = \frac{1}{n-n^{2H-1}}\sum_{i=1}^{n}(X_i - \overline{\mathrm{X}})^2 \tag{2.100}$$

The corresponding autocovariance is:

$$\tilde{c}_\eta = c_\eta + \frac{1}{n^{2H-2}}\tilde{S}^2 = c_\eta + \frac{n-1}{n^{3-2H}-n}S^2 \tag{2.101}$$

And the autocorrelation function is:

$$\tilde{\rho}_\eta = \frac{\tilde{c}_\eta}{\tilde{S}^2} = \rho_\eta\left(1 - \frac{1}{n^{2-2H}}\right) + \frac{1}{n^{2-2H}} \tag{2.102}$$

It is shown (Koutsoyiannis, 2003) that the classical autocorrelation is lower than the theoretical values and after a number of lags disappears, which doesn't reflect the Hurst effect. Contradictory the new estimator is very close to the theoretical one and exhibits the long-term persistence of the process.

3. Methodology

3.1 Symmetric Moving Average

(Koutsoyiannis, 2000) developed and implemented a generating scheme for either single variable or multivariable simulation and forecast, which is either for short term or long- term memory processes and preserves the Hurst coefficient. This scheme incorporates short and long memory models. As aforementioned in Section (2.4) a stochastic process can be formed as an infinite summation of independent and identically distributed random variables V_i (white noise):

$$X_i = \sum_{j=-\infty}^{0} a_{-j} V_{i+j} = \cdots + a_2 V_{t-2} + a_1 V_{t-1} + a_0 V_i \tag{3.1}$$

Where X_i is the random variable at time i and a_j are coefficients which can be found from the sequence of the autocovariance γ_i (Box et al., 2008).

$$\sum_{j=0}^{\infty} a_j a_{i+j} = \gamma_i \tag{3.2}$$

This model is known as Moving Average or specifically Backward Moving Average model. In practice there is no need to define X_i as an infinite summation rather than a limited summation because it is computationally impossible to produce infinite terms and from a certain point and beyond these terms can be neglected because as $j \to -\infty$ a_{-j} decreases (Koutsoyiannis, 2000). With that said (3.1) can be written as:

$$X_i = \sum_{j=-s}^{0} a_{-j} V_{i+j} = a_s V_{i-s} + \cdots + a_2 V_{i-2} + a_1 V_{i-1} + a_0 V_i \tag{3.3}$$

And the sequence of the autocovariance:

$$\sum_{j=0}^{s-i} a_j a_{i+j} = \gamma_i \tag{3.4}$$

The population of a_j j=0, 1, 2,…,s coefficients depends on simulation length and the accuracy of the model and the decay rate of the autocovariance function (Koutsoyiannis, 2000). By generalizing (3.1) we can relate X_i with previous and next white noise error terms V_i and we conclude in a Backward-forward moving average scheme (Koutsoyiannis, 2000) described by:

$$X_i = \sum_{j=-\infty}^{\infty} a_{-j} V_{i+j} = \cdots + a_{-1} V_{i-1} + a_0 V_i + a_1 V_{i+1} + a_2 V_{i+2} + \cdots \tag{3.5}$$

and the coefficients a_j are found by:

$$\sum_{j=-\infty}^{\infty} a_j a_{i+j} = \gamma_i \tag{3.6}$$

BFMA model (3.5) is a generalized form of the Backward Moving Average (3.1), which contains infinite possible solutions i.e. number of a_j combinations satisfying (3.6). For the case that $a_j = 0$ for $j < 0$ BFMA deduces to BMA and for the case that $a_{|j|} = a_j \; \forall \, j > 0$ we have the Symmetric Moving Average (SMA) (Koutsoyiannis, 2000) with equation:

$$\begin{aligned} X_t = \sum_{j=-\infty}^{\infty} a_{|j|} V_{i+j} \\ = \cdots + a_2 V_{i-2} + a_1 V_{i-1} + a_0 V_i + a_1 V_{i+1} + a_2 V_{i+2} \\ + \cdots \end{aligned} \tag{3.7}$$

And the coefficients a_j are associated to γ_i with the equation:

$$\sum_{j=-\infty}^{\infty} a_{|j|} a_{|i+j|} = \gamma_i \tag{3.8}$$

With infinite white noise error terms and with limited parameters:

$$\begin{aligned} X_t = \sum_{j=-s}^{s} a_{|j|} V_{i+j} \\ = a_s V_{i-s} + \cdots + a_1 V_{i-1} + a_0 V_i + a_1 V_{i+1} + \cdots \\ + a_s V_{i+s} \end{aligned} \tag{3.9}$$

and the coefficients a_j are related to γ_i with equation:

$$\sum_{j=-s}^{s-i} a_{|j|} a_{|i+j|} = \gamma_i \tag{3.10}$$

Despite the computation of the coefficients a_j either in the BMA model or the SMA model, it is necessary to compute the statistical characteristics of the random variable V_i(Koutsoyiannis, 2000). Then we define the mean $\mu_v = E[V_i]$, the variance of the random variable $var[V_i] = 1$ and $\xi_v = E[(V_i - \mu_v)^3]$ which is the coefficient of skewness. Respectively the statistical characteristics of the random variable X_i is the mean $\mu_X = E[X_i]$, the variance of the random variable

$var[X_i] = \gamma_0$ and the skewness coefficient $\xi_X = E[(X_i - \mu_X)^3]$. For the BMA model these parameters are related by:

$$\left(\sum_{j=0}^{s} a_j\right)\mu_v = \mu_X \tag{3.11}$$

$$\left(\sum_{j=0}^{s} a_j^3\right)\xi_v = \xi_X \gamma_0^{3/2} \tag{3.12}$$

And for the symmetric moving average model the parameters are related by:

$$\left(a_0 + \sum_{j=1}^{s} a_j\right)\mu_v = \mu_X \tag{3.13}$$

$$\left(a_0^3 + 2\sum_{j=1}^{s} a_j^3\right)\xi_v = \xi_X \gamma_0^{3/2} \tag{3.14}$$

The advantages of the SMA model over the BMA model according to(Koutsoyiannis, 2000) is that the SMA approach has a closed-form solution for determining a_j coefficients in contrast with the BMA that has only numerical solutions. The coefficient of skewness of the white noise terms in the SMA model is lower than that of the BMA model which means it better preserves skewness. Also, the a_j coefficients of the SMA model for large values of j are smaller than those of the BMA model, tested in both Markovian and FGN process and because of that we can neglect a_j coefficients of the SMA model for $j > s$ (finite summation variant).
The coefficients a_j of the SMA model can be approximated using the power spectrum of the X_i process. Let $s_\gamma(\omega)$ denote the power spectrum of the process X_i which is the discrete fourier transform of the autocovariance series γ_j

$$s_\gamma(\omega) = 2\gamma_0 + 4\sum_{j=1}^{\infty} \gamma_j \cos(2\pi j\omega) = 2\sum_{j=-\infty}^{\infty} \gamma_j \cos(2\pi j\omega) \tag{3.15}$$

and the power spectrum of the a_j coefficients:

$$s_a(\omega) = 2a_0 + 4\sum_{j=1}^{\infty} a_j \cos(2\pi j\omega) = 2\sum_{j=-\infty}^{\infty} a_j \cos(2\pi j\omega) \tag{3.16}$$

$s_a(\omega)$ and $s_\gamma(\omega)$ are related by:

$$s_a(\omega) = \sqrt{2s_\gamma(\omega)} \tag{3.17}$$

The solution of the a_j coefficients can be found with the inverse transform

$$a_j = \int_0^{1/2} s_a(\omega)\cos(2\pi j\omega)\, d\omega \tag{3.18}$$

In the case of FGN process (Koutsoyiannis, 2002) the sequence of the a_j coefficients can be approximated using equation:

$$a_0 = \frac{\sqrt{(2-2H)\gamma_0}}{1.5-H} \quad for\ j = 0 \tag{3.19}$$

And for $j > 0$

$$a_j \approx \frac{a_0}{2}[(j+1)^{H+0.5} + (j-1)^{H+0.5} - 2j^{H+0.5}] \tag{3.20}$$

3.2 Hurst-Kolmogorov process

As aforementioned a stochastic process can be either continuous or discrete meaning that it either describes a system over a continuous time or at distinct points in time. The difference is that the underlying parameter, time in our case, takes values from a continuous set or from a discrete set respectively. In nature all the atmospheric and hydroclimatic processes are being evolved in continuous time so its logical we model them as continuous stochastic systems. First we will define some properties of stochastic processes of the cumulative process $\underline{X}(t)$ and the discrete process $\underline{x}_\tau$(Koutsoyiannis, 2021).

The discrete time process $\underline{x}_\tau$ is:

$$\underline{x}_\tau = \frac{1}{D}\int_{(\tau-1)D}^{\tau D} \underline{x}(u)du \tag{3.21}$$

where D is the time step of the continuous time interval. The cumulative process $\underline{X}(t)$ is:

$$\underline{\underline{X}}(t) = \frac{1}{D}\int_{0}^{t} \underline{x}(u)du \tag{3.22}$$

The discrete time process representation is:

$$\underline{x}_{\tau}(D) = \frac{1}{D}\int_{(\tau-1)D}^{\tau D} \underline{x}(u)du = \frac{\underline{\underline{X}}(\tau D) - \underline{\underline{X}}((\tau-1)D)}{D} \tag{3.23}$$

And the discrete time process representation at multiple time scales:

$$\underline{x}_{\tau}{}^{(\kappa)} = \underline{x}_{\tau}{}^{(\kappa)}(\kappa D) = \frac{1}{\kappa D}\int_{(\tau-1)\kappa D}^{\tau\kappa D} \underline{x}(u)du = \frac{\underline{\underline{X}}(\tau\kappa D) - \underline{\underline{X}}((\tau-1)\kappa D)}{\kappa D} \tag{3.24}$$

The observations of the hydroclimatic processes and the generation of synthetic time series are made of discrete-time data, but, because these processes are being evolved in continuous time, we have to derive discrete time processes from the continuous ones and take account of the discretization.

The variance at time t is the cumulative climacogram:

$$\Gamma(t) = var\left[\underline{\underline{X}}(t)\right] = t^2\gamma(t) \tag{3.25}$$

And the variance at time scale k of the time averaged process is the continuous time climacogram (D Koutsoyiannis, 2017):

$$\gamma(k) = var\left[\frac{\underline{\underline{X}}(k)}{k}\right] = \frac{\Gamma(k)}{k^2} \tag{3.26}$$

The climacogram was introduced by (Koutsoyiannis, 2011) and its advantages where studied by (Dimitriadis and Koutsoyiannis, 2015) against the autocovariance function and the power spectrum of either a markovian or Hurst-Kolmogorov process. One major advantage is that the climacogram is not affected by discretizing a continuous process because the equation characteristics are similar in either case (Koutsoyiannis, 2017). The climacogram is also proved to have the smallest estimation error among the autocovariance and the power spectrum and also its bias can be estimated by an analytical equation. The autocovariance of a continuous stochastic process of time lag h is related to the climacogram (Koutsoyiannis, 2016) with equation:

$$c(h) = cov[x(t), x(t+h)] = \frac{1}{2}\frac{d^2\Gamma(h)}{dh^2} \tag{3.27}$$

The power spectrum $s(w)$ is defined as the cosine Fourier Transform and the inverse:

$$s(w) = 4\int_0^\infty c(h)\cos(2\pi wh)\,dh, \quad c(h) = \int_0^\infty s(w)\cos(2\pi wh)\,dw \tag{3.28}$$

The discrete time autocovariance of time lag h(Koutsoyiannis, 2016) is:

$$c_\eta = \frac{1}{D^2}\left(\frac{\Gamma(|\eta+1|D) + \Gamma(|\eta-1|D)}{2} - \Gamma(|\eta|D)\right) \tag{3.29}$$

And it's related to the discrete power spectrum with:

$$s_d(\omega) = 2c_0 + 4\sum_{\eta=1}^{\infty} c_\eta\cos(2\pi\eta\omega)\,, c_\eta = \int_0^{1/2} s_d(\omega)\cos(2\pi\omega\eta)\,d\omega \tag{3.30}$$

General properties and second order characteristics of continuous and discrete time process can be found in (Koutsoyiannis, 2021). The Hurst-Kolmogorov process is attributed to (Hurst, 1951) who introduced this idea while working on the statistical properties of hydrological time series and to (Kolmogorov, 1940) who developed a similar mathematical process. The continuous Hurst-Kolmogorov process described by its climacogram is:

$$\gamma(k) = \lambda\left(\frac{a}{k}\right)^{2-2H} \tag{3.31}$$

Because the instantaneous variance of the process is infinite, the process in discrete time is not physically possible and that is why we construct the average process which behaves well, meaning it has finite properties (Koutsoyiannis, 2016). For $H = \frac{1}{2}$ the process becomes white noise, for ½ < $H < 1$ the process is persistent (long-range dependence) and for $0 < H < \frac{1}{2}$ the process is antipersistent. In the latter case, in the autocovariance function for time lag $h = 0$ we get the variance of the process $c(0) = +\infty$. For time lag $h > 0$ autocovariance is negative $c(h) < 0$ which for natural processes is not feasible. That's why the Hurst-Kolmogorov can reproduce these processes at $1/2 < H < 1$ in large scales.

Table 3.1: Second order characteristics of Hurst-Kolmogorov process at continuous and discrete time adapted from (Koutsoyiannis, 2016).

Property	Formula	Eq.
Variance Continuous Time process	$\gamma_0 = \gamma(0) = c(0) = +\infty$	(3.32)

Averaged process at scale (climacogram)	$\gamma(k) = \lambda(\alpha/\kappa)^{2-2H}$	(3.33)
Autocovariance function Continuous time, lag h	$c(h) = \begin{cases} \lambda H(2H-1)\left(\frac{a}{h}\right)^{2-2H}, & H > 1/2 \\ \lambda\delta\left(\frac{h}{a}\right), & H = 1/2 \\ \lambda H(2H-1)\left(\frac{a}{h}\right)^{2-2H} + \delta''\left(\frac{h}{a}\right), & H < 1/2 \end{cases}$	(3.34)
Discrete time, lag $\eta = h/D$	$c_\eta = \lambda(\alpha/D)^{2-2H}\left(\frac{\lvert j-1\rvert^{2H} + \lvert j+1\rvert^{2H}}{2} - \lvert j\rvert^{2H}\right)$	(3.35)
Power Spectrum Continuous time freq. w	$s(w) = \frac{2a\lambda\Gamma(2H+1)\sin(\pi H)}{(2\pi\alpha w)^{2H-1}})$	(3.36)

The above functions describe a simple scaling law(Koutsoyiannis, 2017; Koutsoyiannis, 2021, 2016), when plotted in double-logarithmic plot, are presented as straight lines and the slope equation is:

$$\gamma^{\#}(\Delta) = \frac{d(ln\gamma(\Delta))}{d(ln\Delta)} = \frac{\Delta\gamma'(\Delta)}{\gamma(\Delta)} \tag{3.37}$$

For the climacogram and respectively $c^{\#}(h)$ and $s^{\#}(w)$ for the autocovariance and power spectrum. It is shown (Koutsoyiannis, 2016) that the HK process maintains all the log-log derivatives of the second order properties constant for any scale, lag and frequency, while in other processes these slopes vary with a change of scale, lag or frequency. In order to handle some inconsistencies of the HK process (Koutsoyiannis, 2016) proposed an alternative process (Hybrid HK) based on its climacogram described by a Cauchy type equation:

$$\gamma(\kappa) = \lambda(1 + (\kappa/a)^{2M})^{(H-1)/M} \tag{3.38}$$

In this form, a second parameter M is introduced defined in the interval $[0,1]$ which describes local properties of the process while H describes its global properties. The variance of the Hybrid HK process is in this case finite $\gamma(0) = \lambda$ and by introducing a second parameter κ the model can achieve for small time scales a Markov type behaviour and for large time scales Hurst behaviour in analogy of the set of the parameters. Other types of HK process can be found in (Koutsoyiannis, 2017) described by their climacogram such as:

The Dagum-type (FHK-D) climacogram:

$$\gamma k = \lambda(1 - (1 + (k/a)2(H-1))M/(H-1) \tag{3.39}$$

The mixed Cauchy-Dagum type (FHK-CD) climacogram:

$$\gamma(k) = \lambda_1(1 + k/a_1)^{2H-2} + \lambda_2(1 - (1 + a_2/k)^{-2M} \tag{3.40}$$

In the case of a Hurst-Kolmogorov Process with $H > 0.5$ the a_j coefficients of the SMA model, can be calculated analytically (Koutsoyiannis, 2016) through equation:

$$a_j = \sqrt{\frac{2\Gamma(2H+1)\sin(\pi H)\,\gamma(D)}{\Gamma^2(H+1/2)(1+\sin(\pi H))}}\left(\frac{|j+1|^{H+1/2} + |j-1|^{H+1/2}}{2} - |j|^{H+1/2}\right) \tag{3.41}$$

Where Γ is the gamma function, and H the Hurst index.

3.3 Extension of the Symmetric Moving Average

In this present thesis we will refer and use also the extension of the SMA model derived by (Dimitriadis and Koutsoyiannis, 2018), which preserves its first four central moments which is found to be useful in cases where there is intermittent data (e.g. rainfall data, seasonal river discharge). These high order central moments will be found theoretically by choosing a distribution function which will be fitted in data.

Classical statistics as aforementioned assume independence of the variables, which is not the case for hydrological processes, therefore it is required much larger samples to obtain estimates of similar reliability with classical statistics and it is suggested that it should be avoided using classical moments of order $q > 2$ (Lombardo et al., 2014). Also (Dimitriadis and Koutsoyiannis, 2018) showed that for an HK process, with $H = 0.8$ to estimate the average of the process with an error approximate to $10\%\sigma$ we will need a time series of length $n = 10^5$. In (Koutsoyiannis, 2021) (Digression 4.B) proves with a simple numerical problem that the estimators of the noncentral moments $\hat{\mu}'_q$ are not a knowable quantity and we cannot define its value from a sample. Therefore, we will calculate theoretically from a chosen distribution

function their moments and preserve these moments in the simulation. The pth moment of the SMA model is:

$$E\left[\underline{x}_i^p\right] = E\left[\left(\sum_{j=-l}^{l} a_{|j|}\underline{v}_{i+j}\right)^p\right] \tag{3.42}$$

For $p = 4$ we get:

$$\begin{aligned} E\left[\underline{x}_i^4\right] &= E\left[\left(\sum_{j=-l}^{l} a_{|j|}\underline{v}_{i+j}\right)^4\right] \\ &= E\left[\underline{v}^4\right]\sum_{j=-l}^{l} a_{|j|}^4 + 6\sum_{j=-l}^{l-1}\sum_{k=j+1}^{l} a_{|j|}^2 a_{|k|}^2 \end{aligned} \tag{3.43}$$

Extending to the fourth central moment the coefficient of kurtosis is:

$$C_{k,v} = \frac{\left(\sum_{j=-l}^{l} a_{|j|}^2\right)^2}{\left(\sum_{j=-l}^{l} a_{|j|}^4\right)} C_{k,x} - 6\frac{\sum_{j=-l}^{l-1}\sum_{k=j+1}^{l} a_{|j|}^2 a_{|k|}^2}{\sum_{j=-l}^{l} a_{|j|}^4} \tag{3.44}$$

Where $C_{k,x}$ is the coefficient of kurtosis of $\underline{x}$. As an example (Dimitriadis and Koutsoyiannis, 2018) applied the symmetric moving average scheme using various two-parameter distributions and preserving the first four moments to white noises processes and found that this methodology can approximate very well the theoretical distribution. Current stochastic generation schemes can hardly deal with non-Gaussian distributions and can hardly handle moments higher than second order. The SMA scheme can approximate the probability density function by preserving the first four central moments which is sufficient for various distributions used in geophysical processes. The coefficients a_j of the SMA scheme are numerically calculated through the Fourier transform of the discrete power spectrum of the coefficients which is related to the discrete power spectrum of the process (Koutsoyiannis, 2000):

$$s_d^a(\omega) = \sqrt{2s_d(\omega)}$$

where $s_d^a(\omega)$ and $s_d(\omega)$ are the power spectra of the SMA coefficients and of the discrete time process respectively. The steps we followed were:

a) Remove the mean and divide by the standard deviation for each time series, in order to create a dimensionless data set (this is done because the Generalized HK model cannot handle the periodicities of the data).

b) Create the climacogram for each dimensionless time series and find Hurst Index.
c) Estimation of the first four central moments from the dimensionless time series.
d) The parameters of the GHK model that were the input of our model are: the mean value, the standard deviation, the coefficient of skewness, the coefficient of kurtosis, Hurst index and the length of the time series.
e) Synthesis of standardized time series through the SMA scheme and GHK model.
f) Add the mean value and multiply with the standard deviation to remove the standardization from the synthetic time series.

4.Applications

4.1 Case Study and Energy Harvesting

The global transition towards renewable energy sources is an imperative need to secure sustainable energy access. For island communities, the significance of renewable energy is even more pronounced due to their limited land availability and dependence on imported fossil fuels. Island communities face unique energy challenges due to their geographical isolation and reliance on imported fossil fuels. Achieving energy independence is not only a strategic goal for reducing dependence on costly imports but also a means of enhancing resilience to external shocks. In this project we explore the stochastic nature of renewable energy resources, emphasizing the need to assess them through data driven methodologies for the quantification of energy resources essential for achieving energy independence on islands. The methodology discussed is implemented in three different time series of meteorological data of the island of Astypalea which includes, Solar irradiance, Wind speeds and Wave heights and the statistical characteristics of the results of the stochastic simulation are compared with the ones of the historical time series to assess the efficiency of the model to provide adequate results.

Renewable energy resources comprise a group of diverse sources such as solar, wind, waves and hydroelectric. Unlike fossil fuel, renewable sources derive their energy from natural processes that are continuously replenished making them inexhaustible. However, one of the primary challenges associated with renewable energy is its stochastic nature, characterized by variability and intermittency. Solar, wind and wave energy, for instance are highly dependent on weather conditions, resulting in fluctuations in energy generation. This inherent variability poses challenges for integrating renewable energy into existing grids. To quantify and harness the potential of renewable energy resources effectively, stochastic approaches are indispensable. Meteorological data, such as solar irradiance, wind speeds and wave heights provide valuable insights into the temporal and spatial distribution of renewable energy resources. For example, a comprehensive energy assessment of an island may reveal abundant solar potential, making solar photovoltaic systems a viable solution for meeting a significant portion of its electricity demand. Wind energy assessments could identify favorable locations for wind farm installations, taking into account factors such as wind speed and terrain. Moreover, integrating multiple renewable energy sources into hybrid energy systems can enhance reliability and resilience, leveraging the complementary nature of different resources to mitigate variability and ensure a stable power supply.

The transition towards renewable energy requires an approach encompassing technological innovation, policy support and community engagement. Governments and policymakers play an important role in creating enabling frameworks that promote renewable energy investments, energy efficiency measures and facilitate grid modernization.

Astypalea island is located in the Southern Aegean Sea as part of the Dodecanese island group of Greece. Spanning an area of approximately 97 km^2 and the coastline extends over 110 km, the island's landscape is characterized by rugged cliffs, rolling hills and coves, all contributing to its distinct geomorphology. Geologically, Astypalea is primarily composed of limestone and volcanic rocks, shapes over time by seismic activity and erosion processes (ΧΡΙΣΤΟΔΟΥΛΟΥ). This geological makeup has given rise to a diverse terrain, featuring rocky outcrops, fertile valleys and limestone formations. The island's Mediterranean climate ensures hot, dry summers and mild winters, with over 300 days of sunshine annually.

Astypalea is home to a permanent population of approximately 1300 residents, primarily concentrated in the main town and several coastal villages. Tourism plays a significant role in Astypalea's economy, particularly during the summer months when the island experiences a notable influx of visitors. According to the Hellenic Statistical Authority in 2018 there was more than 11000 arrivals of visitors in the island with average overnight stays of 4 days. This means that in summer in certain periods the total population of the island is doubled and that results in the majority of the energy consumed to cover needs for tourists rather than the permanents residents.

Figure 4.1: Island of Astypalea

In this project we will use three historical time series covering wind speeds, wave heights and solar irradiance of the island of Astypalea and we will compare the historical statistical characteristics with the simulated ones in order to quantify the uncertainty. These datasets where acquired from the Hellenic National Meteorological Service. The proposed model and the statistical description of the results could help in cases where there is need to decide what renewable energy resources could be used in cases such as Astypalea, either if it is solar panels, wind turbines or wave generators or a mix of the above technologies. These technologies will be explained further in the next section.

Solar Irradiance can be harvested through solar panels (photovoltaic panels) which generate electric current in certain materials when exposed to

light. The basic components of a solar panel include a) photovoltaic cells, which are the building blocks of solar panels. Typically made of semiconductor materials, these cells absorb photons from sunlight, exciting electrons and creating an electric current. Metal contacts, located on the top and bottom of the cell, metal contacts facilitate the flow of electrons generated by the photovoltaic effect. Anti-reflective coating, a layer on the top surface of the cell reduces reflection of sunlight, allowing more photons to enter and be absorbed by the semiconductor material(Zaidi, 2018). Finally solar cells are encapsulated within a protective material such as tempered glass and polymer to shield them from environmental factors and ensure durability.

Figure 4.2: Photovoltaic Panels Setup

When sunlight hits the photovoltaic cells, it knocks electrons loose from their atoms, allowing them to flow freely. This flow of electrons creates an electric current, which can he harnessed for various applications.
Solar panels come in different types, each with its unique characteristics and applications. Monocrystalline solar panels are made from single-crystal silicon, giving them a uniform appearance. The offer high efficiency and power output, making them suitable for limited roof space installations, they are durable and have a long lifespan. Polycrystalline solar panels are manufactured from multiple silicon crystals, giving them a blue appearance. They are less expensive to produce than monocrystalline panels making them a cost-effective option. Although slightly less efficient than monocrystalline panels, they still offer good performance and are suitable for large scale installations where space is not a constraint.

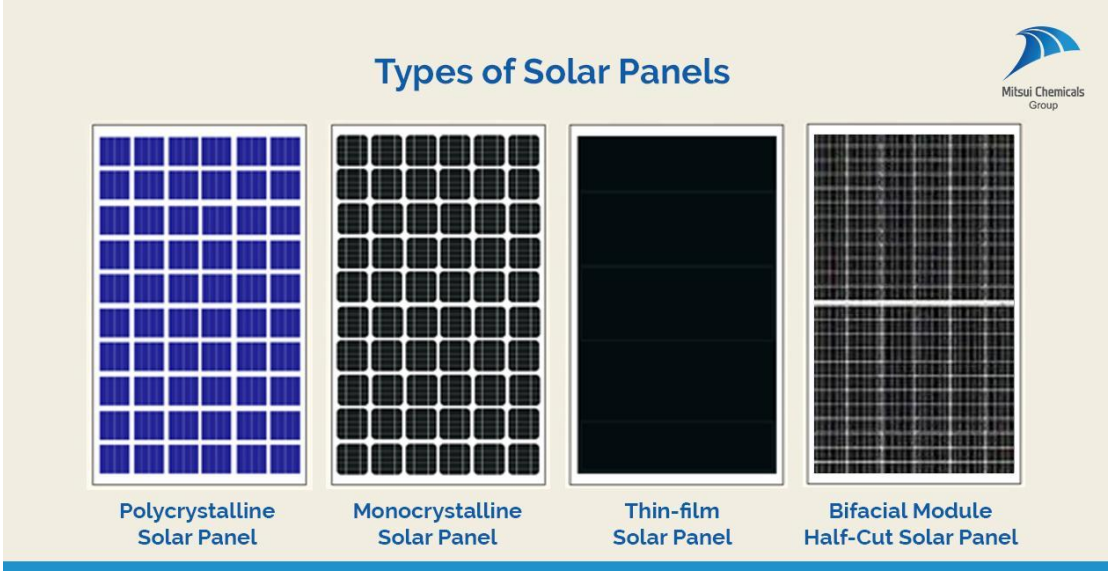

Figure 4.3 Types of Solar Panels

Thin-film panels use thin layers of photovoltaic material deposited onto a substrate such as glass, plastic or metal. They are lightweight and flexible, allowing for easy integration into various applications such as building-integrated photovoltaics. Thin-film panels have lower efficiency compared to crystalline silicon panels but perform better in low light conditions. They are cost-effective to produce and are increasingly used in large-scale solar farms and portable solar chargers.

Lastly bifacial solar panels can capture sunlight from both sides, maximizing energy generation. They have transparent backsheets, allowing sunlight to pass through and be reflected onto the rear side of the panel, increasing overall efficiency. Bifacial panels are suitable for installations with reflective surfaces or elevated mounting structures(Kenu E., 2020). They offer increased energy yield compared to traditional monofacial panels but come at a higher cost.

While all solar panels work on the same basic principle of converting sunlight into electricity, they differ in terms of efficiency, cost, durability and suitability for various applications.

1) **Efficiency**: Monocrystalline panels typically have the highest efficiency, followed by polycrystalline and thin film. Bifacial panels offer increased efficiency due to their ability to capture sunlight from both sides.
2) **Cost**: Polycrystalline panels are generally more affordable than monocrystalline panels, making them a popular choice for large-scale installations. Thin film panels offer a cost-effective alternative, especially for applications requiring flexibility and lightweight design.
3) **Durability**: Monocrystalline and polycrystalline panels are known for their durability and long lifespan. Thin film panels have shorter lifespans and may degrade faster over time.
4) **Applications**: The choice of solar panel type depends on the specific requirements of the application. Monocrystalline panels are suitable for residential rooftops and small scale installations where space is limited.

Thin film panels are preferred for portable applications and building-integrated photovoltaics and bifacial panes are ideal for commercial and utility-scale installations where maximizing energy yield is paramount(Vodapally and Ali, 2022).

Wind energy can be harvested through Wind turbines. Wind turbines operate on the principle of converting kinetic energy from wind into mechanical power, subsequently transformed into electricity. The basic components of a wind turbine include the rotor blades, hub, generator, gearbox and tower. The rotor blades are aerodynamically designed blades which capture wind energy. Blade length and shape significantly influence turbine efficiency. The hub is the connection of the blades with the rotor shaft, transmitting rotational energy. The generator converts mechanical energy into electricity and the gearbox increases rotational speed to optimize generator performance. The tower provides height to maximize wind exposure. When wind strikes the rotor blades, it causes them to rotate. The kinetic energy from the rotating blades is transferred through the shaft to the generator. Then the generator converts this mechanical energy into electrical energy, which is subsequently transmitted to the power grid.

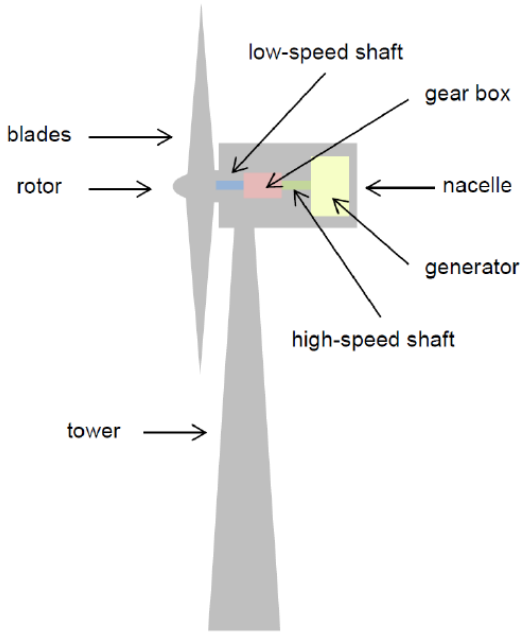

Figure 4.4: Wind Turbine Basic parts

Wind turbines can be classified based on various factors, including axis orientation, size and design. The two major types are horizontal axis wind turbines and vertical axis wind turbines. The first is the most common used type, where, blades rotate around a horizontal axis, is efficient in moderate to high wind speeds and some require a rotating mechanism to adjust to wind direction changes. The latter is suitable for areas with turbulent or inconsistent wind patterns(Konstantinidis and Botsaris, 2016). Its blades rotate around a vertical axis and it is less efficient than horizontal axis wind turbine but offer advantages in certain scenarios. Each type of wind turbine possesses distinct

characteristics, influencing performance, efficiency and suitability for various environments.

1) **Size and Capacity:** Horizontal axis turbines have larger capacities and are commonly used in utility-scale projects while vertical axis turbines have smaller capacities and are utilized in smaller-scale applications, such as residential or remote locations.
2) **Efficiency and Performance:** Horizontal axis turbines exhibit higher efficiency, especially in consistent wind conditions, rather than vertical axis turbines which are more suitable in areas with turbulent winds due to their omnidirectional blade orientation.
3) **Cost and Maintenance:** Horizontal axis turbines often require higher initial investments but may offer better economies of scale in large-scale projects. Vertical axis turbines have lower upfront costs and potentially lower maintenance requirements due to simple design.
4) **Adaptability to wind conditions:** Horizontal axis turbines may require a rotation mechanism to adjust to wind direction changes ensuring optimal performance. Vertical axis turbines may be more adaptable to varying wind directions without the need for complex mechanisms(Wagner, 2018).

Advancements in wind turbine technology continue to drive improvements in efficiency, performance and reliability. An example is, floating wind turbines designed for offshore installations, floating turbines utilize buoyancy systems to remain afloat and are advantageous in cases where there is limited land area. Multi rotor turbines incorporate multiple rotors on a single tower, potentially increasing energy capture efficiency and reducing land footprint. As technology continues to evolve, ongoing innovation promises further enhancements in efficiency, performance and environmental sustainability in the field of wind energy.

Wave energy can be harvested through wave generators. Wave energy derived from the kinetic and potential energy of ocean waves, presents a promising avenue for renewable energy generation. Wave generators, also known as wave energy converters, are devices designed to capture and convert wave energy into electricity. Wave generators operate based on several fundamental principles of wave energy conversion. The most common mechanism involves the conversion of the oscillatory motion of waves into mechanical or hydraulic energy, which is transformed into electricity through generators(McCormick, 2007). Key components of wave generators include wave absorbers, power take-off systems and electrical conversion units.

Figure 4.5: Wave Generator

Wave absorbers are the first point of contact between the incoming waves and the wave generator. They are designed to capture the energy from the waves and convert it into mechanical or hydraulic energy. Depending on the type of wave generator, wave absorbers may take various forms, such as oscillating water columns, floats, or submerged structures. Electrical generation units consist of generators and associated components necessary for converting mechanical energy into electricity. Many wave energy converters utilize buoyant structures to float on the surface of the water and interact with the incoming waves. Mooring and anchoring systems are essential for securing wave generators in place and preventing them from being displaced by wave action.

These systems typically involve anchors, cables and attachment points that connect the wave generator to the seabed or other fixed structures. Control and monitoring systems are responsible for optimizing the performance of wave generators and ensuring safe and efficient operation(Falcão, 2010). These systems include sensors for measuring wave height, frequency and direction as well as control algorithms for adjusting the operation of the device in response to changing environmental conditions. Understanding the interplay between these key components is essential for designing efficient and reliable wave energy converters capable of harnessing the potential of ocean waves for renewable energy generation.

Main types of wave generators consist of: oscillating water columns which utilize the vertical movement of water columns within a chamber to drive air turbines. As waves enter the chamber, the water level rises and falls, causing the air trapped above to oscillate and drive the turbine.

Point absorbers are buoyant devices tethered to the seabed, designed to absorb wave energy from multiple directions. These devices oscillate or rotate in response to wave motion, driving hydraulic or mechanical systems to generate electricity. Point absorbers are adaptable to various wave conditions and can be deployed in offshore or nearshore environments.

Attenuators, consist of multiple floating segments connected by flexible joints. As wave pass through the device, the segments move relative to each other, converting wave energy into mechanical motion. Attenuators offer high

energy capture efficiency and are suitable for locations with moderate to high wave intensity. Wave generators offer several advantages, including renewable resource availability, low environmental impact and potential for energy dependence(Falcão, 2010). However, there are challenges with these solutions such as high installations and maintenance costs, variable wave conditions and technological limitations.

4.2 Application to Wave Height

Our data included three times series of 7 years period (wind speed, wave height and period) and one time series of 1 year of solar irradiance. We will present for each time series the histogram of the historical series as well as the statistical characteristics of the historical and the simulated ones. We will also present each climacogram and the average values per month per hour of each dataset. Before applying the methodology we standardized the timeseries because the GHK cannot handle the periodicity and we created the dimensionless climacogram and estimated the first four statistical moments. The procedure that was followed to created the synthetic time series is, for each data set we calculated the first four statistical moments and the climacogram. These findings were the input of our model in order to create the synthetic time series via the Symmetric Moving Average scheme and the GHK model which is a special case of the HHK model. In the next sections we will present these findings. The generated time series have length of 30 years.

.

Table 4.1: Statistical Characteristics of Wave Height

	Historical	Synthetic
Average	0.97	0.98
Standard Deviation	0.77	0.77
Skewness	1.61	1.53
Kurtosis	3.38	3.60
Min Value	0.02	0.03
Max Value	5.94	7.35

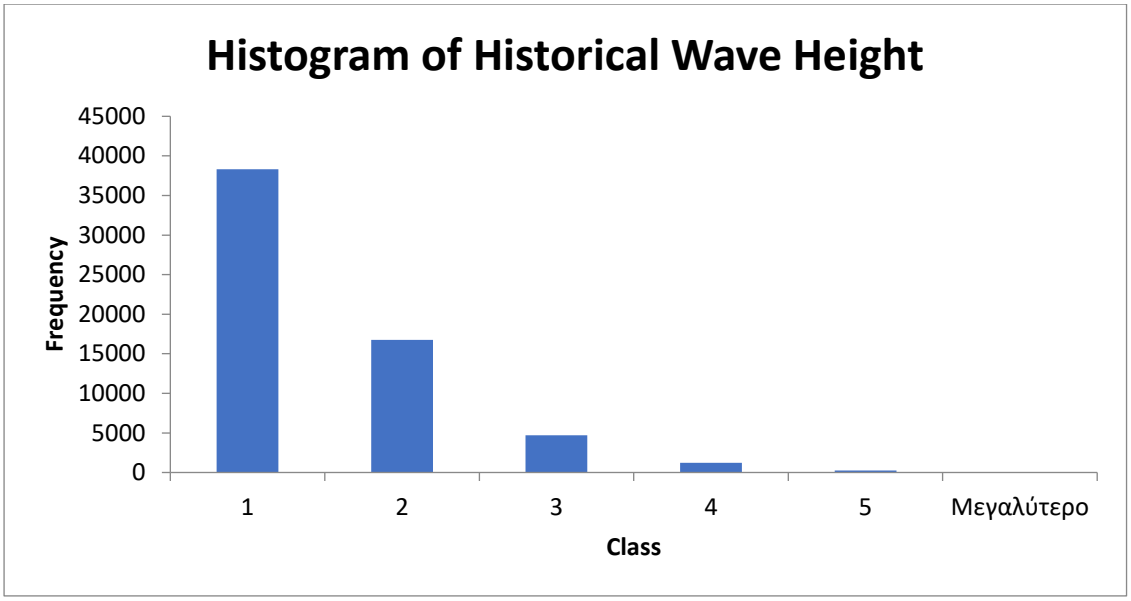

Figure 4.6: Histogram of Historical Wave Height

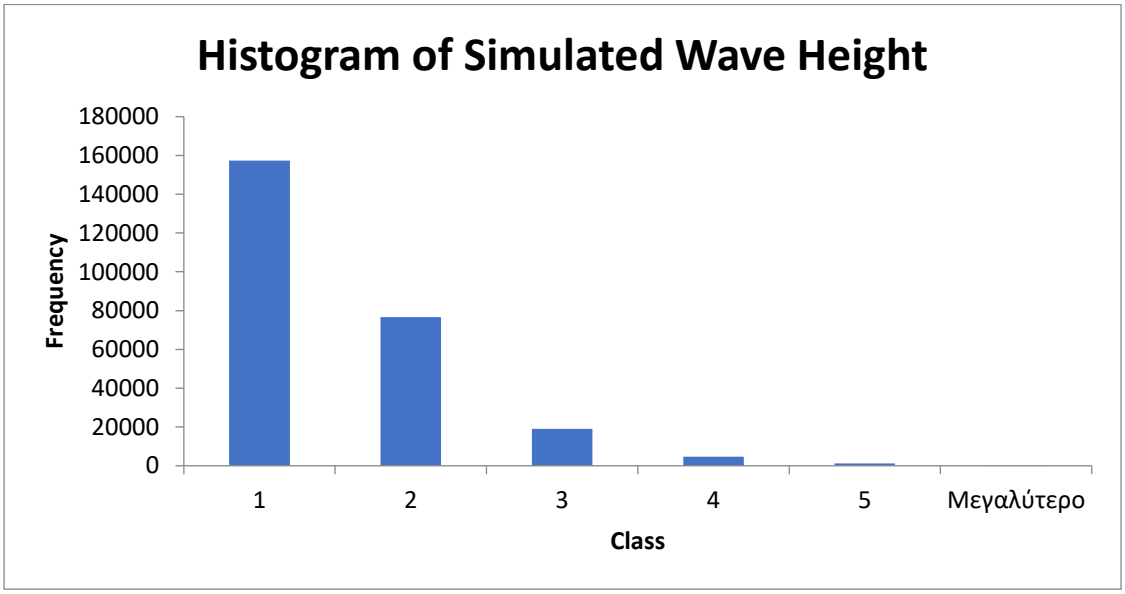

Figure 4.7: Histogram of Simulated Wave Height

As we can see from Table (4.1), the SMA scheme well preserved the first four central moments of the wave height dataset. The maximum value of the simulated wave height differed from the maximum value of the historical one and that is due to the greater length of the simulation comparing with the historical one. The main bodies of the histograms are the same except the maximum values, which in the case of the simulation we find larger values.

Table 4.2: Average Historical Wave Height per Month per Hour

	Jan	Feb	Mar	Apr	May	Jun	Jul	Sep	Aug	Oct	Nov	Dec
0	1.07	1.19	1.05	0.97	0.69	0.98	1.22	1.14	1.02	0.82	0.92	0.99
1	1.08	1.19	1.04	0.98	0.68	0.98	1.20	1.12	1.02	0.82	0.92	0.99
2	1.07	1.18	1.04	0.99	0.68	0.98	1.19	1.11	1.02	0.82	0.93	1.00
3	1.07	1.17	1.02	1.00	0.68	0.97	1.16	1.09	1.00	0.83	0.92	1.00
4	1.06	1.17	1.01	1.00	0.67	0.95	1.13	1.07	0.99	0.85	0.92	1.01
5	1.06	1.17	0.99	1.00	0.66	0.94	1.10	1.05	0.97	0.86	0.91	1.01
6	1.05	1.16	0.98	0.99	0.64	0.91	1.07	1.03	0.96	0.85	0.90	1.02
7	1.04	1.15	0.97	0.98	0.63	0.89	1.05	1.01	0.94	0.84	0.89	1.03
8	1.04	1.14	0.97	0.97	0.62	0.86	1.03	0.99	0.92	0.84	0.89	1.04
9	1.03	1.13	0.97	0.95	0.61	0.84	1.02	0.98	0.91	0.84	0.88	1.05
10	1.03	1.12	0.98	0.94	0.60	0.83	1.03	0.97	0.91	0.83	0.88	1.06
11	1.03	1.12	0.98	0.93	0.60	0.82	1.03	0.97	0.91	0.84	0.87	1.07
12	1.06	1.14	0.99	0.93	0.61	0.82	1.05	0.98	0.93	0.84	0.87	1.07
13	1.06	1.16	1.03	0.93	0.62	0.83	1.08	1.00	0.96	0.84	0.88	1.08
14	1.06	1.17	1.04	0.92	0.62	0.84	1.10	1.02	0.99	0.84	0.88	1.08
15	1.05	1.16	1.05	0.92	0.63	0.86	1.13	1.04	1.00	0.85	0.88	1.07
16	1.05	1.18	1.05	0.94	0.64	0.87	1.15	1.07	1.01	0.85	0.88	1.06
17	1.05	1.19	1.06	0.94	0.65	0.89	1.18	1.10	1.02	0.84	0.88	1.04
18	1.05	1.17	1.06	0.95	0.66	0.90	1.20	1.12	1.02	0.84	0.88	1.03
19	1.05	1.17	1.05	0.96	0.67	0.92	1.22	1.13	1.02	0.84	0.89	1.02
20	1.05	1.18	1.05	0.96	0.68	0.94	1.24	1.16	1.03	0.83	0.89	1.01
21	1.04	1.17	1.05	0.97	0.69	0.97	1.24	1.16	1.02	0.83	0.89	1.01
22	1.05	1.18	1.05	0.97	0.69	0.98	1.24	1.16	1.02	0.83	0.89	1.01
23	1.05	1.19	1.04	0.96	0.68	0.99	1.23	1.15	1.01	0.82	0.89	1.01

Table 4.3: Average Simulated Wave Height per Month per Hour

	Jan	Feb	Mar	Apr	May	Jun	Jul	Sep	Aug	Oct	Nov	Dec
0	1.03	1.18	1.03	1.01	0.68	1.01	1.22	1.24	1.01	0.84	0.95	0.93
1	1.04	1.17	1.01	1.01	0.68	1.00	1.20	1.22	1.01	0.84	0.96	0.94
2	1.04	1.15	1.02	1.02	0.69	1.00	1.18	1.20	0.99	0.84	0.97	0.95
3	1.03	1.14	1.00	1.03	0.70	0.99	1.16	1.17	0.98	0.85	0.97	0.95
4	1.02	1.15	0.99	1.02	0.69	0.98	1.11	1.15	0.98	0.87	0.96	0.96
5	1.02	1.15	0.98	1.01	0.69	0.96	1.09	1.14	0.97	0.88	0.94	0.96
6	1.00	1.16	0.97	1.02	0.67	0.94	1.06	1.12	0.96	0.88	0.93	0.98
7	0.99	1.15	0.96	1.01	0.65	0.92	1.04	1.09	0.94	0.88	0.93	1.00
8	0.99	1.14	0.97	1.00	0.63	0.88	1.02	1.07	0.92	0.89	0.92	1.00
9	0.99	1.14	0.98	0.99	0.63	0.86	1.01	1.06	0.90	0.89	0.91	1.01
10	0.99	1.13	0.98	0.99	0.62	0.85	1.01	1.06	0.89	0.88	0.89	1.03
11	1.00	1.14	0.99	0.97	0.63	0.83	1.02	1.07	0.89	0.87	0.87	1.03
12	1.02	1.16	1.00	0.97	0.64	0.83	1.05	1.07	0.91	0.87	0.88	1.04
13	1.01	1.17	1.04	0.97	0.64	0.84	1.08	1.09	0.94	0.88	0.89	1.04
14	1.00	1.17	1.06	0.97	0.65	0.86	1.11	1.11	0.97	0.88	0.90	1.04
15	0.99	1.18	1.07	0.96	0.66	0.87	1.13	1.13	1.00	0.87	0.90	1.03
16	0.99	1.20	1.06	0.97	0.66	0.89	1.16	1.15	1.01	0.87	0.89	1.00
17	1.00	1.21	1.07	0.97	0.66	0.91	1.18	1.17	1.02	0.87	0.90	0.99
18	1.00	1.20	1.06	0.99	0.67	0.93	1.20	1.21	1.02	0.86	0.90	0.98
19	1.00	1.20	1.06	1.00	0.69	0.95	1.22	1.21	1.02	0.85	0.90	0.97
20	1.00	1.19	1.05	1.01	0.69	0.97	1.24	1.25	1.02	0.87	0.91	0.96
21	0.99	1.19	1.05	1.00	0.70	1.00	1.24	1.26	1.02	0.86	0.90	0.97

22	1.00	1.19	1.05	1.00	0.69	1.01	1.24	1.26	1.01	0.85	0.91	0.95
23	1.00	1.18	1.01	1.01	0.68	1.02	1.22	1.25	1.01	0.84	0.92	0.95

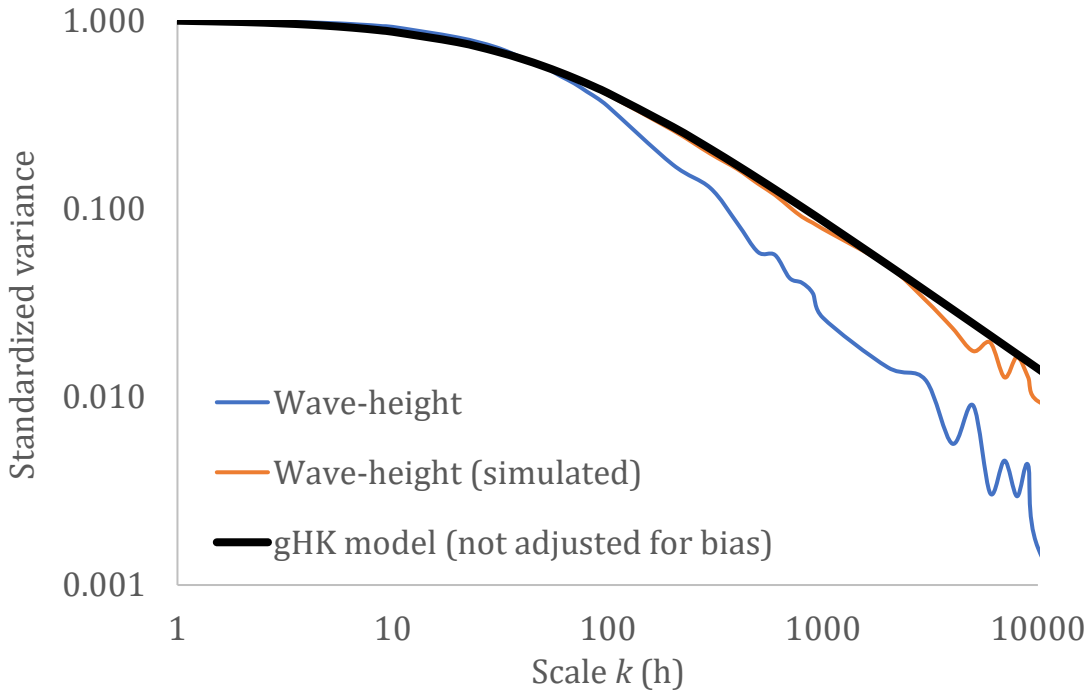

Figure 4.8: Climacogram of Wave Height

The GHK model was able to capture the climacogram of the synthetic time series. Any inconsistencies are met only in large scales and this is due to the length of the simulation. The greater the length of the simulation the better approximation of the climacogram. In our case a 30 year simulation was sufficient for the purpose of the project.

4.3 Application to Wave Period

Wave period data set consisted of 7 year and the length of the synthetic time series is 30 years. Comparing the statistical characteristics of the historical and simulated series we can see that the model was able to preserve the statistical characteristics in the synthetic time series.

Table 4.4: Statistical Characteristics of Wave Period

	Historical	Synthetic
Average	4.45	4.42
Standard Deviation	0.92	0.91
Skewness	0.75	0.79
Kurtosis	0.72	0.92
Min Value	2.46	1.9
Max Value	8.7	10.98

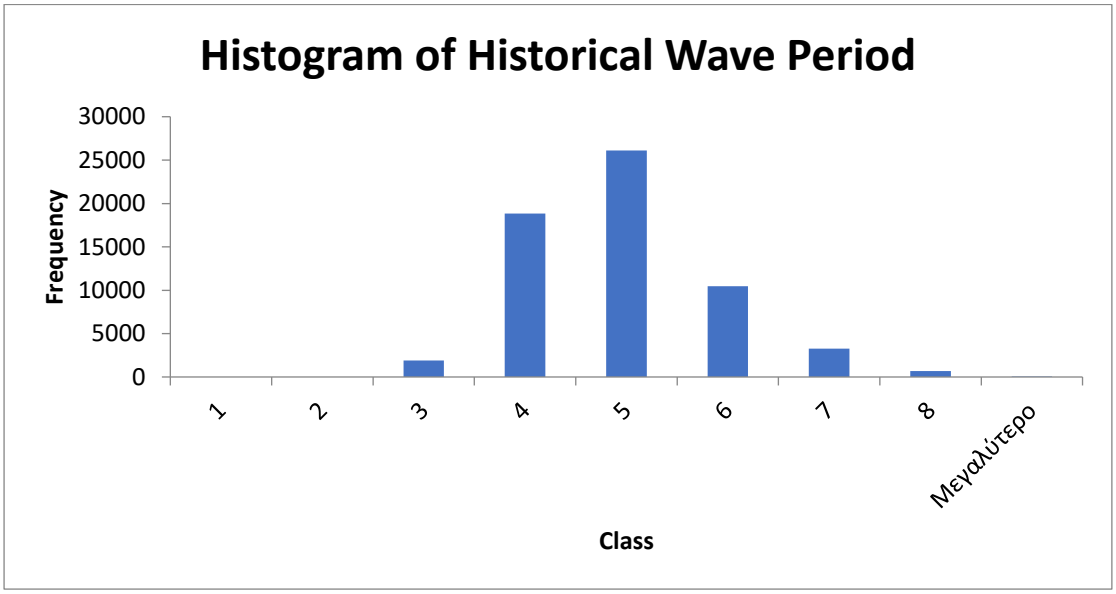

Figure 4.9: Histogram of Historical Wave Period

The maximum value of the simulated wave period differed from the maximum value of the historical one and that is due to the greater length of the simulation comparing with the historical one. The main bodies of the histograms are the same except the maximum values, which in the case of the simulation we find larger values.

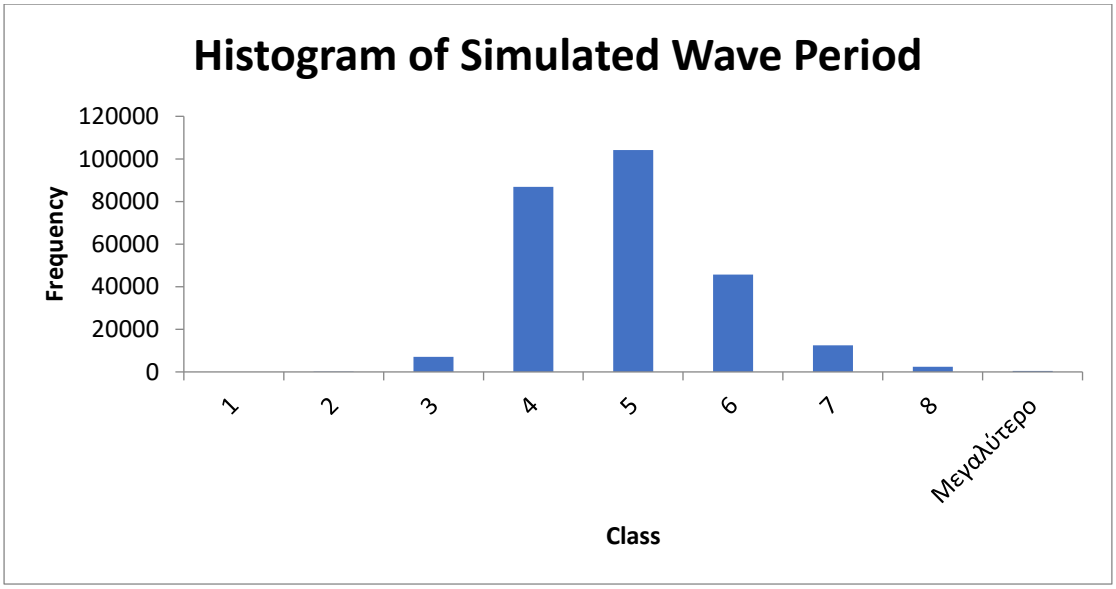

Figure 4.10: Histogram of Simulated Wave Period

Table 4.5: Average Historical Wave Period per Month per Hour

	Jan	Feb	Mar	Apr	May	Jun	Jul	Sep	Aug	Oct	Nov	Dec
0	4.56	4.66	4.53	4.52	4.20	4.27	4.48	4.50	4.43	4.42	4.42	4.44
1	4.56	4.65	4.52	4.55	4.19	4.28	4.48	4.50	4.44	4.42	4.41	4.43
2	4.56	4.64	4.52	4.57	4.19	4.29	4.47	4.49	4.44	4.41	4.42	4.43
3	4.57	4.63	4.53	4.61	4.19	4.31	4.47	4.49	4.45	4.41	4.41	4.45
4	4.57	4.64	4.53	4.63	4.18	4.32	4.47	4.48	4.46	4.43	4.41	4.46
5	4.58	4.63	4.53	4.65	4.17	4.34	4.46	4.48	4.46	4.45	4.41	4.48
6	4.59	4.66	4.53	4.65	4.15	4.35	4.46	4.49	4.46	4.47	4.41	4.50
7	4.61	4.65	4.51	4.64	4.14	4.34	4.46	4.49	4.45	4.48	4.43	4.52
8	4.61	4.64	4.51	4.62	4.12	4.32	4.45	4.47	4.43	4.48	4.43	4.53
9	4.62	4.64	4.51	4.61	4.11	4.30	4.44	4.46	4.41	4.49	4.44	4.54
10	4.63	4.63	4.50	4.59	4.11	4.27	4.44	4.43	4.39	4.48	4.43	4.56
11	4.62	4.62	4.47	4.56	4.09	4.23	4.42	4.40	4.38	4.49	4.43	4.57
12	4.61	4.61	4.45	4.53	4.08	4.20	4.41	4.37	4.37	4.48	4.41	4.57
13	4.61	4.61	4.47	4.51	4.07	4.17	4.39	4.34	4.37	4.46	4.39	4.58
14	4.61	4.62	4.48	4.49	4.05	4.15	4.39	4.33	4.39	4.46	4.39	4.59
15	4.61	4.64	4.48	4.48	4.05	4.14	4.39	4.33	4.39	4.45	4.39	4.59
16	4.60	4.67	4.49	4.48	4.05	4.13	4.40	4.35	4.40	4.45	4.39	4.59
17	4.59	4.69	4.51	4.47	4.06	4.13	4.41	4.38	4.41	4.45	4.39	4.58
18	4.58	4.69	4.52	4.49	4.07	4.14	4.43	4.40	4.42	4.46	4.40	4.58
19	4.58	4.69	4.52	4.50	4.10	4.16	4.45	4.42	4.42	4.46	4.41	4.56
20	4.57	4.68	4.53	4.51	4.13	4.18	4.47	4.45	4.43	4.46	4.41	4.54
21	4.56	4.68	4.53	4.53	4.16	4.21	4.49	4.48	4.43	4.45	4.40	4.52
22	4.56	4.68	4.53	4.52	4.18	4.25	4.49	4.49	4.43	4.45	4.40	4.50
23	4.55	4.68	4.52	4.52	4.19	4.26	4.49	4.50	4.43	4.45	4.39	4.48

Table 4.6: Average Simulated Wave Period per Month per Hour

	Jan	Feb	Mar	Apr	May	Jun	Jul	Sep	Aug	Oct	Nov	Dec
0	4.48	4.71	4.50	4.58	4.20	4.20	4.40	4.39	4.45	4.39	4.39	4.43
1	4.48	4.69	4.49	4.59	4.20	4.20	4.40	4.39	4.47	4.38	4.40	4.43
2	4.48	4.66	4.49	4.62	4.19	4.22	4.39	4.37	4.48	4.38	4.42	4.44
3	4.48	4.65	4.52	4.67	4.19	4.24	4.39	4.38	4.48	4.38	4.42	4.45
4	4.48	4.64	4.54	4.71	4.17	4.26	4.36	4.37	4.49	4.39	4.43	4.46
5	4.50	4.64	4.54	4.72	4.15	4.27	4.37	4.36	4.48	4.41	4.42	4.46
6	4.52	4.67	4.52	4.72	4.14	4.27	4.37	4.37	4.48	4.44	4.41	4.48
7	4.53	4.68	4.50	4.71	4.12	4.27	4.37	4.36	4.48	4.44	4.43	4.50
8	4.54	4.67	4.50	4.66	4.11	4.24	4.37	4.34	4.46	4.44	4.44	4.52
9	4.56	4.65	4.51	4.65	4.11	4.22	4.36	4.33	4.44	4.44	4.44	4.51
10	4.57	4.64	4.49	4.64	4.10	4.19	4.35	4.29	4.41	4.43	4.42	4.51
11	4.56	4.64	4.46	4.62	4.08	4.16	4.34	4.27	4.39	4.42	4.41	4.54
12	4.55	4.64	4.43	4.58	4.07	4.13	4.33	4.25	4.38	4.41	4.40	4.54
13	4.53	4.63	4.44	4.57	4.07	4.11	4.32	4.21	4.38	4.40	4.39	4.56
14	4.52	4.65	4.45	4.55	4.05	4.09	4.31	4.19	4.39	4.40	4.40	4.58
15	4.52	4.67	4.45	4.53	4.05	4.07	4.31	4.20	4.39	4.40	4.40	4.59
16	4.50	4.70	4.45	4.54	4.05	4.06	4.31	4.22	4.40	4.41	4.38	4.58
17	4.52	4.70	4.48	4.52	4.06	4.06	4.33	4.24	4.40	4.39	4.38	4.57
18	4.50	4.69	4.50	4.54	4.07	4.07	4.34	4.27	4.42	4.40	4.38	4.59
19	4.50	4.70	4.50	4.56	4.10	4.09	4.36	4.29	4.43	4.41	4.40	4.57
20	4.48	4.70	4.50	4.56	4.13	4.11	4.39	4.33	4.45	4.42	4.41	4.54
21	4.47	4.69	4.51	4.58	4.17	4.14	4.41	4.37	4.44	4.41	4.41	4.52
22	4.48	4.70	4.51	4.59	4.18	4.18	4.42	4.37	4.44	4.41	4.40	4.51
23	4.49	4.71	4.51	4.58	4.18	4.20	4.42	4.39	4.43	4.42	4.38	4.47

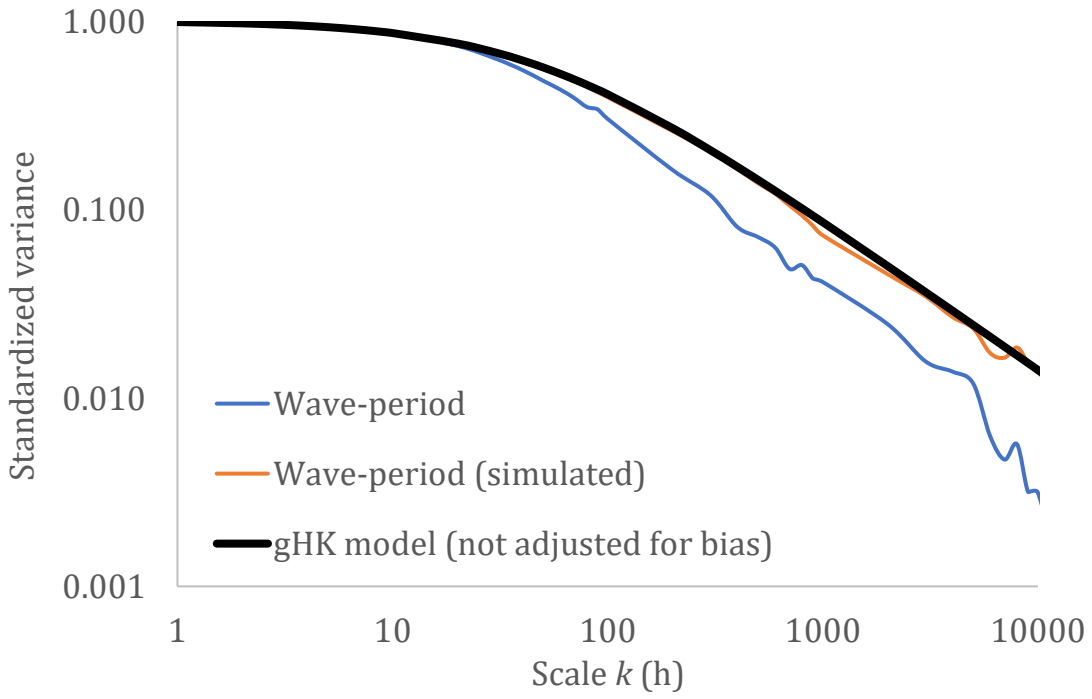

Figure 4.11: Climacogram of Wave Period

The model was able to capture the climacogram of the synthetic time series. Any inconsistencies are met only in large scales and this is due to the length of the simulation. The greater the length of the simulation the better approximation of the climacogram.

4.4 Application to Wind Speed

Wind Speed data set consisted of 7 year and the length of the synthetic time series is 30 years. Comparing the statistical characteristics of the historical and simulated series we can see that the model was able to preserve the statistical characteristics in the synthetic time series.

Table 4.7: Statistical Characteristics of Wind Speed

	Historical	Synthetic
Average	6.46	6.39
Standard Deviation	3.08	3.07
Skewness	0.32	0.13
Kurtosis	-0.23	0.12
Min Value	0.01	0.01
Max Value	19.49	20.60

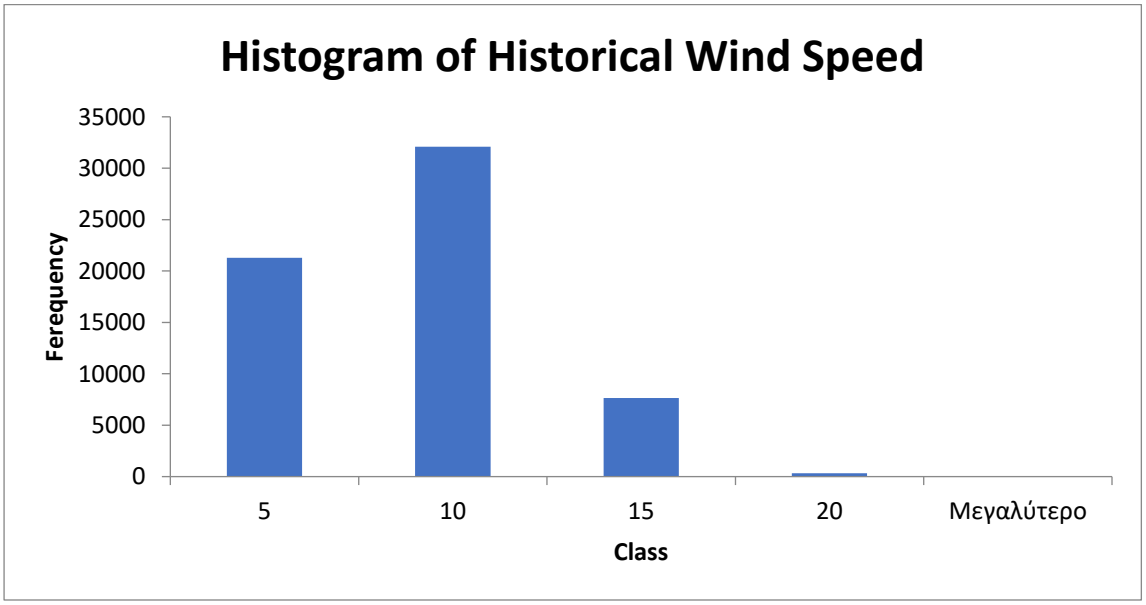

Figure 4.12: Histogram of Historical Wind Speed

Table (4.7), shows that the SMA scheme well preserved the first four central moments of the wind speed dataset. The main bodies of the histograms are the same except the maximum values, which in the case of the simulation we find larger values.

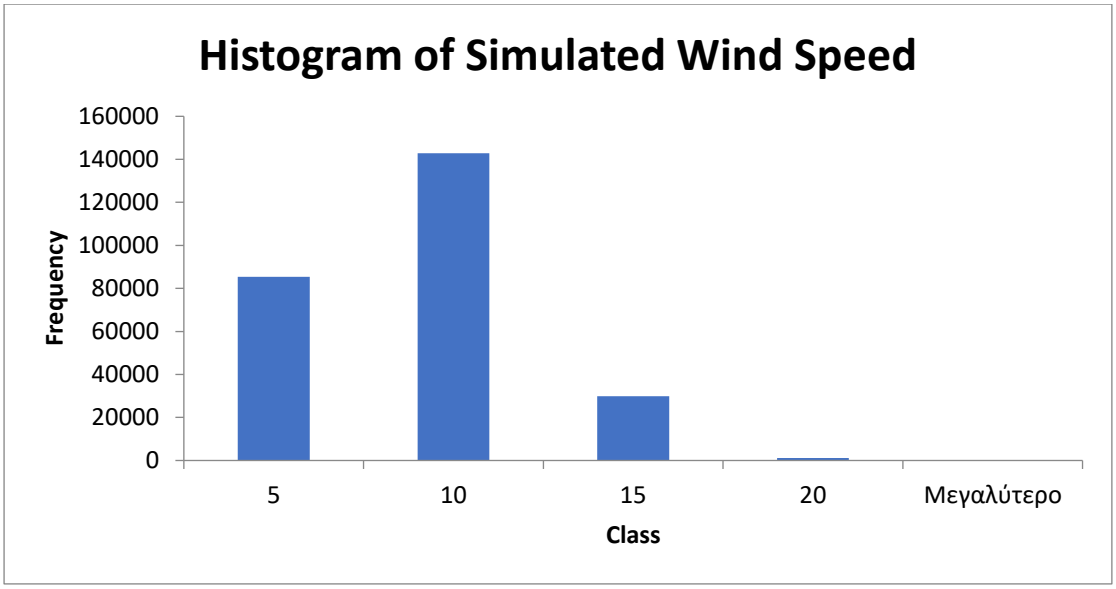

Figure 4.13: Daily Average Wind Speed

Table 4.8: Average Historical Wind Speed per Month per Hour

	Jan	Feb	Mar	Apr	May	Jun	Jul	Sep	Aug	Oct	Nov	Dec
0	7.33	7.56	6.70	6.37	5.15	6.32	7.52	7.20	6.83	6.31	6.55	7.03
1	7.14	7.38	6.53	6.22	4.96	6.12	7.32	7.01	6.66	6.13	6.38	6.88
2	6.98	7.25	6.39	6.09	4.79	5.95	7.13	6.82	6.50	5.97	6.23	6.76
3	6.87	7.15	6.30	6.00	4.67	5.79	6.95	6.65	6.35	5.84	6.10	6.67
4	6.81	7.11	6.25	5.93	4.58	5.65	6.78	6.49	6.23	5.74	6.01	6.62
5	6.79	7.10	6.25	5.89	4.54	5.53	6.63	6.33	6.13	5.67	5.94	6.59
6	6.81	7.14	6.30	5.89	4.53	5.44	6.49	6.19	6.05	5.65	5.91	6.60
7	6.80	7.09	6.30	5.88	4.54	5.54	6.64	6.34	6.10	5.69	5.90	6.60
8	6.83	7.08	6.34	5.91	4.58	5.66	6.81	6.50	6.18	5.75	5.92	6.62
9	6.89	7.11	6.42	5.98	4.69	5.82	7.01	6.68	6.29	5.84	5.98	6.68
10	6.99	7.17	6.54	6.11	4.84	6.00	7.22	6.88	6.42	5.95	6.08	6.76
11	7.12	7.27	6.70	6.29	5.03	6.20	7.45	7.10	6.58	6.09	6.21	6.88
12	7.28	7.40	6.90	6.50	5.25	6.42	7.70	7.33	6.76	6.24	6.37	7.03
13	7.15	7.29	6.74	6.39	5.17	6.40	7.68	7.28	6.69	6.14	6.27	6.88
14	7.05	7.21	6.62	6.31	5.12	6.39	7.66	7.24	6.63	6.07	6.19	6.76
15	6.98	7.17	6.55	6.27	5.10	6.41	7.65	7.21	6.59	6.02	6.13	6.68
16	6.95	7.18	6.52	6.26	5.12	6.43	7.65	7.18	6.57	5.99	6.11	6.65
17	6.95	7.23	6.52	6.28	5.18	6.48	7.65	7.16	6.56	5.99	6.10	6.66
18	6.98	7.33	6.56	6.33	5.26	6.53	7.66	7.15	6.57	6.02	6.12	6.70
19	6.96	7.27	6.49	6.26	5.18	6.47	7.62	7.13	6.58	6.00	6.12	6.68
20	6.97	7.26	6.46	6.21	5.12	6.42	7.59	7.13	6.60	6.01	6.14	6.69
21	7.01	7.28	6.45	6.19	5.09	6.39	7.57	7.13	6.64	6.04	6.19	6.75
22	7.08	7.35	6.49	6.20	5.08	6.38	7.56	7.14	6.68	6.09	6.28	6.84
23	7.18	7.45	6.58	6.24	5.10	6.38	7.56	7.16	6.74	6.17	6.40	6.95

Table 4.9: Average Simulated Wind Speed per Month per Hour

	Jan	Feb	Mar	Apr	May	Jun	Jul	Sep	Aug	Oct	Nov	Dec
0	6.53	6.34	6.45	6.47	6.41	6.58	6.98	6.97	7.02	7.07	7.13	6.92
1	6.43	6.19	6.27	6.33	6.22	6.38	6.78	6.76	6.88	6.86	6.98	6.74
2	6.28	6.01	6.09	6.17	6.03	6.25	6.61	6.62	6.70	6.72	6.84	6.60
3	6.13	5.90	5.94	6.04	5.88	6.12	6.49	6.47	6.55	6.60	6.75	6.53
4	6.08	5.84	5.83	5.93	5.76	6.01	6.38	6.42	6.45	6.49	6.74	6.51
5	6.09	5.81	5.73	5.87	5.67	5.86	6.32	6.32	6.41	6.52	6.73	6.46
6	6.10	5.81	5.67	5.75	5.59	5.73	6.22	6.30	6.41	6.57	6.75	6.48
7	6.11	5.86	5.73	5.81	5.66	5.80	6.26	6.36	6.36	6.59	6.75	6.47
8	6.14	5.92	5.82	5.92	5.74	5.92	6.38	6.37	6.39	6.63	6.81	6.50
9	6.18	5.99	5.92	6.05	5.89	6.08	6.44	6.44	6.52	6.69	6.85	6.52
10	6.29	6.12	6.06	6.20	6.04	6.26	6.60	6.58	6.66	6.79	6.91	6.61
11	6.48	6.33	6.26	6.40	6.23	6.44	6.77	6.74	6.84	6.96	7.06	6.76
12	6.69	6.54	6.45	6.61	6.44	6.66	6.95	7.00	7.06	7.13	7.27	6.90
13	6.55	6.46	6.32	6.53	6.41	6.62	6.88	6.87	6.96	6.99	7.11	6.73
14	6.46	6.40	6.25	6.48	6.36	6.52	6.79	6.74	6.85	6.90	6.97	6.63
15	6.42	6.39	6.29	6.46	6.38	6.49	6.76	6.65	6.78	6.83	6.88	6.63
16	6.40	6.42	6.35	6.50	6.38	6.45	6.71	6.66	6.72	6.78	6.83	6.65
17	6.42	6.45	6.41	6.50	6.37	6.47	6.69	6.63	6.66	6.80	6.81	6.67
18	6.48	6.50	6.48	6.56	6.38	6.51	6.71	6.64	6.70	6.81	6.84	6.71
19	6.42	6.42	6.43	6.52	6.35	6.51	6.74	6.57	6.70	6.78	6.76	6.65
20	6.36	6.39	6.40	6.47	6.31	6.50	6.74	6.61	6.70	6.80	6.77	6.61
21	6.34	6.39	6.36	6.45	6.35	6.52	6.75	6.67	6.75	6.84	6.80	6.67
22	6.36	6.33	6.38	6.44	6.40	6.54	6.81	6.78	6.79	6.89	6.88	6.73
23	6.39	6.31	6.42	6.43	6.41	6.55	6.89	6.87	6.92	6.96	6.96	6.81

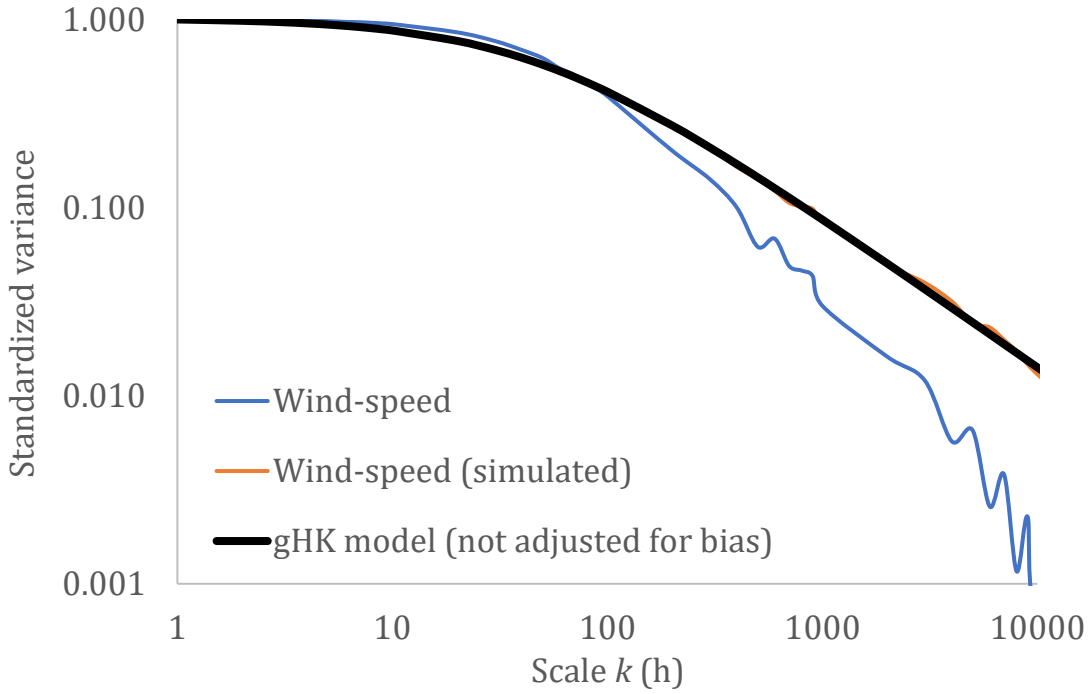

Figure 4.14: Climacogram of Wind Speed

30 years of simulation for the wind speed data set were adequate and the GHK model captured very well the synthetic time series.

4.5 Application to Solar Irradiance

Solar Irradiance data set consisted of 1 year and the length of the synthetic time series is 30 years. Comparing the statistical characteristics of the historical and simulated series we can see that the model was able to preserve the statistical characteristics in the synthetic time series

Table 4.10: Statistical Characteristics of Solar Irradiance

	Historical	Synthetic
Average	205.82	207.62
Standard Deviation	290.21	292.88
Skewness	1.21	1.20
Kurtosis	0.09	0.26
Min Value	0.02	0.01
Max Value	1023.5	1342.49

As aforementioned the main bodies of the histograms are the same except the maximum values, which in the case of the simulation of the solar irradiance we find larger values.

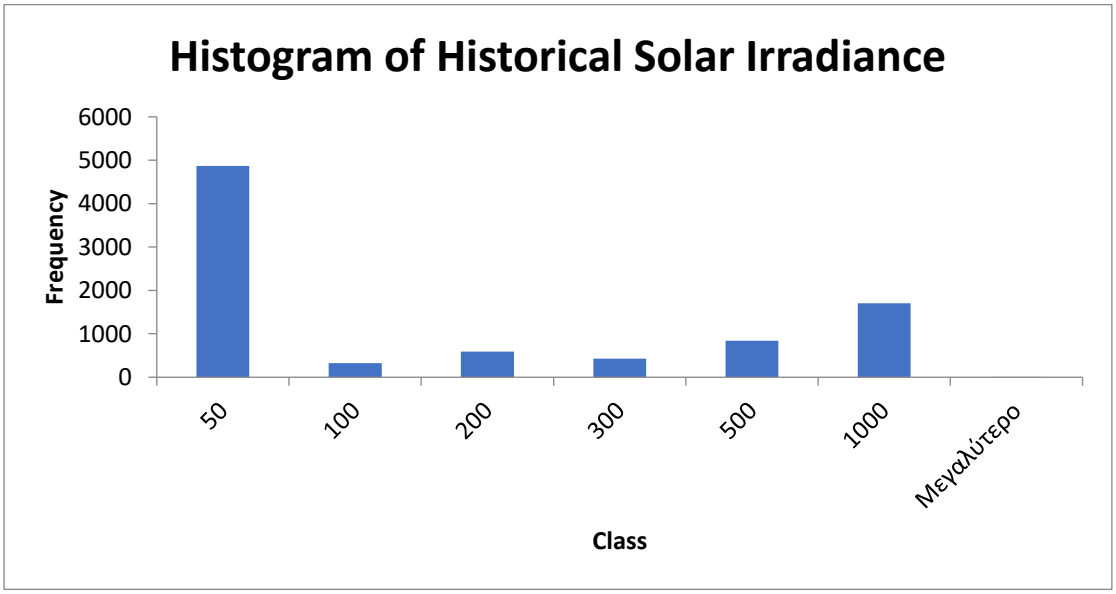

Figure 4.15: Histogram of Historical Solar Irradiance

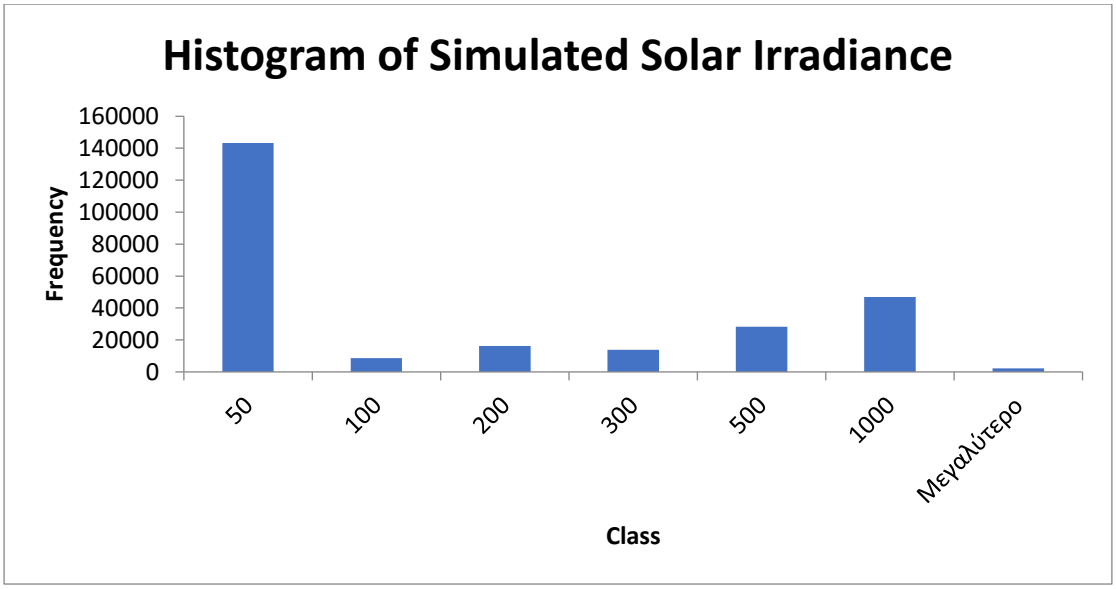

Figure 4.16: Daily Average Solar Irradiance

Table 4.11: Average Historical Solar Irradiance per Month per Hour

	Jan	Feb	Mar	Apr	May	Jun	Jul	Sep	Aug	Oct	Nov	Dec
0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
2	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
3	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
4	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
5	0.1	0.1	0.1	0.1	0.1	0.4	0.2	0.1	0.1	0.1	0.1	0.1
6	0.1	0.1	0.1	2.3	17.2	27.7	15.3	3.8	0.4	0.1	0.1	0.2
7	0.1	0.1	6.5	62.2	122.2	169.1	133.0	78.6	29.8	6.1	0.4	0.7
8	3.8	16.7	89.2	222.0	274.3	354.9	313.8	253.9	171.1	96.3	30.7	9.0
9	80.1	112.3	264.7	418.2	425.4	534.1	496.7	451.7	351.3	251.8	139.8	97.7
10	201.3	240.2	426.8	604.0	535.9	712.2	676.1	637.5	511.3	402.0	250.2	220.7
11	319.5	374.9	570.7	726.4	671.8	845.1	778.6	774.9	612.8	531.5	370.4	312.7
12	375.2	436.4	648.4	792.1	736.1	909.7	870.1	872.9	728.7	537.4	425.4	360.7
13	412.6	478.1	708.1	774.8	764.8	915.6	914.4	902.8	759.0	568.5	410.4	376.3
14	415.9	470.5	657.9	694.0	704.7	907.8	857.7	840.6	735.7	571.9	358.7	344.4
15	348.1	409.7	588.2	619.9	645.1	795.9	772.1	755.0	592.6	456.1	279.6	270.0
16	228.3	296.8	459.3	527.0	570.2	674.9	640.9	615.2	460.0	315.2	185.1	171.3
17	106.4	164.3	281.8	373.7	406.2	494.6	511.9	455.3	307.6	151.4	73.8	69.9
18	13.7	45.5	124.2	192.7	241.6	320.6	338.8	260.3	113.7	30.4	3.3	9.8
19	0.6	0.8	13.6	47.9	88.5	151.6	152.8	84.5	12.6	0.3	0.1	1.5
20	0.1	0.1	0.1	0.8	8.1	25.2	23.1	4.0	0.1	0.1	0.1	0.1
21	0.1	0.1	0.1	0.1	0.1	0.2	0.2	0.1	0.1	0.1	0.1	0.1
22	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
23	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1

Table 4.12: Average Simulated Solar Irradiance per Month per Hour

	Jan	Feb	Mar	Apr	May	Jun	Jul	Sep	Aug	Oct	Nov	Dec
0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
2	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
3	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
4	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
5	0.1	0.1	0.1	0.1	0.1	0.4	0.2	0.1	0.1	0.1	0.1	0.1
6	0.1	0.1	0.1	2.6	17.0	27.7	16.1	4.0	0.4	0.1	0.1	0.2
7	0.1	0.1	7.4	66.5	121.8	169.2	136.0	79.5	30.9	6.2	0.4	0.8
8	3.8	17.0	95.9	230.4	271.7	355.1	319.3	255.4	175.7	96.4	29.4	9.2
9	81.3	113.3	279.1	429.5	421.9	534.2	506.5	453.0	359.7	252.0	134.2	98.5
10	203.3	243.6	448.7	618.1	531.9	713.1	682.2	638.4	523.5	402.1	239.6	222.7
11	322.1	380.3	597.6	743.3	670.6	845.0	795.4	776.7	629.3	531.1	358.2	316.1
12	378.9	440.5	678.2	809.7	741.1	908.6	885.8	873.3	744.1	535.9	409.5	364.9
13	416.7	482.6	738.2	799.6	769.6	914.2	926.9	903.0	773.5	568.1	393.5	378.1
14	419.4	474.8	686.4	725.1	711.4	907.1	878.3	840.5	747.3	570.2	343.6	344.9
15	350.9	410.9	613.8	649.4	653.0	795.6	794.3	755.7	605.4	452.4	267.1	271.2
16	230.8	296.6	480.1	553.3	574.6	674.8	662.3	615.8	472.2	312.8	176.0	172.7
17	108.3	165.1	297.2	394.2	407.6	493.5	526.5	455.7	315.6	150.8	69.9	71.6
18	14.5	46.0	131.4	205.6	242.7	320.2	346.6	261.5	118.1	30.2	3.1	11.1
19	0.7	0.8	14.7	52.0	89.3	151.8	157.5	85.0	13.5	0.3	0.1	1.8
20	0.1	0.1	0.1	0.9	8.3	25.2	24.1	4.1	0.1	0.1	0.1	0.1
21	0.1	0.1	0.1	0.1	0.1	0.2	0.2	0.1	0.1	0.1	0.1	0.1
22	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
23	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1

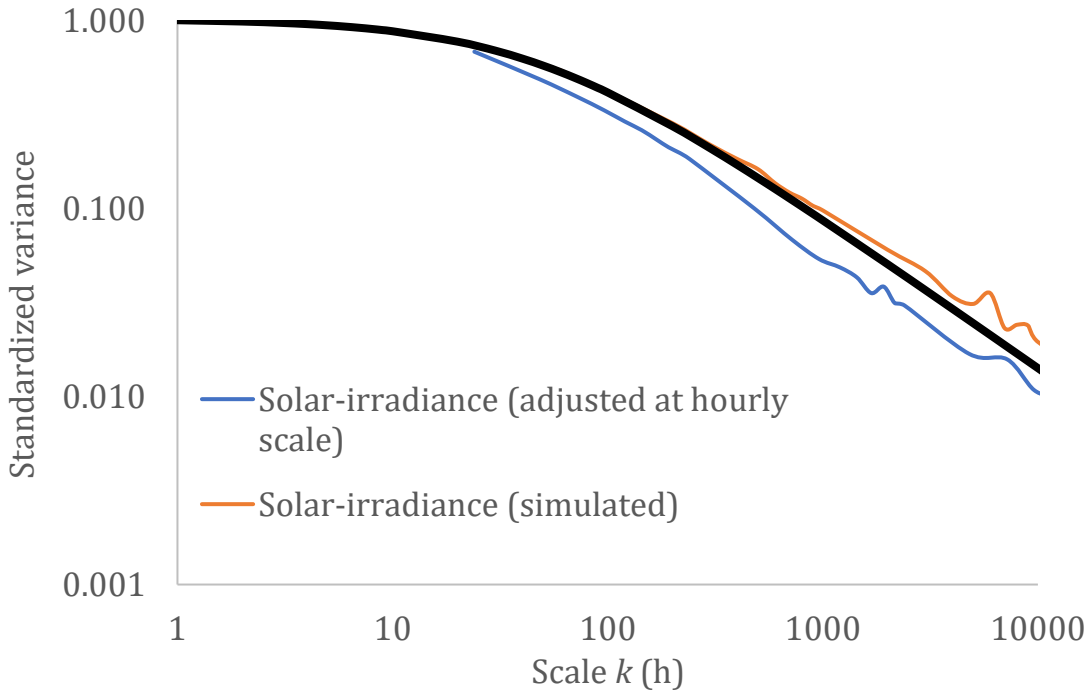

Figure 4.17: Climacogram of Solar Irradiance

Finally in the case of the Solar Irradiance the GHK model , performed well and from the distribution of the synthetic data we can conclude that the SMA scheme well captured the first four statistical moments such as the intermittent behavior of the dataset.

5. Conclusions

- We presented an extension of the Symmetric Moving Average Scheme in order to preserve the first four statistical moments including the coefficient of kurtosis in the simulation.
- The synthetic time series of the GHK model were produced by the SMA scheme.
- Before applying the GHK model, all of the datasets were standardized and estimated the climacogram and the statistical moments of the standardized time series because the GHK model cannot handle the periodicities of the examined data.
- All the datasets exhibit long range dependence, which means clustering of low or high values (Hurst – Kolmogorov Dynamics) with Hurst index of $H = 0.6$

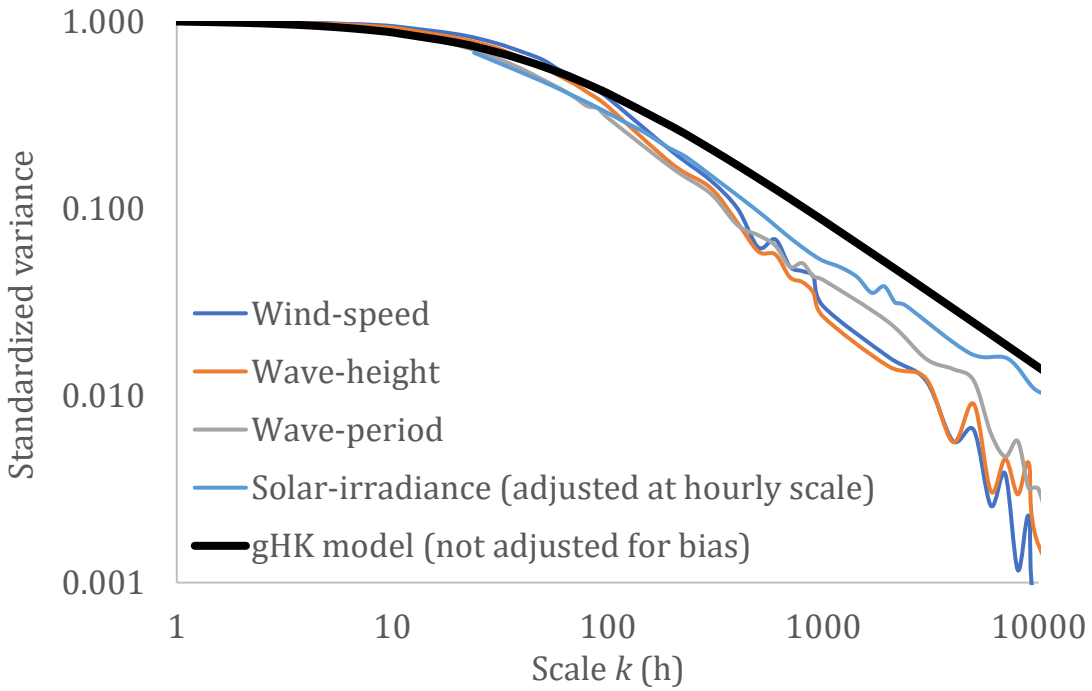

Figure 5.1: Climacogram of all processes

- The GHK model was able to capture the climacogram of the synthetic time series, only to large scales there was a deviation which could be dealt by increasing the length of the simulation.
- The preservation up to the fourth moment was adequate, since preservation of additional moments slightly improve the distribution simulation.
- As we can see from the statistical characteristics of historical and synthetic time series (Table 5.1) the Symmetric Moving average scheme well preserved the first four statistical moments.

Table 5.1: Statistical Characteristics of all datasets

	Wind (H)	Wind (S)	Wave H. (H)	Wave H. (S)	Wave P. (H)	Wave P. (S)	Solar (H)	Solar (S)
Average	6.46	6.39	0.97	0.98	4.45	4.42	205.82	207.62
St. Deviation	3.08	3.07	0.77	0.77	0.92	0.91	290.21	292.88
Skewness	0.32	0.13	1.62	1.53	0.75	0.79	1.21	1.20
Kurtosis	-0.32	0.12	3.39	3.60	0.72	0.92	0.09	0.26
Min Value	0.01	0.01	0.02	0.03	2.46	1.9	0.02	0.01
Max Value	19.49	20.60	5.94	7.35	8.7	10.98	1023.5	1342.49

- The statistical moments of the historical and simulated are close, but, the maximum values of the synthetic series are greater than those of the historical ones and this is due to the length of the simulation being greater than the historical ones.
- The extended SMA scheme as we can see also preserved the intermittent behavior of the processes.
- The distribution of the data is approximated to a desired degree rather than precisely preserved.

A more precise way on estimating the desired distribution of the historical data is proposed by (Koutsoyiannis, 2021) using K-moments (Koutsoyiannis, 2023, 2019) which provide a framework for model fitting, using the entire dataset rather than relying on a few moments. With this method we estimate the K-moments of a dataset and we choose a marginal distribution for the process based on the K-moments and then we calculate theoretically the classical moments of the process. K-moments offer a better approximation of a distribution rather than classical moments and are advantageous in the stochastic theory and its applications in geophysical phenomena especially when dealing with extremes.

6.References

Bloomfield, P., 2000. Fourier analysis of time series: an introduction, 2nd ed. ed, Wiley series in probability and statistics. Applied probability and statistics section. Wiley, New York.

Box, G.E.P., Jenkins, G.M., Reinsel, G.C., 2008. Time series analysis: forecasting and control, Fourth edition. ed. J. Wiley & Sons, Hoboken, N.J.

Bracewell, R.N., 2000. The Fourier transform and its applications, 3rd ed. ed, McGraw-Hill series in electrical and computer engineering. McGraw Hill, Boston.

Brockwell, P.J., Davis, R.A., 2010. Introduction to time series and forecasting, Second edition, (corrected at 8th. printing 2010). ed, Springer texts in statistics. Springer, New York, NY.

David, H.A., Nagaraja, H.N., 2003. Order statistics, 3rd ed. ed. John Wiley, Hoboken, N.J.

Dimitriadis, P., Koutsoyiannis, D., 2018. Stochastic synthesis approximating any process dependence and distribution. Stoch. Environ. Res. Risk Assess. 32, 1493–1515. https://doi.org/10.1007/s00477-018-1540-2

Dimitriadis, P., Koutsoyiannis, D., 2015. Climacogram versus autocovariance and power spectrum in stochastic modelling for Markovian and Hurst–Kolmogorov processes. Stoch. Environ. Res. Risk Assess. 29, 1649–1669. https://doi.org/10.1007/s00477-015-1023-7

Falcão, A.F.D.O., 2010. Wave energy utilization: A review of the technologies. Renew. Sustain. Energy Rev. 14, 899–918. https://doi.org/10.1016/j.rser.2009.11.003

Hosking, J.R.M., 1990. L-Moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics. J. R. Stat. Soc. Ser. B Methodol. 52, 105–124.

Hurst, H.E., 1951. Long-Term Storage Capacity of Reservoirs. Trans. Am. Soc. Civ. Eng. 116, 770–799. https://doi.org/10.1061/TACEAT.0006518

Johnson, N.L., Kemp, A.W., Kotz, S., 2005. Univariate discrete distributions, 3rd ed. ed. Wiley, Hoboken, N.J.

Karlin, S., Taylor, H.M., 1975. A first course in stochastic processes, 2d ed. ed. Academic Press, New York.

Kenu E. Sarah, University of Port Harcourt, 2020. A Review of Solar Photovoltaic Technologies. Int. J. Eng. Res. V9, IJERTV9IS070244. https://doi.org/10.17577/IJERTV9IS070244

Kolmogorov, A.N., 1940. Wiener spirals and some other interesting curves in a Hilbert space. Dokl. Akad. Nauk SSSR, 26, 115-118.

Konstantinidis, E.I., Botsaris, P.N., 2016. Wind turbines: current status, obstacles, trends and technologies. IOP Conf. Ser. Mater. Sci. Eng. 161, 012079. https://doi.org/10.1088/1757-899X/161/1/012079

Koopmans, L.H., 1995. The spectral analysis of time series, 2nd ed. ed, Probability and mathematical statistics. Academic Press, San Diego.

Körner, T.W., 2004. Fourier analysis, 1. paperback ed., 8. print. ed. Cambridge Univ. Pr, Cambridge.

Koutsoyiannis, D., 2023. Knowable Moments in Stochastics: Knowing Their

Advantages. Axioms 12, 590. https://doi.org/10.3390/axioms12060590
Koutsoyiannis, D., 2021. Stochastics of Hydroclimatic Extremes - A Cool Look at Risk. Kallipos, Open Academic Editions. https://doi.org/10.57713/KALLIPOS-1
Koutsoyiannis, D., 2019. Knowable moments for high-order stochastic characterization and modelling of hydrological processes. Hydrol. Sci. J. 64, 19–33. https://doi.org/10.1080/02626667.2018.1556794
Koutsoyiannis, D, 2017. Encolpion of stochastics.
Koutsoyiannis, Demetris, 2017. Entropy Production in Stochastics. Entropy 19, 581. https://doi.org/10.3390/e19110581
Koutsoyiannis, D., 2016. Generic and parsimonious stochastic modelling for hydrology and beyond. Hydrol. Sci. J. 61, 225–244. https://doi.org/10.1080/02626667.2015.1016950
Koutsoyiannis, D., 2011. Hurst-Kolmogorov Dynamics and Uncertainty1: Hurst-Kolmogorov Dynamics and Uncertainty. JAWRA J. Am. Water Resour. Assoc. 47, 481–495. https://doi.org/10.1111/j.1752-1688.2011.00543.x
Koutsoyiannis, D., 2006. Nonstationarity versus scaling in hydrology. J. Hydrol. 324, 239–254. https://doi.org/10.1016/j.jhydrol.2005.09.022
Koutsoyiannis, D., 2003. Climate change, the Hurst phenomenon, and hydrological statistics. Hydrol. Sci. J. 48, 3–24. https://doi.org/10.1623/hysj.48.1.3.43481
Koutsoyiannis, D., 2002. The Hurst phenomenon and fractional Gaussian noise made easy. Hydrol. Sci. J. 47, 573–595. https://doi.org/10.1080/02626660209492961
Koutsoyiannis, D., 2000. A generalized mathematical framework for stochastic simulation and forecast of hydrologic time series. Water Resour. Res. 36, 1519–1533. https://doi.org/10.1029/2000WR900044
Lombardo, F., Volpi, E., Koutsoyiannis, D., Papalexiou, S.M., 2014. Just two moments! A cautionary note against use of high-order moments in multifractal models in hydrology. Hydrol. Earth Syst. Sci. 18, 243–255. https://doi.org/10.5194/hess-18-243-2014
Mandelbrot, B.B., Van Ness, J.W., 1968. Fractional Brownian Motions, Fractional Noises and Applications. SIAM Rev. 10, 422–437. https://doi.org/10.1137/1010093
Mandelbrot, B.B., Wallis, J.R., 1969a. Computer Experiments With Fractional Gaussian Noises: Part 1, Averages and Variances. Water Resour. Res. 5, 228–241. https://doi.org/10.1029/WR005i001p00228
Mandelbrot, B.B., Wallis, J.R., 1969b. Computer Experiments with Fractional Gaussian Noises: Part 2, Rescaled Ranges and Spectra. Water Resour. Res. 5, 242–259. https://doi.org/10.1029/WR005i001p00242
Mandelbrot, B.B., Wallis, J.R., 1969c. Computer Experiments with Fractional Gaussian Noises: Part 3, Mathematical Appendix. Water Resour. Res. 5, 260–267. https://doi.org/10.1029/WR005i001p00260
McCormick, M.E., 2007. Ocean wave energy conversion, Dover ed. ed. Dover Publications, Mineola, N.Y.
Osgood, B., 2019. Lectures on the Fourier transform and its applications, Pure

and applied undergraduate texts. American Mathematical Society, Providence, Rhode Island.
Papoulis, A., Pillai, S.U., 2009. Probability, random variables, and stochastic processes, 4. ed., internat. ed., Nachdr. ed. McGraw-Hill, Boston, Mass.
Salas, J.D. (Ed.), 1980. Applied modeling of hydrologic time series. Water Resources Publications, Littleton, Colo.
Salas, J.D., Boes, D.C., Smith, R.A., 1982. Estimation of ARMA Models with seasonal parameters. Water Resour. Res. 18, 1006–1010. https://doi.org/10.1029/WR018i004p01006
Salas, J.D., Obeysekera, J.T.B., 1982. ARMA Model identification of hydrologic time series. Water Resour. Res. 18, 1011–1021. https://doi.org/10.1029/WR018i004p01011
Smith, P.J., 1995. A Recursive Formulation of the Old Problem of Obtaining Moments from Cumulants and Vice Versa. Am. Stat. 49, 217. https://doi.org/10.2307/2684642
Stark, H., Woods, J.W., Stark, H., 2012. Probability, statistics, and random processes for engineers, 4th ed. ed. Pearson, Boston.
Tao, P.-C., Delleur, J.W., 1976. Seasonal and Nonseasonal ARMA Models in Hydrology. J. Hydraul. Div. 102, 1541–1559. https://doi.org/10.1061/JYCEAJ.0004637
Vodapally, S.N., Ali, M.H., 2022. A Comprehensive Review of Solar Photovoltaic (PV) Technologies, Architecture, and Its Applications to Improved Efficiency. Energies 16, 319. https://doi.org/10.3390/en16010319
Wagner, H.-J., 2018. Introduction to wind energy systems. EPJ Web Conf. 189, 00005. https://doi.org/10.1051/epjconf/201818900005
Zaidi, B. (Ed.), 2018. Solar Panels and Photovoltaic Materials. InTech. https://doi.org/10.5772/intechopen.72061
ΧΡΙΣΤΟΔΟΥΛΟΥ, Γ., n.d. ΕΠΙ ΤΗΣ ΓΕΩΛΟΓΙΑΣ ΤΗΣ ΝΗΣΟΥ ΑΣΤΥΠΑΛΑΙΑΣ.