Stochastic weather generator for the design and reliability evaluation of desalination systems with Renewable Energy Sources

Pierre Ailliot, Marie Boutigny, Eftichis Koutroulis, Athanasios Malisovas, Valérie Monbet

Stochastic Weather Generator for the Design and Reliability Evaluation of Desalination Systems with Renewable Energy Sources

Pierre Ailliot[a,*], Marie Boutigny[a,1], Eftichis Koutroulis[c], Athanasios Malisovas[c], Valérie Monbet[d]

[a] *Univ Brest, CNRS, LMBA - UMR 6205, Brest, France*
[b] *Eau du Ponant, Brest, France*
[c] *School of Electrical and Computer Engineering, Technical University of Crete, Chania GR-73100, Greece*
[d] *Univ Rennes, CNRS, IRMAR - UMR 6625, F-35000 Rennes, France*

Abstract

The operation of Renewable Energy Sources (RES) systems is highly affected by the continuously changing meteorological conditions and the design of a RES system has to be robust to the unknown weather conditions that it will encounter during its lifetime. In this paper, the use of Stochastic Weather Generators (SWGENs) is introduced for the optimal design and reliability evaluation of hybrid Photovoltaic / Wind-Generator (PV-W/G) systems providing energy to desalination plants. A SWGEN is proposed, which is based on parametric Markov-Switching Auto-Regressive (MSAR) models and is capable to simulate realistic hourly multivariate time series of solar irradiance, temperature and wind speed of the target installation site. Numerical results are presented, demonstrating that: (i) SWGENs enable to evaluate the reliability of RES-based desalination plants during their operation over a 20 years lifetime period and (ii) using an appropriate time series simulated with a SWGEN as input to the design optimization process results in a RES-based desalination plant configuration with higher reliability compared to the configurations derived when the other types of meteorological datasets are used as input to the design optimization process.

[*] Corresponding author

Keywords: Renewable Energy Sources, desalination, Stochastic Weather Generators, Markov-Switching Autoregressive models, non-parametric resampling, design optimization.

Highlights

- Renewable Energy Sources (RES) systems are highly sensitive to weather conditions
- Stochastic weather generators (SWGENs) permit to generate realistic weather conditions and enrich available meteorological datasets
- SWGENs enable to estimate the reliability of RES systems and design RES systems which are robust to climate variability

1. Introduction

Nowadays, Renewable Energy Sources (RES) are widely used for energy production purposes across the World in order to enhance the energy supply security and reduce the environmental impact of fossil-fuel-based energy sources [1]. RES systems are currently used in a wide variety of applications ranging from autonomous systems covering the electric energy requirements of buildings or isolated electric loads (e.g. telecommunication stations, measurement systems etc.) to Smart Grids and Microgrids (see e.g. [2, 3, 4]). The combination of RES and desalination technologies has also been extensively studied, as a means of solving the potable water scarcity problem of major areas of the World (e.g. Mediterranean region, Middle-East etc.) [5]. Both the performance and operational reliability of RES systems is highly sensitive to the continuously changing weather conditions, exhibiting hourly, daily and yearly variations during their prolonged lifetime period (typically > 20 years).

RES systems are designed considering historic meteorological data of the weather conditions that prevail at the target installation site [6]. These data affect the structure of the resulting configuration of the RES system in terms of the number and size of the energy production and storage devices which are required for covering a predetermined energy demand. Therefore, for the successful operation of a RES system, it is crucial to use the appropriate set of meteorological data during its design phase. The input dataset must resemble the actual meteorological conditions prevailing at the target installation field during the lifetime period of the RES system.

Due to the computational cost of running RES system models, a usual approach to design a RES system is based on the Typical Meteorological Year (TMY) – also called Test Reference Year. A TMY is defined as a concatenation of typical months where each month is selected among the available data set so that it is the most representative of the meteorological conditions of that month. One of the first works on this subject is the one of [7], where the comparison of the monthly average value of the meteorological parameter of interest (e.g. solar irradiance, temperature, wind speed, etc.) is proposed for that purpose. The most widely used approach is the Sandia National Laboratory method, presented in [8] for solar energy, that uses the Finkelstein-Schafer metric, based on cumulative distribution functions. Many other methods have been developed, such as the Danish method presented in [9], which has been developed for the design of buildings and calculates a distance based on mean and standard deviation, or its variation presented in [10], which has been developed for solar irradiance and includes a metric based on the distributions with the Kolmogorov-Smirnov distance. To handle several meteorological variables simultaneously, the most common approach consists in using a weighted sum of the distances measured for each meteorological variable. TMYs have been mainly used in the field of energy systems [11]. An expected drawback occurring when using TMY as an input to dynamic energy simulation is that the TMY will not reproduce the actual year-to-year fluctuations.

Stochastic Weather Generators (SWGENs) give the ability to estimate climate-related risks by simulating long sequences of weather, consistent with specific aspects of climate variability [12]. The simulated sequences of meteorological variables (e.g. rainfall, wind speed and direction, temperature, solar radiation, etc.) are typically used as inputs into complex environmental and ecosystem models. They have historically been used for hydrological applications [13, 14] with a strong interest for rainfall generation – the main input of run-off simulations – and agronomy for crop-growth modeling [15]. Other examples can be found in environmental management fields such as ecosystem modeling [16], or pollution modeling [17]. In building design, where TMYs are widely used, SWGENs have been applied either for generating long time series based on short or incomplete data records in order to calculate the TMY [18], or for assessing the impact of climate changes on buildings [19].

SWGENs have also become a useful tool in wind energy production studies. For example in [20, 21], wind generation is used for sizing and assessing the reliability of wind power systems, demonstrating the value of SWGENs in terms of capturing the uncertainty of wind conditions. The work of [22] showed the capacity of SWGENs to extend available records of wind conditions over short periods to statistically significant data that can be used then as inputs for simulating the operation of a wind energy conversion system. However to the authors knowledge, SWGENs have not yet been been applied for studying the performance of other types of RES systems. In order to fill this gap, the use of SWGENs is introduced in this paper for the first time in the existing literature, for the optimal design and reliability evaluation of hybrid Photovoltaic (PV) / Wind-Generator (W/G) systems which cover the electric energy demands of desalination plants.

Existing SWGENs can be broadly divided into parametric and non-parametric (or resampling) methods. Many parametric models have been proposed for weather variables in the literature. Most models were developed for rainfall, which is a key variable for many applications such as hydrology and agriculture. As concerns the variables of interest for RES applications, there is also a literature available on univariate models for wind speed U (see [23] for a review), temperature T [24] and solar radiation SR [25] time series at a single location. The chain dependent (or "Richardson") model, originally proposed in [13], is a popular multivariate SWGEN for daily weather variables, including precipitation, wind, temperature, and solar radiation. In this approach, the time evolution of the multivariate process (U, T, SR) is modeled conditionally to the rainfall occurrence as a mixture of two Vector AutoRegressive (VAR) models with different VAR models for dry and wet days, and the rainfall occurrence is modeled as a first order Markov chain.

Non-parametric SWGENs, initially proposed in [26], simulate daily synthetic time series using data-driven methods in order to capture deviations from standard distributions and complex dependencies between the variables. Most of the non-parametric SWGENs are based on K- Nearest Neighbors (K-NN) bootstrap. Given the weather conditions at a time $t-1$, one of the K most similar conditions (analogs) is sampled in the historical database and the successor of the selected day provides the simulated weather condition for time t.

In contrast to past-proposed SWGEN formulation approaches, an original SWGEN is proposed in this paper, where the strengths of parametric and non-parametric approaches are combined, in order to simulate realistic hourly multivariate time series (U, T, SR) of the target installation site during the lifetime period of the RES-based desalination system. Developing a full parametric model for the multivariate process (U, T, SR) at the hourly scale would require a considerable modeling effort, in particular for the sub-daily temporal variations which exhibit complex features. This may explain why, to the best of our knowledge such generator has never been proposed before in the literature. A limitation inherent to any resampling scheme is its difficulty in generating "new" situations since only observed situations can be resampled (see [26]). For example, the operation of a RES system is sensitive to long periods with low energy production and it is not clear how to modify the resampling approaches to generate long periods with low wind and solar radiation in a realistic way. In the proposed SWGEN, a parametric Markov Switching AutoRegressive (MSAR) model, which shares the same probabilistic structure with the chain-dependent model except that the Markov chain describing the switching between the different weather regimes is introduced as a hidden (or latent) variable, is first used to simulate the variables at the daily scale. A non-parametric approach is then used to simulate the hourly weather conditions conditionally to the daily time series.

The aim of this paper is to show that SWGENs can provide useful information for the optimal design and reliability evaluation of RES systems. The rest of this paper is organized as follows: an original SWGEN for hourly multivariate time series (U, T, SR) is presented in Section 2. The RES system considered in this study, which is an autonomous desalination plant using both PV arrays and W/Gs is described in Section 3, together with the design optimization procedure. Numerical results are given in Section 4 where the realism of the meteorological sequences simulated with the SWGEN over a prolonged operational lifetime period of the RES-based desalination system is initially assessed. Then, it is shown that the SWGEN enables to assess the impact of the climate variability on the reliability of the designed RES-based desalination plant during its lifetime period. Finally, it is demonstrated how SWGENs can be used to design RES-based desalination systems with higher operational reliability during their lifetime period, compared to their counterparts designed using other sources of meteorological data. Finally, the conclusions are discussed in Section 5.

2. Stochastic weather generators for RES applications

2.1. Weather type model for daily data

The most usual approach for simulating (U, T, SR) at the daily time step is the chain dependent model [13, 27]. This is a Markov model which includes both a discrete and a continuous variable. The discrete variable, denoted S_t hereafter, is the rainfall occurrence. It has values in $\{1, 2\}$ with $S_t = 1$ if the day t is dry and $S_t = 2$ if it is wet. S_t is assumed to be a first order Markov chain (i.e. the rainfall occurrence for the day t only depends on the rainfall occurrence of the previous day) and different autoregressive models are used in the dry and wet days. More precisely, if $Y_t \in \mathbb{R}^d$ denotes the multivariate weather variables for the day t, it is assumed that

$$Y_t = A^{(S_t)} Y_{t-1} + B^{(S_t)} + \left(\Sigma^{(S_t)}\right)^{1/2} \epsilon_t \quad (1)$$

with $(\epsilon_t)_t$ a Gaussian white noise sequence with values in $\mathbb{R}^d$. $A^{(s)} \in \mathbb{R}^{d,d}$ and $B^{(s)} \in \mathbb{R}^d$ describe the evolution of the weather variables conditionally to the rainfall occurrence $s \in \{1,2\}$ and $\Sigma^{(s)} \in \mathbb{R}^{d,d}$ is a covariance matrix which describes the corresponding temporal variability. It may seem arbitrary to define the regimes S_t based only on rainfall occurrence, especially when the rainfall in not of interest as it is the case in this work. Weather type models (see [28]) have been proposed in this context, where the variable S_t is intended to represent "weather types" or "meteorological regimes". The weather regimes can be defined a priori using some knowledge of local weather conditions, or a clustering algorithm. An alternative approach, which is used in this work, consists in introducing the weather regime as a latent variable and using the EM algorithm [29] to maximize the likelihood function and identify the optimal regimes in a statistical sense. The proposed model is a MSAR model where the latent weather type S_t is assumed to be a first order Markov chain and the conditional evolution of the weather variables is modeled using the AR(1) model (1). These are classical assumptions in the literature on weather type models (see [28] and references therein).

The considered meteorological time series are non-stationary with important and complex daily, seasonal and eventually interannual components. A pre-processing step can be applied to the data in order to remove these components, or they can be included directly in the models using coefficients which evolve periodically (see e.g. [13]). In this work, the interannual components (related e.g. to climate change) are neglected and daily components are reproduced using the non-parametric approach introduced in the next section. Seasonal components are taken into account by fitting different models separately for the 12 calendar months. This approach is appropriate when the length of the time series is long enough and when it is reasonable to assume that the meteorological conditions are approximately stationary on a given month. A post-processing is applied to the data simulated by the fitted MSAR model to force the simulated time series to have the same monthly marginal distribution as the data using the probability integral transform on each variable.

2.2. *Non-parametric approach for simulating daily blocs of* (U, T, SR)

The model introduced in the Section 2.1 is used to generate artificial sequences of daily mean for U, T and SR, whereas for the simulation of the RES system under study, time series at the hourly time step are needed. For that purpose, a non-parametric SWGEN is proposed, with a conditioning by the daily time series in the spirit of [30]. In practice, for each day t, the K days with more similar mean meteorological conditions are searched in the historical data and the daily blocks are simulated by choosing randomly the daily block among these neighbors. The distance used for the K-NN search is the Euclidean distance between a multivariate query and the multivariate catalog. It is computed on both the daily mean meteorological value of (U, T, SR) and the day of the year. The value $K = 5$ is used in practice and neighbors are searched in a window of 11 days centered on the current date and then weighted with an exponential decay in order to increase the probability to sample the nearest ones. The K and window width values were chosen empirically to produce the most realistic sequences according to the validation criteria discussed in Section 4.

The search of nearest neighbors in large databases may lead to large computational time. To avoid this issue, k-d tree nearest neighbor search is used [31].

A post-processing step, based on an additive correction, is performed to ensure that the daily means of the time series at the hourly time step are the same as the ones generated at the daily time step.

3. Design of the RES system

3.1. *RES system modeling for optimization*

The desalination plant under study is power-supplied by an autonomous hybrid RES system (i.e. not connected with the electric grid). A block diagram of the overall RES-based desalination plant is illustrated in Figure 1. It consists of multiple PV arrays and W/Gs, each connected through a battery charger to a common DC bus that is also interconnected with a battery bank. Multiple DC/AC converters (i.e. inverters) are used to produce the AC voltage required by the desalination units to operate. The water produced by the desalination units covers the potable water requirements of the consumers. Any water surplus is stored in a water tank in order to be used during the time periods of high water demand by the consumers.

The RES-based desalination system is designed according to the optimization process described in [32], such that the present value of its total lifetime cost (including the installation and maintenance costs), C_{total} (€), is minimized and simultaneously the water demand of the consumers is completely covered. The value of C_{total} is calculated as follows:

$$\begin{aligned} C_{total}(X) = {} & N_{PV}C_{PV} + N_{BAT}C_{BAT} + N_{CH}C_{CH} + W_{TANK}C_{TANK} \\ & + N_{WG}(C_{WG} + hC_h) + N_{DU}C_{DU} + N_{INV}C_{INV} + \sum_{j=1}^{Y}(N_{PV}M_{pV} \\ & + N_{BAT}M_{BAT} + N_{CH}M_{CH} + W_{TANK}M_{TANK} \\ & + N_{WG}(M_{WG} + hM_h) + N_{DU}M_{DU} + N_{INV}M_{INV})\frac{(1+g)^j}{(1+i)^j} \\ & + R_{BAT} + R_{CH} + R_{INV} \end{aligned} \tag{2}$$

where $X = (N_S, N_B, N_{BAT}, \beta, W_{TANK}, N_{DU}, N_{WG}, h)$ is the vector of design variables, N_S is the number of PV modules connected in series in each PV array, N_B is the number of PV arrays, which is also equal to the total number of battery chargers N_{CH}, N_{BAT} is the total number of batteries comprising the battery bank, β (in degrees) is the tilt angle of the PV modules, $N_{PV} = N_SN_PN_B$ is the total number of PV modules of the RES system with N_P being the number of parallel-connected PV strings in each PV array, C_{pV}, C_{BAT}, C_{CH}, C_{WG}, C_{DU} and C_{INV} are the purchase and installation costs (€) of each PV module, battery, PV battery charger, W/G, desalination unit and DC/AC inverter, respectively, W_{TANK} (lt) is the total volume of the water tank, C_{TANK} (€lt^{-1}) is the construction cost of the water tank per lt of water tank volume, N_{WG} is the total number of W/Gs, h is the height of the W/Gs installation tower, C_h (€m^{-1}) is the purchase and installation cost of the W/Gs installation tower per meter of height, N_{DU} is the total number of desalination units, Y (years) is the lifetime period of the RES-based desalination system, M_{pV}, M_{BAT}, M_{CH}, M_{WG}, M_{DU} and M_{INV} (€) are the present values of the annual maintenance cost of each PV module, battery, PV battery charger, W/G, desalination unit and DC/AC inverter, respectively, M_{TANK} (€lt^{-1}) is the present value of the annual maintenance cost of the water tank per lt of water tank volume, M_h (€m^{-1}) is the present value of the annual maintenance cost of the W/Gs installation tower per meter of height, R_{BAT}, R_{CH} and R_{INV} (€) are the present values of the total costs of replacing the batteries, PV battery chargers and DC/AC inverters, respectively, during the desalination system lifetime period, g (%) is the annual inflation rate and i (%) is the interest rate. The values of R_{BAT}, R_{CH} and R_{INV} are calculated as described in [33]. The value of N_P is calculated based on the number of PV modules connected in series, N_S, such that the power rating of the PV array does not exceed the nominal power capability of the corresponding PV battery charger.

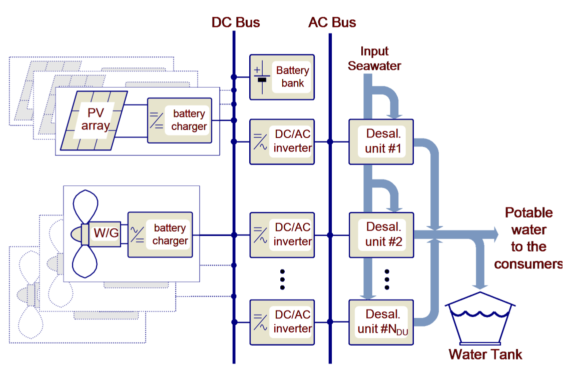

Figure 1: A general diagram of the RES-based desalination system under study.

The target of the design optimization process is to derive the optimal value of vector X which minimizes the present value of the lifetime system cost C_{total}, i.e. compute

$$X^* = \arg\max_{X \in \Omega} \left(C_{total}(X) \right). \tag{3}$$

The optimal value of C_{total} and the corresponding optimal vector of design parameters X^* are derived by using the Particle Swarm Optimization (PSO) algorithm [1]. In comparison with other evolutionary optimization methods (e.g. genetic algorithms, differential evolution etc.), the PSO algorithm has the advantage of relative computational complexity, while, simultaneously, being capable to effectively solve complex non-linear optimization problems [34]. The subset Ω defines a set of constraints:

1. the state of charge of the battery bank never drops below the minimum permissible limit specified by the batteries manufacturer,

2. the level of water stored in the water tank never drops below 10% and

3. the water demand of the consumer is always covered by the water production of the desalination units and/or the water storage tank.

A flowchart of the design optimization process is depicted in Figure 2. The designer provides as inputs the operational characteristics (e.g. nominal power, efficiency, etc.) and cost of the devices (i.e. PV modules, W/Gs, desalination units, etc.) synthesizing the hybrid PV-W/G desalination system. Additionally, the time series of the hourly water demand of the consumer during the year and the hourly mean values of solar irradiance, ambient temperature and wind speed which prevail at the installation site for a predefined time period (e.g. 1 year, 10 years, etc.) are provided by the designer. The aforementioned time series are repeatedly used for simulating the operation of the RES-based desalination plant during each hour of the entire duration of its service lifetime period, which is specified by the designer (e.g. 20 years).

3.2. Simulation of the RES system operation

During the simulation process, both the energy flow from the energy production units to the battery energy storage subsystem and the desalination units, as well as the flow of water from the desalination units to the water storage tank and the consumer, are considered. Therefore, the total power produced by all PV and W/G power production units of the desalination system during hour $t \in \{1, ..., 8760Y\}$ is initially calculated as follows [32]

$$P_{RES}(t) = N_S N_P N_B n_{ch} n_{MPPT} P_{PV}(\beta, t) + N_{WG} P_{WG}(h, t) \tag{4}$$

where n_{ch} (%) and n_{MPPT} (%) is the power conversion efficiency and the MPPT efficiency, respectively, of each PV battery charger, $P_{PV}(\beta, t)$ (W) is the power produced during hour t by a PV module installed at a tilt angle of β degrees and $P_{WG}(h, t)$ (W) is the power produced during hour t by a W/G installed at height h. The values of $P_{PV}(\beta, t)$ and $P_{WG}(h, t)$ are calculated as analyzed in [32].

The total DC input power of the DC/AC inverters, which must be provided by the RES generators and/or the battery bank in order to operate the desalination units, is given by:

$$P_L(t) = N_{DU} \frac{P_U}{n_{INV}} \tag{5}$$

where P_U (W) is the power consumption of each desalination unit and n_{INV} (%) is the power conversion efficiency of the DC/AC inverters.

If $P_{RES}(t) > P_L(t)$, then the power surplus $P_B(t) = P_{RES}(t) - P_L(t)$ is used to charge the battery bank. In case that $P_{RES}(t) < P_L(t)$, then the power deficit $-P_B(t) = P_L(t) - P_{RES}(t)$ is covered by discharging the battery bank. At each time step of the simulation process, the state of charge of the battery bank, $SOC(t)$ (%), is calculated as follows:

$$SOC(t) = min\left(SOC(t-1) + \frac{n_B P_B(t) 1h}{N_{BAT} C_B}, 100\%\right) \tag{6}$$

where C_B (Wh) is the nominal capacity of each battery, n_B is the battery round-trip efficiency (i.e. $n_B = 0.8$ during charging and $n_B = 1$ during discharging) and $P_B(t) < 0$ during discharging of the battery bank.

The minimum permissible limit of the battery bank state of charge is defined as follows:

$$SOC_{min} = 1 - DOD_{MAX} \tag{7}$$

where DOD_{MAX} (%) is the maximum permissible depth of discharge specified by the batteries manufacturer. In case that at any time step t the state of charge of the battery bank drops below SOC_{min} then the operation of the RES-based desalination system is considered unsuccessful and the corresponding vector of decision variables X is rejected as a potential optimal solution of the optimization process.

The total volume of water produced by the desalination units during operation at hour t, W_{RO} (lt), is calculated as follows:

$$W_{RO}(t) = N_{DU} W_U \tag{8}$$

where W_U (lt) is the hourly water production of each desalination unit, specified by its manufacturer.

If the volume of water produced by the desalination units during time step t is higher than the water demand of the consumer $W_D(t)$, then the water surplus $\Delta W(t) = W_{RO}(t) - W_D(t)$ is stored in the water tank, else the water deficit $-\Delta W(t) = W_D(t) - W_{RO}(t)$ is withdrawn from the water tank. Therefore, the volume of water stored in the water tank, W_T (lt), is calculated as follows

$$W_T(t) = min\left(W_T(t-1) + \Delta W(t), W_{TANK}\right) \tag{9}$$

where W_{TANK} (lt) is the total volume of the water tank.

In case that at any time step the volume of water in the tank drops below the minimum permissible limit of $0.1W_{TANK}$, then the operation of the desalination system is also considered to be unsuccessful and the corresponding vector X is not considered as a potential optimal solution.

This simulation process is repeated for multiple values of the decision variables vector $X = (N_S, N_B, N_{BAT}, \beta, W_{TANK}, N_{DU}, N_{WG}, h)$ produced by the PSO algorithm. Among the alternative values of X examined, the set of decision variables which results in the minimum value of C_{total} and simultaneously satisfies the constraints (1-3) described above, is considered as the optimal configuration of the RES-based desalination plant under design.

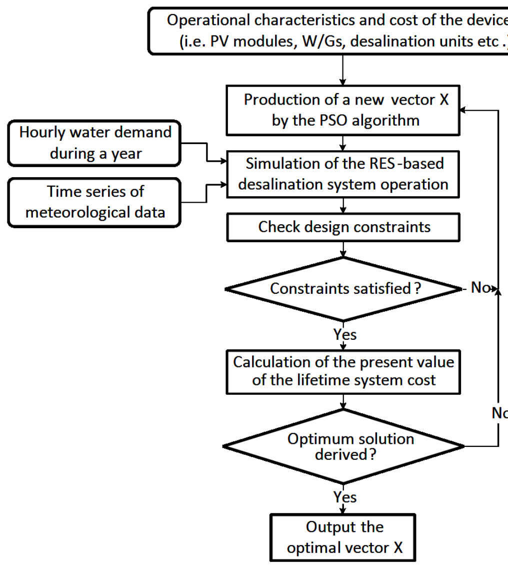

Figure 2: A flowchart of the optimal design process of the hybrid PV-W/G desalination system under study.

4. Numerical results

The SWGEN has been implemented using the free software environment R. The MSAR model of Section 2.1 was implemented using the R package NHMSAR [35]. It was run on a computer with a 2.5-GHz Central Processing Unit (CPU) and 8-GB of Random Access Memory (RAM). The modeling and optimization process of the RES-based desalination system have been implemented in the MATLAB software platform. The minimization of C_{total} in (3) has been performed using the PSO built-in function of the MATLAB Global Optimization Toolbox. Also, it was executed in a computer with a 1.7-GHz CPU and 4-GB of RAM.

4.1. Data

In this section, the ERA5 meteorological dataset is used. ERA5 is a re-analysis product provided by the European Center for Medium Range Weather Forecasts (ECMWF) which can be freely downloaded from the link `https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5`.

The meteorological data are available on a 30 kilometers grid over the Earth and 40 years of data are available (from 1979 to 2018) with a time resolution of 1 hour. The RES-based power supply system considered in this study is located at the point with coordinates $(35.25^o N, 26.25^o E)$ in the North-East of the Crete island (Greece).

The last 20 years of ERA5 meteorological data (period $1999 - 2018$) were initially used as input in the design optimization algorithm of Section 3 and the corresponding optimal configuration for the RES system is presented in Table 1. The resulting optimal capacities of the RES devices, battery bank and desalination units for the configurations in Table 1 are displayed in Table 2. The design optimization process presented in Section 3 was also applied by providing alternative types of meteorological data as inputs and the corresponding design optimization results in terms of the optimal values of the design parameters [i.e. vector X* in (3)] are also presented in Table 1 and further discussed below. Figure 3 illustrates the behavior of the RES-based desalination system. The weather conditions are shown on the left panels and the right panels show the power produced by the PV arrays and W/Gs together with the volume of water in the tank (middle right panel). The colored area superimposed on Figure 3 corresponds to the three weather regimes identified by the MSAR model (see Section 2.1) fitted on the last 20 years of daily meteorological data. The first regime (white periods) corresponds to periods with low wind speed where the water volume in the tank generally decreases due to a lack of wind-generated energy at the desalination plant installation site under consideration. The two other regimes correspond to periods with higher wind speed where the water volume in the tank tends to increase. It suggests that the volume of water in the tank at a given time is related to the succession of the weather types in the previous weeks and that a low volume of available water will be reached when the time spent in the first weather regime is important.

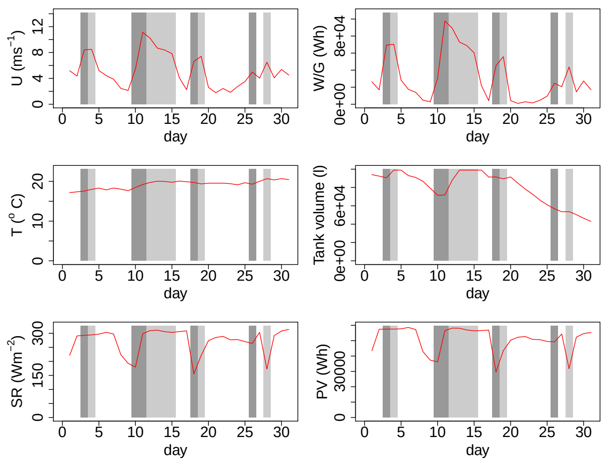

Figure 3: Daily mean meteorological time series (left) and corresponding energy production and water level (right) for the month of May 2007. The colors correspond to the regimes identified by the MSAR model.

4.2. Validation of the proposed SWGEN

The SWGEN described in Section 2 is fitted to the last 20 years (period $1999-2018$) of ERA5 meteorological data and then used to simulate a large number (20 years of operation are simulated 1000 times) of artificial meteorological data. Initially, it was verified that the simulated sequences are "realistic" in the sense that they have statistical properties similar to the ones of the original data. The Quantile-quantile (Q-Q) plots on Figure 4 permit to compare the marginal distribution of the different variables and the off-diagonal plots permit to compare visually the bivariate distributions and the relations between the different variables. The results, which correspond to weather conditions at noon in May, show a good agreement. Similar results where obtained at other dates too.

It was also verified that the model is able to reproduce seasonal components of the meteorological parameters. This is illustrated on Figure 5 where the

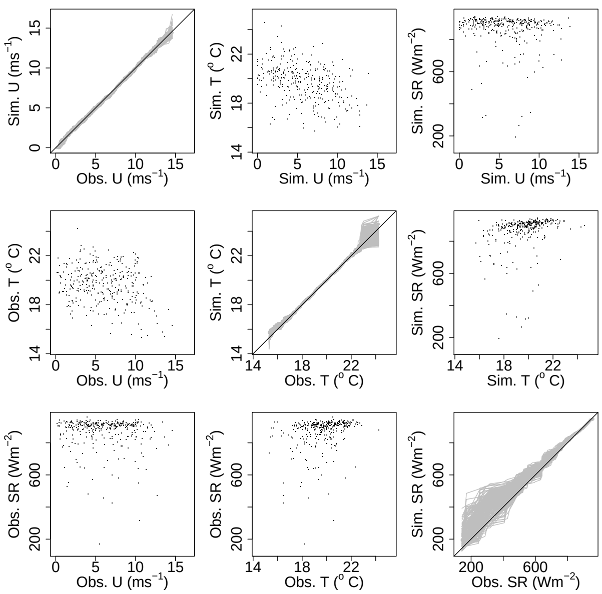

Figure 4: Comparison of the distributions of the observed and simulated weather time series at noon in May. The diagonal panels show Q-Q plots for each variable, where a Q-Q plot is plotted in grey for each of the 1000 20-years simulated time series. The off-diagonal panels show scatter plots for the different pairs of variables. The bottom plots correspond to observations and top plots to 20 years of simulations.

monthly distributions of daily power produced by the PV arrays and W/Gs are displayed for both the original and simulated time series. The agreement is generally good: periods with low energy production are well reproduced, and the variability around the mean seasonal component is also well captured.

In order to validate the dynamical properties of the simulated sequences, the autocorrelation functions (ACF) of the meteorological time series are plotted in Figure 6. These results demonstrate that the SWGEN simulates the wind speed time series with too low autocorrelation for lags below one day. This is due to the breaks created when sampling non-parametrically the daily blocks. Indeed, given the block of the day d, the next block is chosen among the successors of the neighbors of the block d and there may be a jump (also called break) between the wind at the last hours of the day d and the wind at the first hours

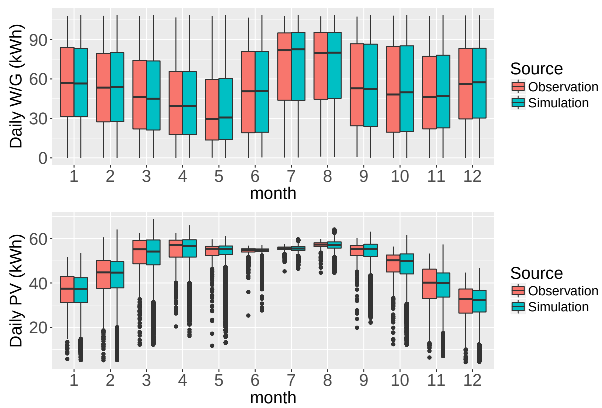

Figure 5: Boxplots of the monthly distribution of daily wind and PV energy produced by the system.

of the $d+1$ chosen block. For instance, the wind can be high at the end of the day d and low at the beginning of the day $d+1$. It leads to low correlations between the last point of a block and the first point of the next block. This effect is not visible on SR (the breaks do not exist because the time series is equal to zero during the night) and less pronounced on T (the variability of temperature in the night is low and thus the breaks are less important). Remark that the long-term evolution of the water level in the water storage tank is not expected to be sensitive to these breaks. This was verified by applying the non-parametric generator conditionally to the time series of daily mean computed on the data. It leads to meteorological time series which have the same daily component with the data but breaks between the daily blocks. The level of water in the water storage tank associated to these simulated time series shows similar characteristics with the ones computed by using the original time series.

A well-known weakness of many SWGENs proposed in the literature in the past is that they tend to underestimate the interannual variability in the weather conditions ("overdispersion" phenomenon, see e.g. [36]). Figure 7 indicates however that the SWGEN is also capable to reproduce the interannual variability in the monthly means. This is an important quantity when designing a RES system since it is related to the probability of having long time periods with low

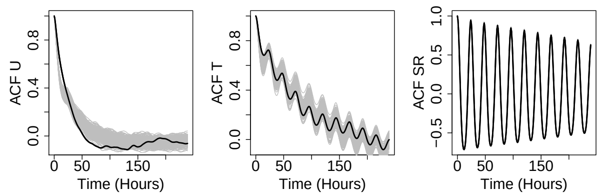

Figure 6: Autocorrelation functions of the different variables in May. The black curve corresponds to the ACF of the data and each grey curve represent the ACF computed on each of the 1000 simulated 20-years time series.

energy production.

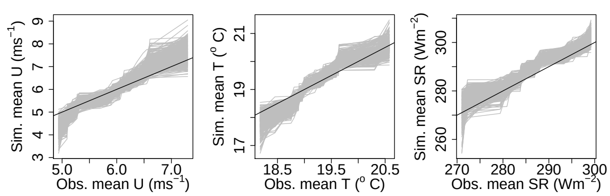

Figure 7: Q-Q plots of the monthly mean of U, T and SR. A Q-Q plot is plotted in grey for each of the 1000 20-years simulated time series.

4.3. Reliability of the RES-based desalination system

The sequences simulated with the SWGEN can be used to assess the reliability of the RES system before its actual installation. Since the goal of a desalination plant is to cover the water demand of the consumer, different metrics have been employed in order to evaluate its operational reliability from different perspectives, as follows:

1. the total Water Deficit during its operational lifetime period (i.e. 20 years), WD (m^3): the value of this parameter is practically useful when the consumer is mostly interested in the overall water supply and, therefore, small quantities of water deficit during some time intervals do not affect significantly the reliability of the desalination system,

2. the total Time with water Deficit when the system cannot provide the water required by the consumer, TD (hours): this parameter quantifies

the desalination system reliability irrespectively of how much is the water deficit, thus considering that even a very small quantity of water not being provided to the consumer is equally critical for the consumer and

3. the total Time that the level in the tank water is Lower than the 10% level, TL (hours): the value of this parameter demonstrates how often the safety limit of 10% of the water storage tank, which was considered during the design optimization process, is violated during the actual operation of the desalination plant.

The results are given in Table 3 for the different optimal configurations described in Table 1 and Table 2. These results have been derived by simulating the operation of the overall RES-based desalination system (i.e. including all of the energy- and water-related components that it comprises) as described in Section 3. They demonstrate that the design results, the lifetime cost and the reliability of the RES-based desalination system depend on the meteorological data set which has been used. The optimal configurations presented in these tables are different since the time-series of meteorological data which has been used in each case affects differently the evolution of the hourly operation of the energy- and water-related subsystems of the RES-based desalination plant during its lifetime period. In turn, this affects differently the profile in time of the energy stored in the battery bank and the water stored in the water tank, as well as the sizes/capacities of the devices which should be included in order to satisfy the design constraints specified in Section 3. Let us first discuss the second line of these tables, which corresponds to the optimal RES system obtained when using the ERA5 data for the period 1999-2018 as input to the optimization procedure (denoted *ERA 1999-2018 configuration* hereafter). Remark that the values of WD, TD and TL are equal to zero when testing the RES system on the dataset which has been used in the design optimization procedure. This is expected from the optimization algorithm (see Section 3) since the level of water stored in the water tank is constrained to never drop below 10% and the water demand of the consumer is constrained to be always covered. When testing the ERA 1999-2018 configuration using the ERA 1979-1998 meteorological data, it is found that the RES-based desalination system cannot always cover the water demand, although the water deficit seems relatively small (less than 1 day over 20 years when the water demand cannot be covered by the RES system). By simulating a large number of meteorological data, SWGENs allow to estimate the full distribution of the metrics WD, TD and TL. According to the second line of Table 3 the probability of having at least one water deficit event (failure) over 20 years of the desalination plant operation with the ERA 1999-2018 configuration is close to 90%. The distributions of WD, TD and TL are summarized using 95% fluctuation intervals in Table 3. It is shown that with a probability of 95%, the total water deficit WD fluctuates in the interval between $0m^3$ and $157m^3$, with a median value of $37m^3$, and the total time that the system cannot provide the water required by the consumer has values between 0 and 8.1 days, with a median value of 2 days.

Next, the results obtained with other configurations were also investigated. Since 40 years of meteorological data are available and the lifetime of the RES system is fixed to 20 years, it was possible to execute the design optimization procedure of Section 3 on two different blocks of meteorological data. The optimal configuration obtained when using the ERA5 data for the period 1979-1998 as input to the design procedure of Section 3 is abbreviated as *ERA 1979-1998 configuration* hereafter and is described in Table 1. According to Table 3, the ERA 1979-1998 configuration is slightly more robust than the ERA 1999-2018 configuration, with typically lower values for WD, TD and TL.

Another possibility is to use a TMY for the optimization procedure, which is frequently performed to design RES systems. A method similar to the Sandia approach [8] was used in this study to define the TMY, where the Kolmogorov-Smirnov distance is used instead of the Finkelstein-Schafer metric, as in [10], and the same weight is given to the three meteorological variables. The obtained typical year is then repeated 20 times to produce the 20 years of meteorological data during the lifetime period of the RES-based desalination plant for the design optimization procedure. The obtained optimal configuration is described in Table 1 and is abbreviated as *TMY configuration* hereafter. According to Table 3, this is the configuration with the lowest reliability. Deficits of water occur for all the observed and simulated 20-years of meteorological data and the reliability metrics have larger values than for the other configurations. This is expected since through the TMY approach the interannual variability of the meteorological conditions is not taken into account during the RES system design process.

The discussion above illustrates that SWGENs can be used to assess the reliability of a RES system. Existing weather data sets describe past weather conditions and using them for the design of a RES system may lead to non-robust design if they are not representative enough of future weather conditions that it will encounter during its lifetime. This is illustrated for example in Table 3: the optimal design obtained with the ERA 1999-2018 data set lead to a period with water deficit for the ERA 1979-1998 data set. SWGENs allow to simulate quickly artificial but realistic weather conditions and assess the impact of the climate variability on the RES system. Moreover, SWGENs may also be useful during its design process by providing multiple input meteorological time series to the design optimization algorithm. However, such an approach would increase the computational complexity of the design optimization procedure, which is already highly demanding, since it requires to simulate the RES system operation at the hourly time scale for its entire lifetime period (i.e. 20 years in this study). Therefore, an alternative approach of incorporating the SWGEN in the design optimization process was also investigated in this paper, where the 20-years sequences simulated with the SWGEN were used individually as input to the design optimization procedure in order to assess the variability of the design results with respect to the meteorological data and identify more robust configurations of the RES-based desalination plant in terms of its reliability. This is illustrated in this study by using two particular simulated 20-years sequences :

- the time series which gives the median total water deficit (criterion WD) for the ERA 1999-2018 configuration. The corresponding optimal configuration is denoted as *SWGEN (50%) configuration* hereafter, and

- the time series which corresponds to the 95% total water deficit for the ERA 1999-2018 configuration. The corresponding optimal configuration is denoted as *SWGEN (95%) configuration* hereafter.

The flowchart of Figure 8 summarizes the proposed methodology, where the sequences simulated with SWGEN are used both for reliability evaluation and optimal design of the RES-based desalination system. The values of the reliability metrics are calculated at the final stage of this process (i.e. "RES simulator" block in Figure 8) by simulating the operation of the RES-based desalination plant as analyzed in Section III.B.

The resulting configurations are presented in Table 1 and Table 2. According to the values of C_{total}, the lifetime cost of the RES-based desalination plant configuration which has been derived by using the SWGEN (95%) data is higher by $2.05 - 17.03\%$ compared to the systems designed by using other types of meteorological data sources. However, the corresponding results presented in Table 3 reveal that the SWGEN (95%) configuration is the most robust one (i.e. with lowest values of WD, TD and TL) compared to the configurations derived when the other types of meteorological datasets are input to the design optimization process, and this is true when testing the various configurations both on ERA data and synthetic data simulated with the SWGEN.

Figure 9 provides a visual comparison of the various configurations. It is a useful tool to choose the configuration which has the best compromise in terms of cost, capacities and reliability. It shows for example that the TMY configuration is slightly less expensive than the other configurations, although it is based on a larger total W/G power, but it is less reliable. On the opposite, the SWGEN (95%) has higher PV power and battery capacities, and this permits to obtain a more reliable configuration but at a higher cost.

Table 1: Optimal configurations corresponding to different calibration meteorological time series.

	Optimal configuration X^*								
Calibration data	N_S	N_B	N_{BAT}	β	W_{TANK}	N_{DU}	N_{WG}	h	C_{total} (€)
ERA 1979-1998	1	16	19	50	97033	1	6	12	78786
ERA 1998-2018	4	15	17	48	98956	1	7	12	78912
TMY	3	5	18	35	95111	1	10	15	69471
SWGEN (50%)	1	18	40	58	98761	1	5	12	79668
SWGEN (95%)	1	18	85	50	98831	1	5	15	81300

Table 2: Optimal capacities of devices for each of the optimized configurations in Table 1.

	Optimal capacities			
Calibration data	Total PV power (KW)	Total W/G power (kw)	Total nominal battery capacity (kWh)	Total nominal water production (lt/h)
ERA 1979-1998	10.4	3.6	42.18	788.54
ERA 1998-2018	9	4.2	37.74	788.54
TMY	3	6	39.96	788.54
SWGEN (50%)	11.7	3	88.8	788.54
SWGEN (95%)	11.7	3	188.7	788.54

5. Conclusion

For the reliable operation of RES systems, it is necessary to design them by using a set of meteorological data resembling the actual meteorological conditions that prevail at the target installation field during their prolonged lifetime period. In this paper, the use of SWGENs has been introduced for the first time in the existing literature, for the optimal design and reliability evaluation of RES-based desalination plants. An original SWGEN has been proposed, combining parametric MSAR models and nonparametric resampling, which is capable to simulate realistic hourly multivariate time series of solar irradiance, temperature and wind speed. The numerical results demonstrated that depending on the type of meteorological parameters dataset input to the design process, different optimal configurations of the RES-based desalination plant are derived and each of them exhibits a different reliability during the lifetime period of the desalination plant. The incorporation of the appropriate time series, among the multiple alternatives produced by the SWGEN, to the design optimization

Table 3: Reliability metrics for optimal configurations corresponding to different calibration meteorological time series. For the SWGEN data, the percentage of failure (simulations where the water demand can not be always covered) is also reported together with the median values and WD, TD and TL over the 1000 simulations. The italic values in parenthesis correspond to the quantiles of order 2.5% and 97.5%.

	Test data					
	ERA 1979-1998			ERA 1999-2018		
Calibration data	WD (m^3)	TD (h)	TL (h)	WD (m^3)	TD (h)	TL (h)
ERA 1979-1998	0	0	0	0	0	0
ERA 1999-2018	16.8	20	10.9	0	0	0
TMY	1.19E3	1.21E3	6.02E3	968	989	4.94E3
SWGEN (50%)	8.06	13	343	0	0	0.5
SWGEN (95%)	0	0	0	0	0	0

	Test data			
	SWGEN			
Calibration data	% fail	WD (m^3)	TD (h)	TL (h)
ERA 1979-1998	73.0%	14.0 *(0;104)*	20 *(0;140)*	271 *(0;1.18E3)*
ERA 1998-2018	90.6%	37.9 *(0;157)*	48 *(0;195)*	465 *(15;1.46E3)*
TMY	100%	1.22E3 *(821;1.70E3)*	1.21E3 *(825;1.72E3)*	6.08E3 *(446; 8.08E3)*
SWGEN (50%)	75.1%	15.9 *(0;106)*	24 *(0;154)*	350 *(0;1.48E3)*
SWGEN (95%)	30.1%	0 *(0;50.5)*	0 *(0;73)*	1 *(0;662)*

process of the RES-based desalination plant in order to improve the reliability of the designed system, was also investigated. It has been demonstrated that by using as input to the design optimization process the SWGEN time series which corresponds to the 95% total water deficit for the ERA 1999-2018 configuration, results in a RES-based desalination plant configuration with higher reliability compared to the configurations derived when the other types of meteorological datasets are employed. The proposed methodology has been applied for the design and reliability evaluation of RES-based desalination systems due to the importance of the environmentally-friendly energy-water nexus for solving the potable-water scarcity problems in major areas of the World. However, it seems general enough to be applicable in other RES-based energy production applications. Current work includes the computationally-efficient incorporation of the SWGEN-based reliability estimation in the RES system design optimization process. It also should be noted that the SWGEN is built to simulate weather

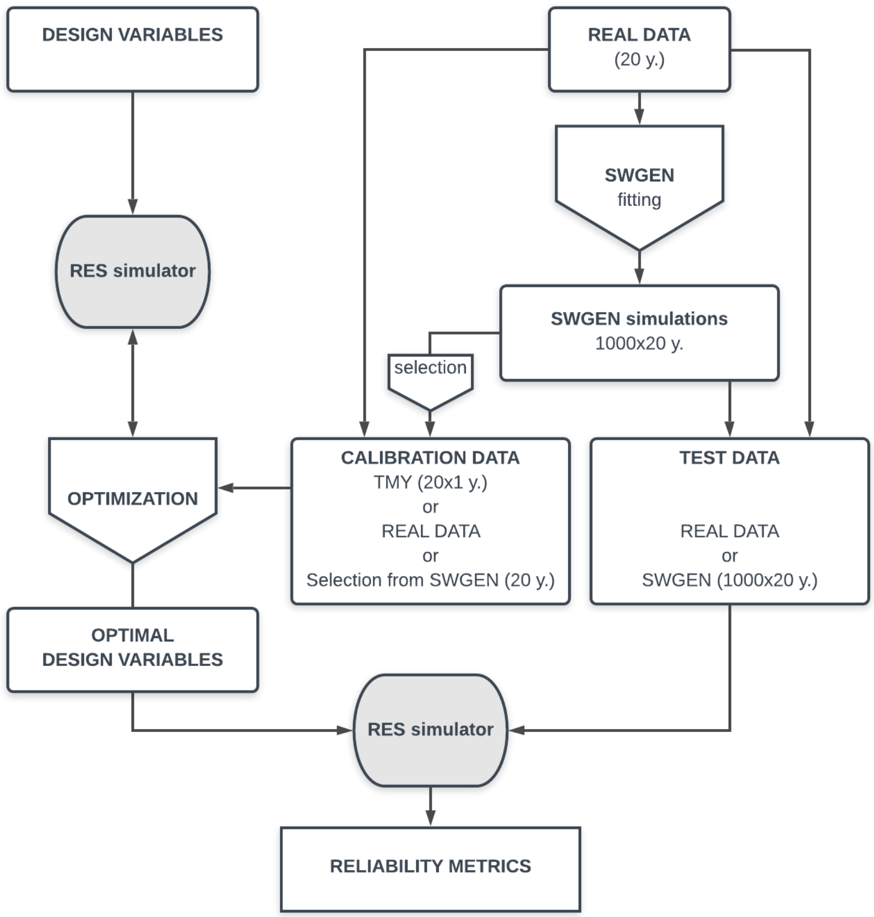

Figure 8: Flowchart of the procedure used for optimal design and reliability evaluation of the RES-based desalination system using SWGEN.

conditions representative of the last decades, hence to study the impact of climate change on the design and reliability of the RES systems, information on climate evolution should be integrated in the SWGEN, see e.g. [37].

Acknowledgements

The work presented in this paper has been implemented within the framework of the research project DES^{2}iRES: DESign of DESalination systems based

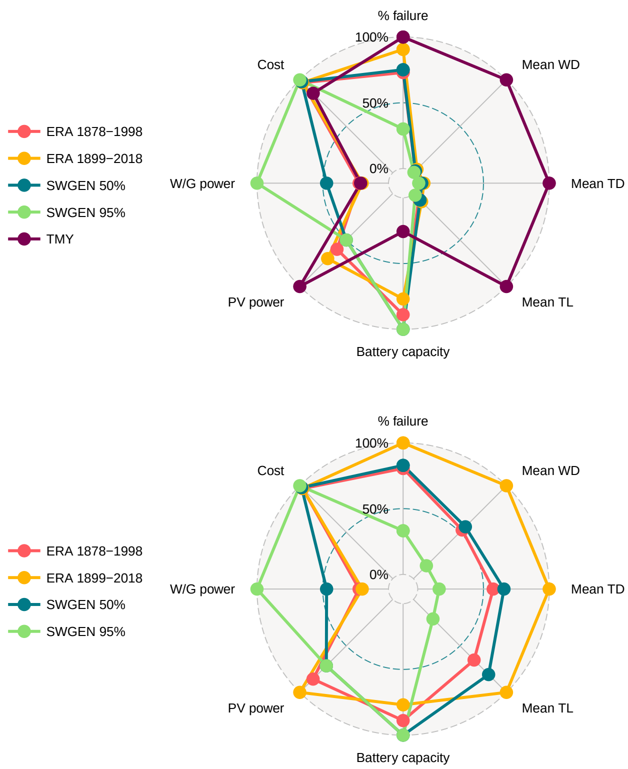

Figure 9: Comparison of the different optimal configurations in terms of cost (see Table 1), capacities (see Table 2) and reliability (see Table 3). In the bottom panel, the TMY configuration has been removed to allow easier comparison between the other configurations.

on optimal usage of multiple Renewable Energy Sources (ERANETMED NEXUS-14-049) supported by the ERA-NET Initiative EURO-MEDITERRANEAN Co-operation through ERANET joint activities and beyond (ERANETMED) under the topic ERANETMED Energy-Water nexus. In Greece, this work was funded by the Operational Program Competitiveness, Entrepreneur-ship and Innovation 2014-2020 (co-funded by the European Regional Development Fund) and managed by the Greek General Secretariat of Research and Technology, Ministry of Education, Research, and Religious Affairs of Greece under the project DES^{2}iRES (T3EPA-00017) of the ERANET action. In France it was funded by the National Research Agency (ANR). These supports are gratefully acknowledged.

References

[1] E. Eriksson, E. M. Gray, Optimization of renewable hybrid energy systems–a multi-objective approach, Renewable Energy 133 (2019) 971–999.

[2] T. Dragičević, H. Pandžić, D. Škrlec, I. Kuzle, J. M. Guerrero, D. S. Kirschen, Capacity optimization of renewable energy sources and battery storage in an autonomous telecommunication facility, IEEE Transactions on Sustainable Energy 5 (2014) 1367–1378.

[3] İ. Şengör, H. C. Kılıçkıran, H. Akdemir, B. Kekezolu, O. Erdinc, J. P. Catalao, Energy management of a smart railway station considering regenerative braking and stochastic behaviour of ess and pv generation, IEEE Transactions on Sustainable Energy 9 (2018) 1041–1050.

[4] A. Hamidi, S. Golshannavaz, D. Nazarpour, D-facts cooperation in renewable integrated microgrids: a linear multi-objective approach, IEEE Transactions on Sustainable Energy 10 (2019) 355–363.

[5] M. Molinos-Senante, D. González, Evaluation of the economics of desalination by integrating greenhouse gas emission costs: An empirical application for chile, Renewable Energy 133 (2019) 1327–1337.

[6] P. Asef, R. B. Perpina, M. Barzegaran, A. C. Lapthorn, A 3-D pareto-based shading analysis on solar photovoltaic system design optimization, IEEE Transactions on Sustainable Energy 10 (2019) 843–852.

[7] R. Benseman, F. Cook, Radiation in New Zealand-Standard year and radiation on inclined slopes, New Zealand Journal of Science 12 (1969) 696.

[8] I. J. Hall, R. Prairie, H. Anderson, E. Boes, Generation of a typical meteorological year, Technical Report, Sandia Labs., Albuquerque, NM (USA), 1978.

[9] H. Lund, The design reference year users manual, thermal insulation laboratory, Technical University of Denmark: Lyngby (1995).

[10] R. Festa, C. F. Ratto, Proposal of a numerical procedure to select reference years, Solar Energy 50 (1993) 9–17.

[11] A. Chan, Generation of typical meteorological years using genetic algorithm for different energy systems, Renewable Energy 90 (2016) 1–13.

[12] D. S. Wilks, R. L. Wilby, The weather generation game: a review of stochastic weather models, Progress in physical geography 23 (1999) 329–357.

[13] C. W. Richardson, Stochastic simulation of daily precipitation, temperature, and solar radiation, Water resources research 17 (1981) 182–190.

[14] K. Breinl, Driving a lumped hydrological model with precipitation output from weather generators of different complexity, Hydrological Sciences Journal 61 (2016) 1395–1414.

[15] M. Dubrovskỳ, J. Buchtele, Z. Žalud, High-frequency and low-frequency variability in stochastic daily weather generator and its effect on agricultural and hydrologic modelling, Climatic Change 63 (2004) 145–179.

[16] T. Kittel, N. Rosenbloom, T. Painter, D. Schimel, The vemap integrated database for modelling united states ecosystem/vegetation sensitivity to climate change, Journal of Biogeography (1995) 857–862.

[17] N. B. Pickering, J. R. Stedinger, D. A. Haith, Weather input for nonpoint-source pollution models, Journal of irrigation and drainage engineering 114 (1988) 674–690.

[18] L. Adélard, H. Boyer, F. Garde, J.-C. Gatina, A detailed weather data generator for building simulations, Energy and Buildings 31 (2000) 75–88.

[19] A. H. van Paassen, Q. Luo, Weather data generator to study climate change on buildings, Building Services Engineering Research and Technology 23 (2002) 251–258.

[20] A. Hering, K. Kazor, W. Kleiber, A markov-switching vector autoregressive stochastic wind generator for multiple spatial and temporal scales, Resources 4 (2015) 70–92.

[21] H. V. Haghi, S. Lotfifard, Spatiotemporal modeling of wind generation for optimal energy storage sizing, IEEE Transactions on Sustainable Energy 6 (2015) 113–121.

[22] B. Klöckl, G. Papaefthymiou, Multivariate time series models for studies on stochastic generators in power systems, Electric Power Systems Research 80 (2010) 265–276.

[23] V. Monbet, P. Ailliot, M. Prevosto, Survey of stochastic models for wind and sea state time series, Probabilistic engineering mechanics 22 (2007) 113–126.

[24] S. D. Campbell, F. X. Diebold, Weather forecasting for weather derivatives, Journal of the American Statistical Association 100 (2005) 6–16.

[25] R. H. Inman, H. T. Pedro, C. F. Coimbra, Solar forecasting methods for renewable energy integration, Progress in energy and combustion science 39 (2013) 535–576.

[26] B. Rajagopalan, U. Lall, A k-nearest-neighbor simulator for daily precipitation and other weather variables, Water resources research 35 (1999) 3089–3101.

[27] M. B. Parlange, R. W. Katz, An extended version of the richardson model for simulating daily weather variables, Journal of Applied Meteorology 39 (2000) 610–622.

[28] P. Ailliot, D. Allard, V. Monbet, P. Naveau, Stochastic weather generators: an overview of weather type models, Journal de la Société Française de Statistique 156 (2015) 101–113.

[29] A. P. Dempster, N. M. Laird, D. B. Rubin, Maximum likelihood from incomplete data via the em algorithm, Journal of the Royal Statistical Society: Series B (Methodological) 39 (1977) 1–22.

[30] S. Apipattanavis, G. Podestá, B. Rajagopalan, R. W. Katz, A semiparametric multivariate and multisite weather generator, Water Resources Research 43 (2007).

[31] J. L. Bentley, Multidimensional binary search trees used for associative searching, Communications of the ACM 18 (1975) 509–517.

[32] E. Koutroulis, D. Kolokotsa, Design optimization of desalination systems power-supplied by pv and w/g energy sources, Desalination 258 (2010) 171–181.

[33] E. Koutroulis, D. Kolokotsa, G. Stavrakakis, Optimal design and economic evaluation of a battery energy storage system for the maximization of the energy generated by wind farms in isolated electric grids, Wind Engineering 33 (2009) 55–81.

[34] T. S. Babu, J. P. Ram, T. Dragičević, M. Miyatake, F. Blaabjerg, N. Rajasekar, Particle swarm optimization based solar pv array reconfiguration of the maximum power extraction under partial shading conditions, IEEE Transactions on Sustainable Energy 9 (2018) 74–85.

[35] V. Monbet, NHMSAR: Non-Homogeneous Markov Switching Autoregressive Models, 2018. R package version 1.12.

[36] R. W. Katz, M. B. Parlange, Overdispersion phenomenon in stochastic modeling of precipitation, Journal of Climate 11 (1998) 591–601.

[37] F. M. Vesely, L. Paleari, E. Movedi, G. Bellocchi, R. Confalonieri, Quantifying uncertainty due to stochastic weather generators in climate change impact studies, Scientific reports 9 (2019).