

Article

Uncertainty Cost Functions in Climate-Dependent Controllable Loads in Commercial Environments

Daniel Losada ¹, Ameena Al-Sumaiti ²  and Sergio Rivera ^{1,*} 

¹ Electrical and Electronic Engineering, Universidad Nacional de Colombia, Sede Bogotá, Bogotá 111321, Colombia; dlosadar@unal.edu.co

² Advanced Power and Energy Center, Electrical Engineering and Computer Science, Khalifa University, Abu Dhabi 127788, United Arab Emirates; ameena.alsumaiti@ku.ac.ae

* Correspondence: srriverar@unal.edu.co

Abstract: This article presents the development, simulation and validation of the uncertainty cost functions for a commercial building with climate-dependent controllable loads, located in Florida, USA. For its development, statistical data on the energy consumption of the building in 2016 were used, along with the deployment of kernel density estimator to characterize its probabilistic behavior. For validation of the uncertainty cost functions, the Monte-Carlo simulation method was used to make comparisons between the analytical results and the results obtained by the method. The cost functions found differential errors of less than 1%, compared to the Monte-Carlo simulation method. With this, there is an analytical approach to the uncertainty costs of the building that can be used in the development of optimal energy dispatches, as well as a complementary method for the probabilistic characterization of the stochastic behavior of agents in the electricity sector.

Keywords: uncertainty cost; electricity demand; controllable load; probability density estimator; probability density function; expected value



Citation: Losada, D.; Al-Sumaiti, A.; Rivera, S. Uncertainty Cost Functions in Climate-Dependent Controllable Loads in Commercial Environments. *Energies* **2021**, *14*, 2885. <https://doi.org/10.3390/en14102885>

Academic Editors: Carlos Henggeler Antunes, Azeddine Houari, Saad Mekhilef and Mohamed Machmoum

Received: 7 March 2021

Accepted: 12 May 2021

Published: 17 May 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The economic dispatch of electrical energy is an activity of great relevance within the sector. In this, the fulfillment of the energy, technical and economic expectations of the different agents that participate in the entire process of energy production, transport and consumption are sought. For this reason, when an energy dispatch is carried out, it is sought to be optimal, to meet the objectives of the different agents, taking into account the efficient use of infrastructure resources and with minimal associated costs.

Due to the above, several efforts have been presented around the world, focused on achieving an optimal energy dispatch. However, the development of these methodologies faces several challenges associated with the different variables that must be taken into account in the dispatch. One of them has to do with the inclusion of agents that bring characteristics of randomness and uncertainty in their behavior. This represents a challenge, since this type of behavior is incompatible with the development of optimal energy dispatch methodologies with analytical and deterministic approaches.

The agents that bring uncertainty variables within the energy system are mainly focused on renewable energy generators, such as solar and wind, and on controllable loads such as electric vehicles or environmental thermal control devices. For the case of generation, the randomness occurs due to renewable energy sources, which depend on solar irradiation [1] and wind speed [2], respectively. Regarding the loads, the uncertainty is present due to the ambient temperature conditions [3–5] and the transportation needs in a given place [6–12]. These types of phenomena are non-deterministic, so it is not possible to directly obtain a certain value and, therefore, try to define the exact behavior of the agents that base their behavior on them.

Now, with the inclusion of these agents in the development of the activities of the energy sector, it is necessary that they also be taken into account when carrying out the economic dispatch of energy, and, moreover, that methodologies be developed so that this dispatch is optimum.

A recent methodology that seeks the inclusion of agents with random behavior in the economic dispatch is the development of uncertainty cost functions. This methodology allows us to obtain analytical cost equations, starting from probability density functions that describe the stochastic nature of these agents. With these equations, it is possible to carry out analytical economic dispatches that include all types of agents and that seek the optimization of both objectives and resources in all the main activities of the energy sector. The main characteristics of the uncertainty cost functions are detailed below.

In the past, different works have been developed that have sought to obtain uncertainty cost functions for both generating agents and demand agents, as can be seen in [10–15]. Most share the characteristic that, to obtain the analytical equations, predetermined probability density functions were assumed or approximated to characterize the random behaviors of the different agents of the sector under study in each work [16]. Although these equations obtained in this way are very useful when taking into account stochasticity in the economic dispatch, in several cases, they tend to skew certain behaviors of the agents they represent. This is because assuming or approximating random consumption or generation curves to probability distributions can eliminate characteristics of the analyzed curves [16–20].

Therefore, the need to find another way to characterize the consumption or generation curves of the agents that give stochasticity to the energy system has become evident, in order to obtain more precise representations of these agents in the development of economic dispatches. In the next sections, the state of the art, the methodological development, the results obtained and the conclusions are described, about an alternative method found to obtain more representative probability distributions of stochastic agents (controllable loads in the case of this work) and, as an ultimate goal, the uncertainty cost functions associated with its stochastic behavior.

2. Material and Methods

2.1. Uncertainty Cost

The costs of uncertainty associated with the economic dispatch of energy refer to the economic sanctions or penalties that may occur, due to the underestimation or overestimation of the energy that needs to be delivered or received at a specific point in the electrical system [12]. These situations of underestimation or overestimation are generally presented by the inclusion of agents of random behavior in the electrical network. As mentioned above, this type of behavior does not allow obtaining predetermined values of energy to be delivered, as in the case of non-conventional generators, or of energy to receive, such as controllable loads. For this reason, to dispatch these agents, it is necessary to obtain an estimate of the energy they can generate or consume. In an economic dispatch with these conditions, cases of non-compliance in the committed energy can occur, both on the generation side and on the demand side, which in turn will lead to economic sanctions.

In its basic form, the costs of under- or overestimation can be represented by the following equations.

$$C_{e,u,i}(P_{e,i}, P_{e,s,i}) = C_{e,u,i} (P_{e,i} - P_{e,s,i}) \quad (1)$$

$$C_{e,o,i}(P_{e,i}, P_{e,s,i}) = C_{e,o,i} (P_{e,s,i} - P_{e,i}) \quad (2)$$

where $C_{e,u,i}$ is the coefficient of costs for underestimating the energy to be consumed or delivered at node i ; $C_{e,o,i}$ is the cost coefficient for overestimating the energy to be consumed or delivered at node i ; $P_{e,i}$ is the real power delivered or demanded at node i ; and $P_{e,s,i}$ is the power programmed to deliver or consume at node i .

As can be seen from both equations, the costs are derived from the difference present between the programmed energy and the actual energy dispatched.

When applying these equations to the agents with stochastic behavior, the probability density functions that describe this behavior must be taken into account and that allow obtaining expected values of generated or consumed energy [10,12]. Therefore, it is necessary to use integrals of expected value in the cost equations for its overestimation or underestimation, having as the main variable the real energy demanded or delivered by the random agents. Considering this, the uncertainty cost equations for agents with stochastic behavior are presented below.

Uncertainty Costs for Overestimating

$$E[C_{e,o,i}(P_{e,i}, P_{e,s,i})] = \int_0^{P_{e,s,i}} C_{e,o,i}(P_{e,s,i} - P_{e,i}) f_{P_{e,i}}(P_{e,i}) dP_{e,i} \quad (3)$$

Uncertainty Costs for Underestimating

$$E[C_{e,u,i}(P_{e,i}, P_{e,s,i})] = \int_{P_{e,s,i}}^{P_{e,\infty}} C_{e,u,i}(P_{e,i} - P_{e,s,i}) f_{P_{e,i}}(P_{e,i}) dP_{e,i} \quad (4)$$

where $f_{P_{e,i}}(P_{e,i})$ is the probability density function used to describe the behavior of the stochastic agent and $P_{e,\infty}$ is the maximum scheduled demand that can be delivered to the controllable load by the system.

2.2. Agents of the Electricity Sector with Stochastic Behavior

As mentioned in Section 2.1, the uncertainty cost functions appear in agents that participate in the electric power dispatch process and that follow a random generation or consumption behavior, that is, not susceptible to being established in a deterministic way. This group of agents includes generators whose main source of energy production is a renewable resource. Renewable energy resources, such as the wind or the sun, are randomly time-varying, so anything that is directly related to them will have this same uncertainty characteristic. The foregoing also applies to another type of agent such as controllable loads. Electric vehicles that are connected within a distribution network in a city or automatic environmental conditioning devices within residential or commercial places, are examples of controllable loads that follow random behaviors associated with phenomena such as the need for mobility of people or the setting of the comfort temperature in a climate-dependent location, respectively.

As shown above, to obtain the uncertainty cost functions of these agents, it is necessary to use the probability density functions of the base phenomena of the agents behavior. In [12], it is possible to observe the procedure to obtain the uncertainty cost functions for photovoltaic and wind power generators. For wind generation, the probability density of wind speed [12] was taken into account, which in [11,15] appears as a Rayleigh distribution. Regarding photovoltaic generation, the behavior of solar irradiance was taken into account, which in [12] is assumed with a log-normal distribution. The above is true for specific regions, as specified in [21]. The process then focused on developing expected value integrals of the uncertainty costs taking into account these probability distributions, as detailed in Section 2.1.

Continuing with the characterization of non-conventional generation agents, in [13], you can see the determination of the uncertainty costs of run-of-river plants operating in a microgrid environment. As in [12], the behavior of these plants is subject to their source of generation, which in this case is the flow of the river from which they are supplied. In the literature [22–24], it is seen that an approximation can be made to the flow behavior through a Gumbel probability density. The remaining part of the process to characterize the uncertainty costs is followed, as in [12].

In the case of controllable loads, in [10], an approximation is made to their uncertainty costs through the analysis of integrals of expected value, taking into account probability distributions inherent to the behavior of controllable demand. The authors of [10–13] considered that the controllable loads can follow normal or beta distributions, so, in [10],

the mathematical analysis is performed with these distributions. As a result, analytical expressions were obtained that can be integrated in the development of an optimal energy dispatch.

2.3. Climate-Dependent Controllable Loads

The work carried out in [10] considers the uncertainty cost functions for controllable loads, particularly those associated with electric vehicles. As a complementary effort, the work detailed in this report sought to achieve the uncertainty costs for climate-dependent controllable loads found in commercial environments. Therefore, we proceed to detail this type of loads.

Over the years, several studies have been developed that have tried to establish a consistent relationship between energy consumption and the ambient temperature of the place where one lives. As can be intuitively understood, non-linearity occurs due to the search for an ambient temperature that offers comfort in the place where it is inhabited, through the use of ambient conditioning devices. Therefore, if the temperature is below or above the optimal comfort temperature, generally, electrical devices such as air conditioners or heaters will be used to reach it again. This will be reflected in energy consumption, which will have a minimum point that corresponds to the optimal comfort temperature and maximum points related to temperatures far from the optimum [16].

Continuing with the work developed (see, e.g., [16–18]), it can be seen that this non-linear relationship between ambient temperature and energy consumption is susceptible to being represented by a mathematical expression, when you have statistical data for the two variables. This is possible thanks to techniques such as linearization of the relationship through the use of variables exogenous to the model, multilinear regression and the use of dummy variables to weight the impact of different phenomena on the relationship to be analyzed.

With this, it is possible to show the existence of controllable loads that are closely related to climate and temperature, such as environmental conditioning devices, which will act to maintain an ambient temperature that is considered comfortable.

The main contribution of the proposed study is its applicability in the calculation of uncertainty cost functions (UCF) and determining the probability functions required to calculate the UCF using kernels. In this way, the cost functions of uncertainty of controllable loads through the investigation of probability distributions and the analytical development of integrals of expected value were determined, for their application in an efficient energy dispatch.

3. Methodology for Probability Density Estimation

This section presents information on how to approximate or obtain the probability density functions that represent the behavior of stochastic agents. These are necessary to be able to develop the uncertainty cost functions of the latter.

3.1. Parametric Probability Density Functions

As mentioned in previous sections, to find the uncertainty costs of a random agent, it is necessary to have a probability density function that best describes the agent's behavior. Doing a check on the different works that have addressed the issue of uncertainty costs [10–15], it can be seen that most of them have used standard parametric probability density functions. These functions can be normal, beta, Rayleigh, Gumbell distributions, etc. They are known as parametric because, with 1–3 parameters for each function, it is possible to define them in their entirety.

It is important to mention that, to define the type of density function that best suits an agent and its base parameters, it is necessary to have statistical information on its behavior. Once this information is available, it is possible to approximate the probability density functions to the consumption or generation curve of the agent in question, through the use of software tools, such as the application of adjustment of the probability distribution of

MatLab, among others. These tools allow the histogram of the statistical data of a random agent and the different probability density curves to be approximated to the histogram, with their parameters based on and calculated from the statistical data of the agent, to be viewed in the same graph. Bearing this in mind, the possibility is presented of seeing the density function that most closely approximates the agent. Figure 1 shows the behavior in one year of the controllable loads of the building that is part of the study of this article (the data come from a random building from over 1000 buildings over a three-year timeframe given by American Society of Heating, Refrigerating and Air Conditioning Engineers (ASHRAE) in the competition “ASHRAE—Great Energy Predictor” [25]; additionally, in Section 6, you can find more information on these loads) and the different probability densities that were chosen to represent it.

As shown in Figure 1, the density functions that would most represent the agent’s behavior would be the normal, Gamma or inverse Gaussian functions (the data are from [10–15]). It is also seen that functions such as Weibull or Rayleigh would not be very useful in this case. Following the methodology developed in [10–15], one could think of using some of the densities closest to the agent to calculate its uncertainty costs. Although the above would represent a good approximation of costs, if you pay more attention to the past graph, you can see that there are sections of the histogram that are oversized or that are not covered by the closest density functions. Then, the uncertainty costs calculated with these functions will omit certain agent-specific details associated with these missing parts of their behavior.

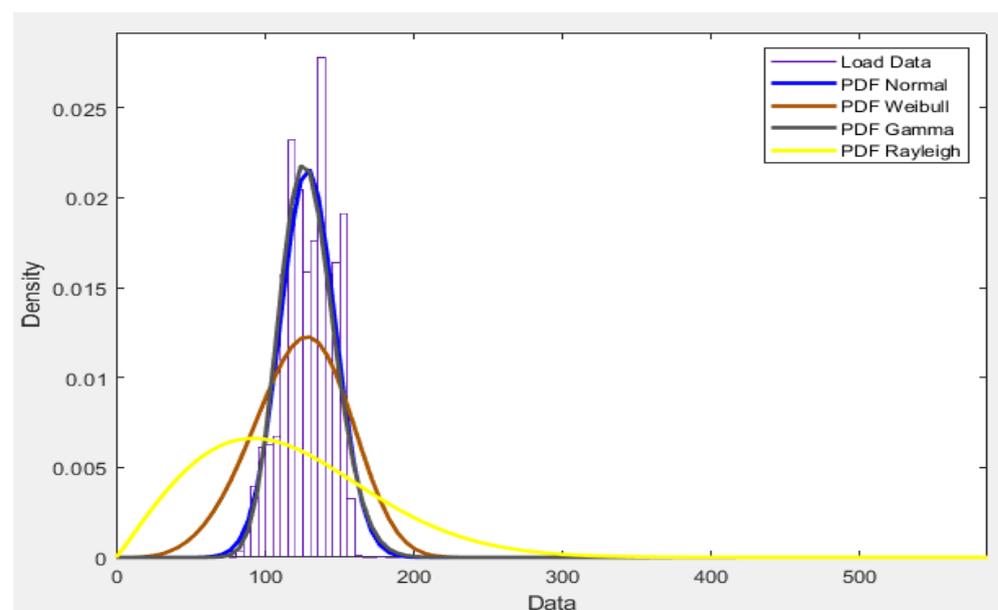


Figure 1. Comparison between agent behavior histogram and assumed probability density functions.

3.2. Nonparametric Probability Density Functions

As mentioned above, although it is possible to try to approximate the behavior of a stochastic agent through parametric probability density functions, this causes certain statistical data characteristic of the agent to be omitted. To obtain analytical uncertainty cost functions that can be used in the development of efficient economic dispatches, it is necessary to use a methodology that allows obtaining probability densities closer to the random behavior of the agent.

A methodology used to estimate probability densities, given a set of statistical data, is the application of a kernel density estimator. This methodology consists of approximating a probability density curve to the behavior data of a stochastic agent, through the use of a function composed of a summation of kernel functions, which seek to represent each

datum in the final density function. The composite function is known as the kernel density estimator. Here is an overview of the methodology mentioned above.

3.2.1. Kernel Density Estimator and Kernel Distribution

A kernel density estimator is a function that represents a special type of probability density known as a kernel distribution. This is characterized by being nonparametric, that is, it is not defined by one or more parameters. The kernel distribution is used when parametric density functions cannot properly describe the behavior of statistical data of a random variable, or when we want to avoid assumptions about their distribution [19].

If we have a series of n observations of an independent and identically distributed random variable (*iid*) of the form $\{x_1, x_2, \dots, x_n\}$, whose probability density $f(x)$ is unknown, a kernel density estimator $\hat{f}(x)$ can be used, which will approximate $f(x)$ by the sum of functions $K(x_i, t)$ (kernel function), assigned to each observation of the series. The above is summarized in the following expression:

$$\hat{f}(t) = \frac{1}{n} \sum_{i=1}^n K(x_i, t) \quad (5)$$

In more detail with respect to the previous expression, it can be seen that the estimator $\hat{f}(x)$ consists of an average of core functions ($K(x_i, t)$) which will have as main and individual parameter, each observation that belongs to the series [20]. It is worth mentioning that the kernel function is symmetric and generally smooth, with an arbitrary own parameter h , known as bandwidth, that controls the amount of smoothness which is wanted in the final estimated probability density curve [26]. The expression for $K(x_i, t)$ is shown below:

$$K(x_i, t) = \frac{1}{h} K_{kernel} \left(\frac{x_i - t}{h} \right) \quad (6)$$

From this expression, it should be noted that K_{kernel} is a parametric probability density function that can be arbitrarily selected, as long as it meets the following characteristics:

1. $K_{kernel}(x)$ is symmetric.
2. $\int_{-\infty}^{\infty} K_{kernel}(x) dx = 1$.
3. $\lim_{x \rightarrow -\infty} K_{kernel}(x) = \lim_{x \rightarrow \infty} K_{kernel}(x) = 0$.

When reviewing these characteristics, especially the second one, it can be seen that we are talking about a probability density function. Then, it should be clarified that, in its most common practical applications, the estimator uses a symmetric $K_{kernel}(x)$; however, recently we have seen the use of asymmetric functions as a kernel for the recreation of probability density functions [20]. Qualitative characteristics of the use of one or the other are detailed below.

3.2.2. How the Density Estimator Works

When there is a series of observations of a random variable, of the form $\{x_1, x_2, \dots, x_n\}$, the density estimator will take each x_i data as a main parameter of a small density function, graphically creating “bumps” of the form of the $K_{kernel}(x)$ function at each x_i . Then, all these functions will be averaged and weighted to obtain the final density function of the set of observations [26]. Next, a graphical example of the operation of the estimator presented in [26] is shown.

Figure 2 shows a dataset located at the points $x_i = \{0.1, 0.15, 0.2, 0.5, 0.7, 0.8\}$, which are represented by the black bars. Each of these points is assigned a proper density function ($K_{kernel}(x)$), which follows the behavior of a normal distribution in this case (purple curves). Then, the sum of these density functions typical of each point will lead to obtaining the estimated density curve for the dataset, represented by the brown curve.

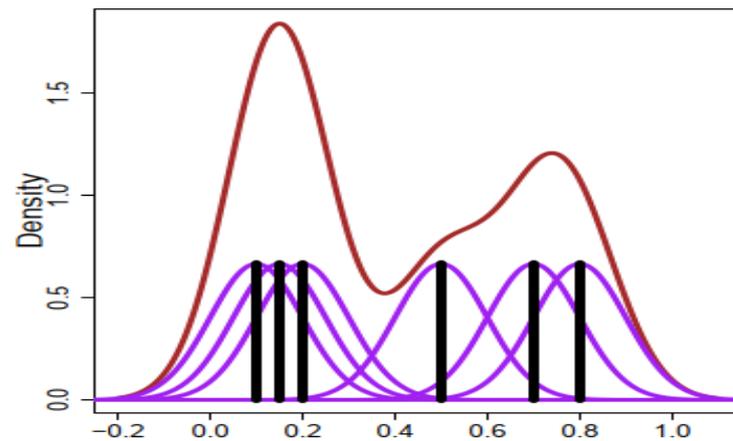


Figure 2. Graphic representation of the operation of the density estimator (simulated and modified from [26]).

3.2.3. Bandwidth h and Kernel Functions

As can be seen from Equation (6), the density estimator has two customization elements that can be adjusted as required in different applications. These elements are the bandwidth h and the kernel function $K_{kernel}(x)$.

In the configuration of a probability density estimator, the choice of the bandwidth parameter h plays an important role. This parameter is known by this name because it determines how much information of the base data series is reflected in the density function obtained from the estimator. It can be interpreted as a noise modulator in the final density curve [20].

A very small value of h can lead to the density curve obtained by the estimator showing insignificant or insignificant details. On the other hand, a very large value of h can cause a very smooth effect on the final curve, which could obviate important characteristics contained in the data information [20]. Figure 3 shows an example of the effect of choosing h .

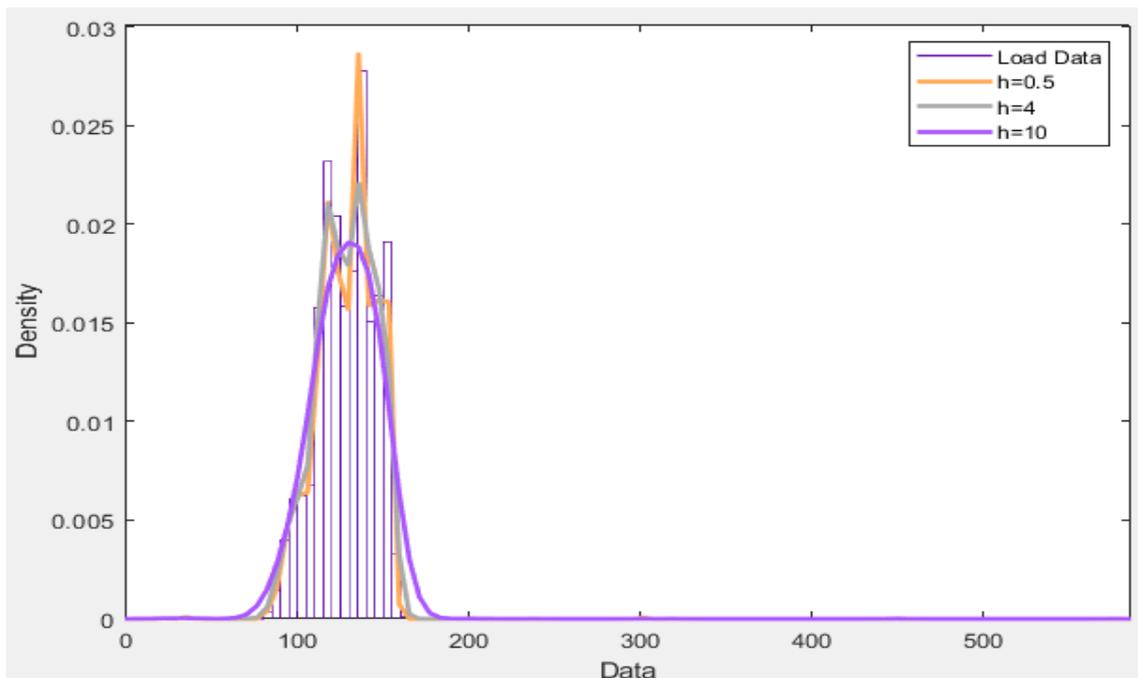


Figure 3. Effect of different values of h on the estimated density curve (simulated and modified from [26]).

Regarding the kernel function $K_{kernel}(x)$, it is worth mentioning that symmetric functions are the most used in the most common practical applications; however, recently there has been an increase in the use of asymmetric functions [20]. The main difference between the use of one or the other has to do with the form that the kernel function takes depending on the location point of the data of the base series. As shown in Figure 2, symmetric functions, e.g., the normal distribution, have shapes that do not vary depending on where they are. On the other hand, asymmetric functions show different shapes for the different points of the data series.

The most common symmetric kernel functions are:

$$\text{Gaussian or Normal} \quad K_{kernel}(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \tag{7}$$

$$\text{Uniform} \quad K_{kernel}(x) = \frac{1}{2} I(-1 \leq x \leq 1) \tag{8}$$

$$\text{Epanechnikov} \quad K_{kernel}(x) = \frac{3}{4} \max\{1 - x^2, 0\} \tag{9}$$

Apart from these functions, there are other symmetric and asymmetric functions that can be used as kernels of the density estimator. Węglarczyk [20] presented different kernel functions.

4. Methodology for Cost of Uncertainty

4.1. Methodology for Cost of Uncertainty from Underestimating Electricity Demand

Looking at the uncertainty cost of underestimation equation below, you can see that the expression is quite similar to the cost of overestimation equation, except for the order of the addends $P_{e,i}$ and $P_{e,s,i}$. Therefore, the methodology for obtaining the cost function for underestimation is the same as that used in the previous section.

$$E[C_{e,u,i}(P_{e,i}, P_{e,s,i})] = \int_{P_{e,s,i}}^{P_{e,\infty}} C_{e,u,i}(P_{e,i} - P_{e,s,i}) f_{P_{e,i}}(P_{e,i}) dP_{e,i} \tag{10}$$

Again, the density estimator with normal kernel function detailed in Section 6.2 is used:

$$f_{P_{e,i}}(P_{e,i}) = \frac{1}{\sqrt{2\pi nh}} \sum_{j=1}^n e^{-\frac{(P_{e,i} - \mu_j)^2}{2h^2}} \tag{11}$$

With this, the process of determining the cost function of the underestimation is:

$$\begin{aligned} E[C_{e,u,i}(P_{e,i}, P_{e,s,i})] &= C_{e,u,i} \left[\int_{P_{e,s,i}}^{P_{e,\infty}} P_{e,i} f_{P_{e,i}}(P_{e,i}) dP_{e,i} - \int_{P_{e,s,i}}^{P_{e,\infty}} P_{e,s,i} f_{P_{e,i}}(P_{e,i}) dP_{e,i} \right] \\ C_{e,u,i} &\left[\int_{P_{e,s,i}}^{P_{e,\infty}} P_{e,i} \frac{1}{\sqrt{2\pi nh}} \left(\sum_{j=1}^n e^{-\frac{(P_{e,i} - \mu_j)^2}{2h^2}} \right) dP_{e,i} - \int_{P_{e,s,i}}^{P_{e,\infty}} P_{e,s,i} \frac{1}{\sqrt{2\pi nh}} \left(\sum_{j=1}^n e^{-\frac{(P_{e,i} - \mu_j)^2}{2h^2}} \right) dP_{e,i} \right] \\ C_{e,u,i} &\left[\frac{1}{\sqrt{2\pi nh}} \int_{P_{e,s,i}}^{P_{e,\infty}} P_{e,i} \left(\sum_{j=1}^n e^{-\frac{(P_{e,i} - \mu_j)^2}{2h^2}} \right) dP_{e,i} - P_{e,s,i} \frac{1}{\sqrt{2\pi nh}} \int_{P_{e,s,i}}^{P_{e,\infty}} \left(\sum_{j=1}^n e^{-\frac{(P_{e,i} - \mu_j)^2}{2h^2}} \right) dP_{e,i} \right] \\ C_{e,u,i} &\left[\frac{1}{\sqrt{2\pi nh}} \sum_{j=1}^n \int_{P_{e,s,i}}^{P_{e,\infty}} P_{e,i} e^{-\frac{(P_{e,i} - \mu_j)^2}{2h^2}} dP_{e,i} - P_{e,s,i} \frac{1}{\sqrt{2\pi nh}} \sum_{j=1}^n \int_{P_{e,s,i}}^{P_{e,\infty}} e^{-\frac{(P_{e,i} - \mu_j)^2}{2h^2}} dP_{e,i} \right] \end{aligned}$$

Next, the proposed indefinite integrals are calculated, with which the following expression is obtained (solving the integrals through the error function identities of Appendix A, the expected value of the cost function is obtained):

$$\begin{aligned}
 & \frac{C_{e,u,i}}{\sqrt{2\pi n h}} \left[\sum_{j=1}^n \left(-h^2 e^{-\frac{(\mu_j - P_{e,i})^2}{2h^2}} - \sqrt{\frac{\pi}{2}} h \mu_j \operatorname{erf} \left(\frac{\mu_j - P_{e,i}}{\sqrt{2} h} \right) \right) \Big|_{P_{e,s,i}}^{P_{e,\infty}} \right. \\
 & \quad \left. - P_{e,s,i} \sum_{j=1}^n \left(-\sqrt{\frac{\pi}{2}} h \operatorname{erf} \left(\frac{\mu_j - P_{e,i}}{\sqrt{2} h} \right) \right) \Big|_{P_{e,s,i}}^{P_{e,\infty}} \right] \\
 & \frac{C_{e,u,i}}{\sqrt{2\pi n h}} \left[\left(-h^2 \sum_{j=1}^n e^{-\frac{(\mu_j - P_{e,i})^2}{2h^2}} \Big|_{P_{e,s,i}}^{P_{e,\infty}} - \sqrt{\frac{\pi}{2}} h \sum_{j=1}^n \mu_j \operatorname{erf} \left(\frac{\mu_j - P_{e,i}}{\sqrt{2} h} \right) \Big|_{P_{e,s,i}}^{P_{e,\infty}} \right) \right. \\
 & \quad \left. + P_{e,s,i} \sqrt{\frac{\pi}{2}} h \sum_{j=1}^n \operatorname{erf} \left(\frac{\mu_j - P_{e,i}}{\sqrt{2} h} \right) \Big|_{P_{e,s,i}}^{P_{e,\infty}} \right] \\
 E[C_{e,u,i}(P_{e,i}, P_{e,s,i})] &= \frac{C_{e,u,i}}{\sqrt{2\pi n h}} \left[-h^2 \sum_{j=1}^n \left(e^{-\frac{(\mu_j - P_{e,\infty})^2}{2h^2}} - e^{-\frac{(\mu_j - P_{e,s,i})^2}{2h^2}} \right) - \sqrt{\frac{\pi}{2}} h \sum_{j=1}^n \mu_j \right. \\
 & \left. \left(\operatorname{erf} \left(\frac{\mu_j - P_{e,\infty}}{\sqrt{2} h} \right) - \operatorname{erf} \left(\frac{\mu_j - P_{e,s,i}}{\sqrt{2} h} \right) \right) + P_{e,s,i} \sqrt{\frac{\pi}{2}} h \sum_{j=1}^n \left(\operatorname{erf} \left(\frac{\mu_j - P_{e,\infty}}{\sqrt{2} h} \right) - \operatorname{erf} \left(\frac{\mu_j - P_{e,s,i}}{\sqrt{2} h} \right) \right) \right]
 \end{aligned}$$

Finally, taking into account that all the terms of the expression carry the sum $\sum_{j=1}^n$, it is possible to combine it to obtain the uncertainty cost function to be underestimated, as follows:

$$\begin{aligned}
 E[C_{e,u,i}(P_{e,i}, P_{e,s,i})] &= \sum_{j=1}^n \frac{C_{e,u,i}}{\sqrt{2\pi n h}} \left[P_{e,s,i} \sqrt{\frac{\pi}{2}} h \left(\operatorname{erf} \left(\frac{\mu_j - P_{e,\infty}}{\sqrt{2} h} \right) - \operatorname{erf} \left(\frac{\mu_j - P_{e,s,i}}{\sqrt{2} h} \right) \right) \right. \\
 & \left. - \sqrt{\frac{\pi}{2}} h \mu_j \left(\operatorname{erf} \left(\frac{\mu_j - P_{e,\infty}}{\sqrt{2} h} \right) - \operatorname{erf} \left(\frac{\mu_j - P_{e,s,i}}{\sqrt{2} h} \right) \right) - h^2 \left(e^{-\frac{(\mu_j - P_{e,\infty})^2}{2h^2}} - e^{-\frac{(\mu_j - P_{e,s,i})^2}{2h^2}} \right) \right] \quad (12)
 \end{aligned}$$

4.2. Uncertainty Cost for Overestimating Electricity Demand

As mentioned in Section 2.1, the determination of the analytical cost function for overestimating the electricity demand can be calculated by means of the following equation.

$$E[C_{e,o,i}(P_{e,i}, P_{e,s,i})] = \int_0^{P_{e,s,i}} C_{e,o,i}(P_{e,s,i} - P_{e,i}) f_{P_{e,i}}(P_{e,i}) dP_{e,i} \quad (13)$$

With the above, we proceed to calculate the expected value of the cost to be overestimated. First, the separation of the integral is performed on the addends of the expression.

$$E[C_{e,o,i}(P_{e,i}, P_{e,s,i})] = C_{e,o,i} \left[\int_0^{P_{e,s,i}} P_{e,s,i} f_{P_{e,i}}(P_{e,i}) dP_{e,i} - \int_0^{P_{e,s,i}} P_{e,i} f_{P_{e,i}}(P_{e,i}) dP_{e,i} \right]$$

Then, the replacement of the density function $f_{P_{e,i}}(P_{e,i})$

$$\begin{aligned}
 C_{e,o,i} & \left[\int_0^{P_{e,s,i}} P_{e,s,i} \frac{1}{\sqrt{2\pi n h}} \left(\sum_{j=1}^n e^{-\frac{(P_{e,i} - \mu_j)^2}{2h^2}} \right) dP_{e,i} - \int_0^{P_{e,s,i}} P_{e,i} \frac{1}{\sqrt{2\pi n h}} \left(\sum_{j=1}^n e^{-\frac{(P_{e,i} - \mu_j)^2}{2h^2}} \right) dP_{e,i} \right] \\
 C_{e,o,i} & \left[P_{e,s,i} \frac{1}{\sqrt{2\pi n h}} \int_0^{P_{e,s,i}} \left(\sum_{j=1}^n e^{-\frac{(P_{e,i} - \mu_j)^2}{2h^2}} \right) dP_{e,i} - \frac{1}{\sqrt{2\pi n h}} \int_0^{P_{e,s,i}} P_{e,i} \left(\sum_{j=1}^n e^{-\frac{(P_{e,i} - \mu_j)^2}{2h^2}} \right) dP_{e,i} \right]
 \end{aligned}$$

$$C_{e,o,i} \left[P_{e,s,i} \frac{1}{\sqrt{2\pi nh}} \sum_{j=1}^n \int_0^{P_{e,s,i}} e^{-\frac{(P_{e,i}-\mu_j)^2}{2h^2}} dP_{e,i} - \frac{1}{\sqrt{2\pi nh}} \sum_{j=1}^n \int_0^{P_{e,s,i}} P_{e,i} e^{-\frac{(P_{e,i}-\mu_j)^2}{2h^2}} dP_{e,i} \right]$$

Next, the indefinite integrals are calculated, with which the following expression is obtained:

$$\frac{C_{e,o,i}}{\sqrt{2\pi nh}} \left[P_{e,s,i} \sum_{j=1}^n \left(-\sqrt{\frac{\pi}{2}} h \operatorname{erf} \left(\frac{\mu_j - P_{e,i}}{\sqrt{2} h} \right) \right) \Big|_0^{P_{e,s,i}} - \sum_{j=1}^n \left(-h^2 e^{-\frac{(\mu_j - P_{e,i})^2}{2h^2}} - \sqrt{\frac{\pi}{2}} h \mu_j \operatorname{erf} \left(\frac{\mu_j - P_{e,i}}{\sqrt{2} h} \right) \right) \Big|_0^{P_{e,s,i}} \right]$$

Subsequently, a reorganization of terms is carried out:

$$\frac{C_{e,o,i}}{\sqrt{2\pi nh}} \left[\left(h^2 \sum_{j=1}^n e^{-\frac{(\mu_j - P_{e,i})^2}{2h^2}} \Big|_0^{P_{e,s,i}} + \sqrt{\frac{\pi}{2}} h \sum_{j=1}^n \mu_j \operatorname{erf} \left(\frac{\mu_j - P_{e,i}}{\sqrt{2} h} \right) \Big|_0^{P_{e,s,i}} \right) - P_{e,s,i} \sqrt{\frac{\pi}{2}} h \sum_{j=1}^n \operatorname{erf} \left(\frac{\mu_j - P_{e,i}}{\sqrt{2} h} \right) \Big|_0^{P_{e,s,i}} \right]$$

Afterwards, the points of the definite integral are replaced in the variable $P_{e,i}$

$$E[C_{e,o,i}(P_{e,i}, P_{e,s,i})] = \frac{C_{e,o,i}}{\sqrt{2\pi nh}} \left[h^2 \sum_{j=1}^n \left(e^{-\frac{(\mu_j - P_{e,s,i})^2}{2h^2}} - e^{-\frac{\mu_j^2}{2h^2}} \right) + \sqrt{\frac{\pi}{2}} h \sum_{j=1}^n \mu_j \left(\operatorname{erf} \left(\frac{\mu_j - P_{e,s,i}}{\sqrt{2} h} \right) - \operatorname{erf} \left(\frac{\mu_j}{\sqrt{2} h} \right) \right) - P_{e,s,i} \sqrt{\frac{\pi}{2}} h \sum_{j=1}^n \left(\operatorname{erf} \left(\frac{\mu_j - P_{e,s,i}}{\sqrt{2} h} \right) - \operatorname{erf} \left(\frac{\mu_j}{\sqrt{2} h} \right) \right) \right]$$

Finally, taking into account that all the terms of the expression carry the sum $\sum_{j=1}^n$, it is possible to combine them to obtain the uncertainty cost function for overestimation, as follows:

$$E[C_{e,o,i}(P_{e,i}, P_{e,s,i})] = \sum_{j=1}^n \frac{C_{e,o,i}}{\sqrt{2\pi nh}} \left[h^2 \left(e^{-\frac{(\mu_j - P_{e,s,i})^2}{2h^2}} - e^{-\frac{\mu_j^2}{2h^2}} \right) + \sqrt{\frac{\pi}{2}} h \mu_j \left(\operatorname{erf} \left(\frac{\mu_j - P_{e,s,i}}{\sqrt{2} h} \right) - \operatorname{erf} \left(\frac{\mu_j}{\sqrt{2} h} \right) \right) - P_{e,s,i} \sqrt{\frac{\pi}{2}} h \left(\operatorname{erf} \left(\frac{\mu_j - P_{e,s,i}}{\sqrt{2} h} \right) - \operatorname{erf} \left(\frac{\mu_j}{\sqrt{2} h} \right) \right) \right] \tag{14}$$

Please note that the uncertainty cost functions for the analytical method does not depend of $P_{e,i}$, since in the mathematical development $f_{P_{e,i}}(P_{e,i})$ is used, i.e., the probability density function used to describe the behavior of the stochastic $P_{e,i}$.

In this way, the main contributions of the proposed methodology are:

- Probability distributions able to handle the behavior of controllable loads in a residential, industrial or commercial context, by means of the applications proposed in the literature review are found.
- The characteristic equations of the uncertainty costs of controllable loads by means of the analytical development of integrals of expected value are obtained .
- An efficient energy dispatch that considers optimization of the costs of uncertainty of controllable loads is performed.

5. Simulation and Validation of Uncertainty Cost Functions

The validation of the cost functions (12) and (14), obtained in the previous section, was carried out from the methodology used in [10], with the exception of obtaining the density function probability, which is the result of the application of the density estimator in the series of statistical data of Building 944, as mentioned above.

5.1. Matlab Kernel Density Estimator

The numerical simulation software tool Matlab has an application with its own interface for the analysis of statistical data from the fitting of probability curves. This application is known as *Distribution Fitter*, and it is capable of fitting parametric or nonparametric density curves to statistical dataset that is to be analyzed and supplied as input to the system. Information about the use of this tool can be found in [27].

Using the Distribution Fitter, it was possible to obtain the probability density function based on the statistical data of Building 944, by using the density estimator with a normal type kernel function. Figure 4 shows the density curve obtained through the estimator and its comparison with parametric density functions. As can be seen, the estimator curve (red) has a more faithful fit to the data from Building 944, compared to the other functions, even with the normal function and the gamma function. This makes it the optimal option to represent the building's probability density function.

In addition, it should be mentioned that the Distribution Fitter automatically provided the value of the h bandwidth, which was 3.08079. As mentioned above, this value can be arbitrarily selected, so that a more or less smooth curve can be obtained, as required. With this value of h together with the dataset for Building 944, it is possible to simulate the cost functions (12) and (14).

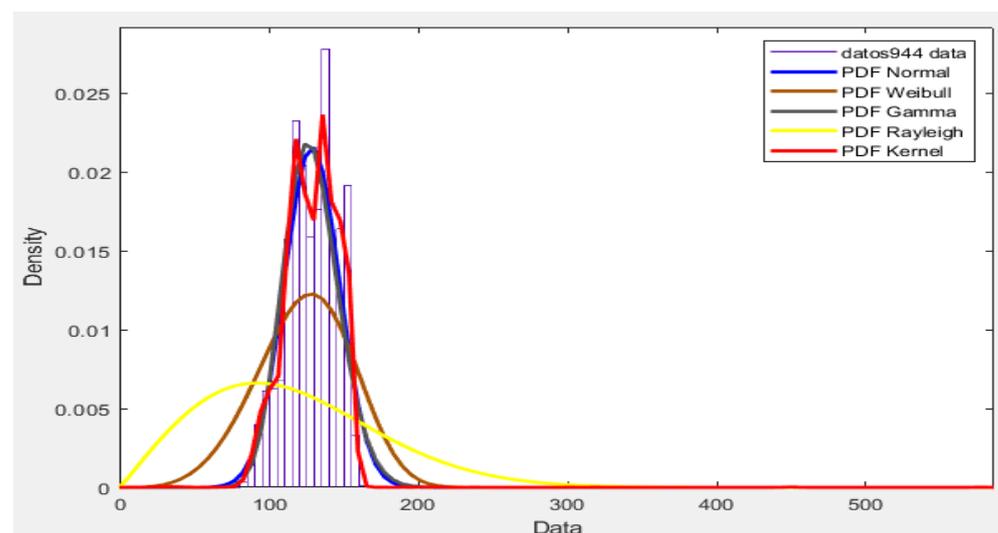


Figure 4. Probability density function obtained from the density estimator for demand (kW) in Building 944.

5.2. Monte-Carlo Simulation

Similar to the methodology used in [10], to validate the uncertainty cost functions found, simulations were carried out with the Monte-Carlo simulation method. With this, a comparison could be made between the costs obtained from the analytical functions and those obtained from the simulations.

We next describe the validation algorithm. In a first stage, the random obtaining of the energy consumption values with nonparametric probability was carried out using the Monte-Carlo simulations. Then, we proceeded to determine if the consumption values fall into the categories of overestimation or underestimation. In case they did not enter either of the two categories, we went back to the first stage. Then, the uncertainty cost vector was formed, from which its mean value was obtained. In the final stage, this mean value was compared with the analytical calculation from the cost functions found, and the percentage error was found.

It is worth mentioning that Matlab's Distribution Fitter application offers the possibility of exporting the data of its density curve fits to the workspace, with which the characteristics

of the nonparametric probability density function of Building 944 could be included in the simulation code.

In this way, the main steps used to perform the Monte-Carlo simulation are as follows:

1. A demand value is established that represents the power programmed by a network operator, normally from an economic dispatch model.
2. A Monte Carlo scenario is generated through a random value generated for a random generator according to the probability distribution (Section 2).
3. Given the value of the previous point, the value of the available demand is determined.
4. In this Monte Carlo scenario, the cost is evaluated: if it is underestimated, the corresponding equation is used, and, similarly, if it is overestimated, the corresponding equation is used.
5. Steps 2–4 are repeated several times.
6. A cost histogram is obtained where all Monte-Carlo scenarios are considered.
7. The expected value of the total accumulated cost is calculated, this quantity is the expected value of the uncertainty cost function and it is compared with the analytical value.

6. Data of the Controllable Load in Study

For the development of the work, statistical data in which hourly records of energy consumption and ambient temperature of buildings of different uses were evidenced, in Florida, USA, for the year 2016. In a first effort, a data treatment was carried out so that daily consumption and temperature records could be obtained in order to facilitate their manipulation. For the case of this article, the data of a building identified with the number 944 (hereinafter, Building 944) were used (a random building from over 1000 buildings over a three-year timeframe given by American Society of Heating, Refrigerating and Air Conditioning Engineers (ASHRAE) in the competition “ASHRAE—Great Energy Predictor”), due to its purpose for commercial use and the dependency relationship between energy consumption and ambient temperature, which is detailed below.

6.1. Climate Dependence

First, the dependence of the energy consumption data of the load under study, with respect to its ambient temperature, was determined. For the case of Building 944, when plotting its energy consumption data against its ambient temperature data, it was possible to visualize, in the first instance, the non-linear relationship with a “v” shape, which is characteristic in loads that present dependence between its consumption and ambient temperature. As mentioned in [17], this type of behavior is the result of the use of climate-controllable loads such as heaters or air conditioners (room temperature control devices), which will activate to reach the temperature of comfort. Figure 5 presents the graph.

Then, as mentioned in Section 2.3, it is possible to obtain a mathematical expression that reflects the dependence of electricity demand on the ambient temperature for a controllable load. The necessary procedures that were carried out for the development of the expression for Building 944 are detailed below.

Multivariate regression

To obtain the mathematical expression of the dependence of the energy consumption on the ambient temperature, the multivariate regression method was used. This method consists of expressing a variable (energy consumption in this case), as the result of the weighted sum of the different variables that may be related to this [17]. Next, the variables used to carry out the multivariate regression are detailed.

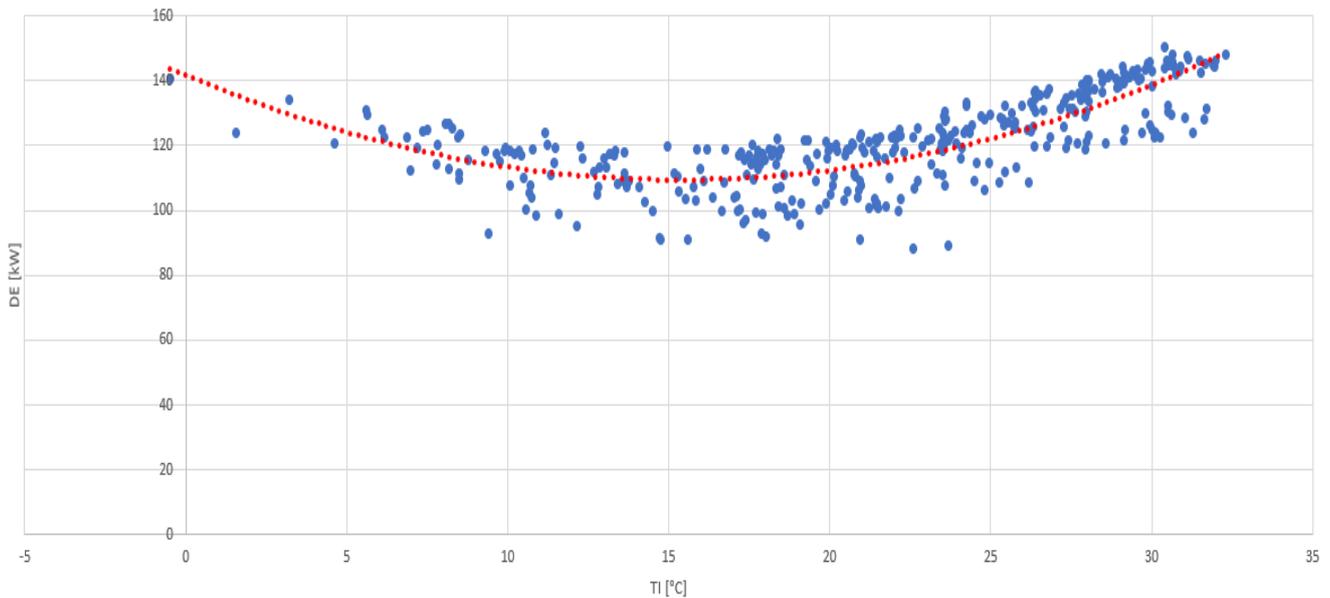


Figure 5. Demand of power electricity vs. temperature curve behavior, Building 944.

Exogenous temperature variables

After visualizing the non-linear behavior of the relationship between energy consumption and ambient temperature in the building, we proceeded to linearize the curve by associating exogenous temperature variables to the expression. These variables are the Cooling Degree Days and the Heating Degree Days (*CDD* and *HDD*, respectively) [17]. These variables are related to temperature through the following expressions:

$$CDD = \max\{0, T_{amb} - T_{opt}\}$$

$$HDD = \max\{0, T_{opt} - T_{amb}\}$$

Below, when reviewing the previous equations, it can be seen that, when one “acts”, the other will be at 0, because the room temperature T_{amb} will be on one side or the other of the optimal temperature T_{opt} . With this, the points of the curve below T_{opt} are covered by *HDD* and the points of the curve above T_{opt} are covered by *CDD*.

Trend and seasonality variables

As can be seen in previous works [16,17], the behavior of energy consumption can present trends and seasonality related to the time series. The above refers to repetitive patterns or increases and decreases in the time series, sustained over time. These effects are reflected in the electricity demand due to certain behaviors that occur in consumption and that depend on the day of the week or the time of the year in which a certain demand record occurs. To take these effects into account, dummy variables (variables that take values of 1 when a condition is met and 0 otherwise) were introduced in the expression that represent the days of the week, the months of the year, holidays and the days immediately preceding holidays. As explained in previous works [16,17], these variables are taken into account to see the impact of both trend and seasonality on the expression.

Autoregression

Another effect that can occur in the treatment and analysis of time series is autoregression, as mentioned in [18]. This refers to the relationship that a datum in the series may have with previously recorded data. In this work, it is the dependence of the energy consumption that occurs in a day, with the consumption that has occurred in previous days. In [28], the presence of autoregression in time series related to electricity demand is evidenced, for which lag’ variables were introduced to be able to observe the dependence of the series up to three days ago.

Structure of the regression expression

Taking into account the above variables, below you can see the equation used to characterize the relationship between consumption and temperature of Building 944.

$$DE = C + e * t + a * CDD(T_{amb}) + b * HDD(T_{amb}) + \sum_{i=1}^7 d_i * Da_i + \sum_{j=1}^{12} m_j * Mes_j + h_1 * Hol + h_2 * Hol_{T-1} + \sum_{k=1}^3 l_k * Lag_k$$

where C is a constant; t is the variable time; e is the coefficient of the variable time; D is the Day; Mes is the Month; Hol is the Holiday; a is the coefficient of the variable CDD ; b is the coefficient of the variable HDD ; d_i is the coefficient of the variable Day i of the week; m_j is the coefficient of the variable Month j of the year; h_1 is the coefficient of the Holiday variable; h_2 is the coefficient of the variable Day before holiday; and l_k is the coefficient of the variable lag k .

With the expression complete, it is possible to construct the regression equation using the Matlab functionality of multiple linear regression “*regress*”. With this, in terms of the similarity of the replicated data from the found expression and the base data of Building 944, a correlation of R^2 of 0.9154 was obtained. Nevertheless, a high value of the correlation coefficient of 0.9154 does not yet prove the adequacy of the regression model; indeed, both F-test and t -test yielded p -values < 0.05 for several buildings (Table 1), which means the null hypotheses is rejected. In this way, the presented results of the multidimensional regression model do not justify the conclusions about the influence of temperature on the electricity demand (consumption), but this study assumes that, for the buildings of ASHRAE “ASHRAE—Great Energy Predictor” [25], there is a direct influence [29–33].

Over the years, several studies have been developed that have tried to determine a consistent relationship between demand and ambient temperature, in order to establish the former as a dependent variable. Taking into account characteristics of the behavior of ambient temperature such as randomness, seasonality and trends, the methods to find its relationship with electricity demand have used statistical (regression) and probabilistic (expected value functions) tools. The proposal is a linear multiple regression statistical procedure to validate this assumption for the set of ASHRAE buildings. Initially, there were hourly data on both demand and the building’s ambient temperature. These were averaged to obtain daily data. Once we have the daily data, we proceed to obtain the characteristic curve of the behavior between electricity demand and temperature. The curve is characterized by having a “V” shape. Intuitively, it is understood that, when a place is inhabited, there will be an optimal comfort temperature (minimum point), and that, when the temperature is far from the optimal one, electrical devices (heaters or air conditioners, depending on the case) are used to return to the temperature again comfort, thus increasing demand at extreme points of temperature.

Additionally, this assumption comes from the comparison of the adjusted model found for the analyzed buildings with and without autoregression effect over ASHRAE buildings (Table 1).

As shown in Table 1, in most cases, the models found through the multiple regressions have a high degree of similarity to the data examined for each building. Most of the models have an R^2 adjustment above 0.8, so the results simulated based on these have a relationship greater than 80% of similarity to the input data of the regressions. Furthermore, it is noteworthy that several models (5) have an R^2 close to or above 0.9 (90%).

Table 1. Analyzed ASHRAE buildings with and without autoregression.

Building	R^2 without AR	R^2 with AR	p -Value for F-Test	p -Value for t -Test
892	0.67119	0.699674	0.057	0
944	0.909277	0.915403	0	0
970	0.907186	0.930692	0.999	0
982	0.89748	0.918719	0	0
984	0.529493	0.717893	0.999	0
1115	0.891994	0.915112	0	0
1118	0.765109	0.878194	1	0
1120	0.924841	0.949972	0	0
1197	0.477236	0.901832	1	0

It is also worth mentioning the impact of the autoregression phenomenon on the models. The results of a first comparison reveal that in all the analyzed cases there is an improvement in the similarity in the models when autoregression is considered as a descriptive variable of energy consumption in buildings. This shows a dependence of energy consumption dependent on consumption in past days, speaking only of statistical data.

In particular, the variation that occurred in Buildings 984 and 1197 is highlighted by considering the effect of autoregression on the models. As shown in Table 1, the first goes from an R^2 of 0.529493 to one of 0.717893 and the second from an R^2 of 0.477236 to one of 0.901832. This implies that the behavior of the buildings is strongly influenced by autoregression, to the point that the found models initially have a low similarity relationship and then one with a high descriptive capacity of their behavior. This confirmed the relevance of the expression developed and therefore the dependence of the consumption data with respect to the ambient temperature of Building 944.

6.2. Building 944 Probability Density Function

After characterizing the dependence of energy consumption with respect to ambient temperature, the work focused on consulting probability distributions to be able to represent the statistical data of energy consumption of Building 944. After a literature review, the kernel density estimator described in Section 3.2.1 is reached. Thus, to perform the representation of the probability density of Building 944, the Gaussian kernel function of the Equation (7) was chosen.

To proceed with the calculation of the expected value integral of the uncertainty cost functions, it is first necessary to establish the probability density function $f_{P_{e,i}}(P_{e,i})$ that describes the behavior of Building 944. As mentioned above, for this density function, the density estimator with a normal type kernel function is used, applied in the statistical data of Building 944. Obtaining the expression of the estimator to be used is achieved by replacing function (7) in Equation (6) and finally in expression (5). This leads to the probability density function for Building 944, represented by the following equation.

$$f_{P_{e,i}}(P_{e,i}) = \frac{1}{\sqrt{2\pi}nh} \sum_{j=1}^n e^{-\frac{(P_{e,i}-\mu_j)^2}{2h^2}} \quad (15)$$

where μ_j is the data series for Building 944; n is the total of data in the series; and $P_{e,i}$ is the real energy demand as an integration variable.

7. Results and Discussion

In this section, the results obtained from simulation and validation are shown, for different conditions of dispatched power and maximum power that can be dispatched. This is to check the operation of the equations obtained, subjected to different conditions. Together, an analysis of the results is developed. It should be remembered that the power dispatched or scheduled power is P_s and the maximum power to dispatch is $P_{e,\infty}$. For these cases, the minimum, average and maximum consumption values of Building 944 are taken into account, identified as $min944 = 8$ kWh (Case 1), $mean944 = 128.2749$ kWh (Case 2) and $max944 = 577$ kWh (Case 3), respectively.

The results shown below are from 25 simulations for each case. It is worth mentioning that, for each simulation, 10,000 generations of random numbers were carried out for the consumption of Building 944, part of the Monte-Carlo simulation method.

7.1. Case 1

In Table 2 is shown the case 1 where $P_s = min944$ and $P_{e,\infty} = max944$.

Table 2. Case 1 Results.

No. Simulation	Analytic Functions [\$]	Monte-Carlo [\$]	Error [%]
1	36,072.6	36,137.22	0.178829
2	36,072.6	36,050.66	0.060865
3	36,072.6	36,039.28	0.092459
4	36,072.6	36,090.44	0.049432
5	36,072.6	36,098.55	0.071874
6	36,072.6	36,008.41	0.178269
7	36,072.6	36,092.01	0.053771
8	36,072.6	36,029.81	0.118759
9	36,072.6	36,152.23	0.220251
10	36,072.6	36,007.75	0.180094
11	36,072.6	36,081.87	0.02568
12	36,072.6	36,126.22	0.148435
13	36,072.6	36,189.79	0.323822
14	36,072.6	36,073.91	0.003618
15	36,072.6	36,051.01	0.059894
16	36,072.6	36,076.51	0.010837
17	36,072.6	36,120.46	0.132505
18	36,072.6	36,073.58	0.002728
19	36,072.6	36,030.64	0.116449
20	36,072.6	36,037.09	0.098528
21	36,072.6	36,127.66	0.152397
22	36,072.6	36,103.9	0.086697
23	36,072.6	36,080.71	0.022481
24	36,072.6	36,095.15	0.062464
25	36,072.6	36,082.2	0.0266

7.2. Case 2

In Table 3 is shown the case 2 where $P_s = \text{mean}944$ and $P_{e,\infty} = \text{max}944$.

Table 3. Case 2 Results.

No. Simulation	Analytic Functions [\$]	Monte-Carlo [\$]	Error [%]
1	7115.163	7049.649	0.929327
2	7115.163	6999.377	1.654234
3	7115.163	7123.705	0.119909
4	7115.163	7003.204	1.598682
5	7115.163	7054.541	0.859343
6	7115.163	7178.322	0.879848
7	7115.163	7153.225	0.532096
8	7115.163	7220.486	1.458668
9	7115.163	7170.19	0.767442
10	7115.163	6986.23	1.845533
11	7115.163	7074.522	0.574472
12	7115.163	7230.857	1.599998
13	7115.163	7209.265	1.305292
14	7115.163	7134.508	0.271147
15	7115.163	7057.038	0.823645
16	7115.163	7059.236	0.79226
17	7115.163	7189.986	1.040647
18	7115.163	7101.062	0.198582
19	7115.163	7085.249	0.422207
20	7115.163	7069.38	0.647635
21	7115.163	7226.585	1.541837
22	7115.163	7112.49	0.037587
23	7115.163	7136.004	0.292049
24	7115.163	7113.682	0.020829
25	7115.163	7104.051	0.156421

7.3. Case 3

In Table 4 is shown the case 3 where $P_s = \text{max}944$ and $P_{e,\infty} = \text{max}944$.

The coefficient of costs for underestimating the energy $c_{e,u,i}$ is 30 USD/MW and the cost coefficient for overestimating the energy $c_{e,o,i}$ is 70 USD/MW using the data from previous studies [10–12]. The reason for large differences in costs for Cases 1–3 (USD 36,072.6, 7115.2 and 314,107.4, respectively) is that for the extreme case (Cases 1 and 3) the uncertainty cost increase due the probability distribution of the real demand ($P_{e,i}$: the real power delivered or demanded at node i). $P_{e,s,i}$ (power programmed to deliver or consume at node i) is the set value for the controllable load to be determined by a optimization problem of economic dispatch, where there are decision variables not only of the power to be injected for the generation agents but also the controllable demand values to be schedules, that is to say, it is a exogenous variable assumed in the regression model.

Table 4. Case 3 Results.

No. Simulation	Analytic Functions [\$]	Monte-Carlo [\$]	Error [%]
1	314,107.4	313,879.4	0.07265
2	314,107.4	314,280.1	0.054924
3	314,107.4	314,133.1	0.008178
4	314,107.4	314,038.3	0.022014
5	314,107.4	314,319.7	0.067513
6	314,107.4	314,210.6	0.032829
7	314,107.4	314,169.5	0.019739
8	314,107.4	314,103.5	0.001266
9	314,107.4	314,069.1	0.012208
10	314,107.4	314,275.2	0.053363
11	314,107.4	313,977.7	0.041311
12	314,107.4	314,042.7	0.020603
13	314,107.4	313,985.6	0.03882
14	314,107.4	314,227.9	0.038319
15	314,107.4	314,052.4	0.01752
16	314,107.4	314,166.2	0.018694
17	314,107.4	314,226.6	0.037919
18	314,107.4	314,059.5	0.015256
19	314,107.4	314,039	0.021782
20	314,107.4	313,847.1	0.08294
21	314,107.4	314,411.9	0.096821
22	314,107.4	314,161	0.017051
23	314,107.4	314,066.3	0.013106
24	314,107.4	313,958.9	0.047306
25	314,107.4	314,226	0.037736

7.4. Analysis of Results

The results obtained for Cases 1–3 show that the percentage differentials are minimal. Figures 6–8 present detailed graphical comparisons between the analytical functions and the Monte Carlo method for these cases. Visually, it can be seen that the values tend to be quite similar, in some simulations almost the same. It should be mentioned that Cases 1–3 have in common that the maximum energy to dispatch is equal to the record of maximum energy consumed by Building 944.

Table 5 presents the average error in the 25 simulations for Cases 1–3. These errors do not exceed 1%, which confirms that the analytical functions found represent a good approximation to the uncertainty costs modeled from the Monte-Carlo simulation method.

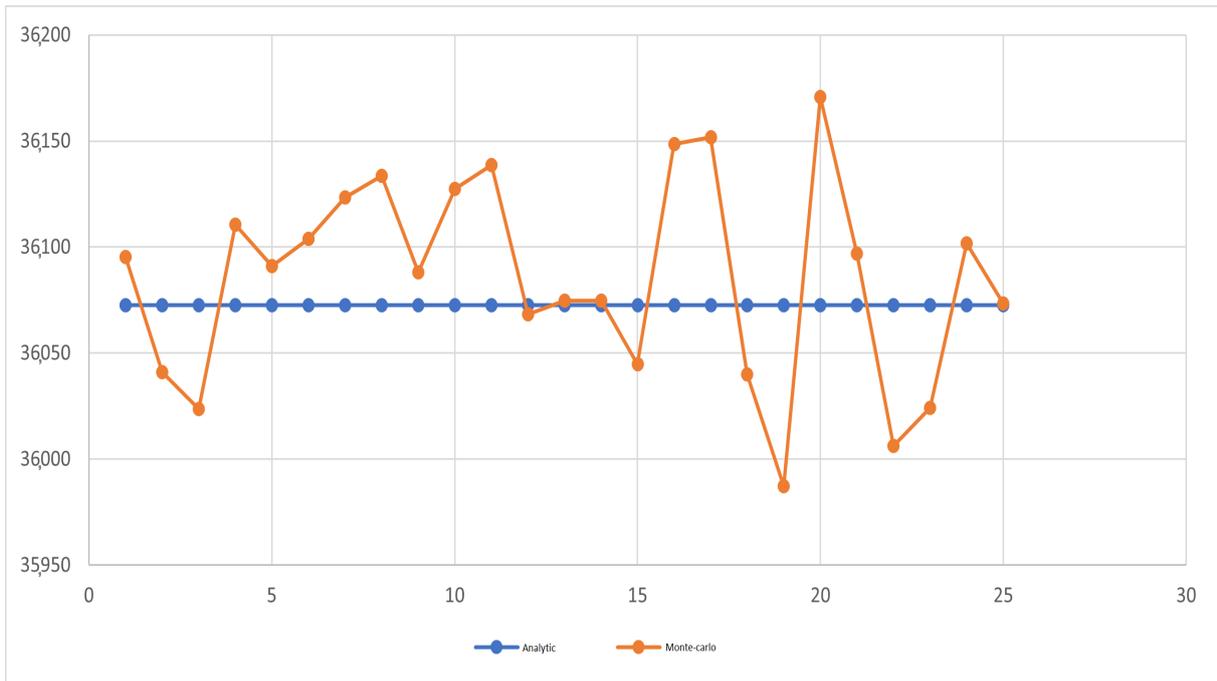


Figure 6. Comparison between analytical functions and Monte Carlo Case 1.

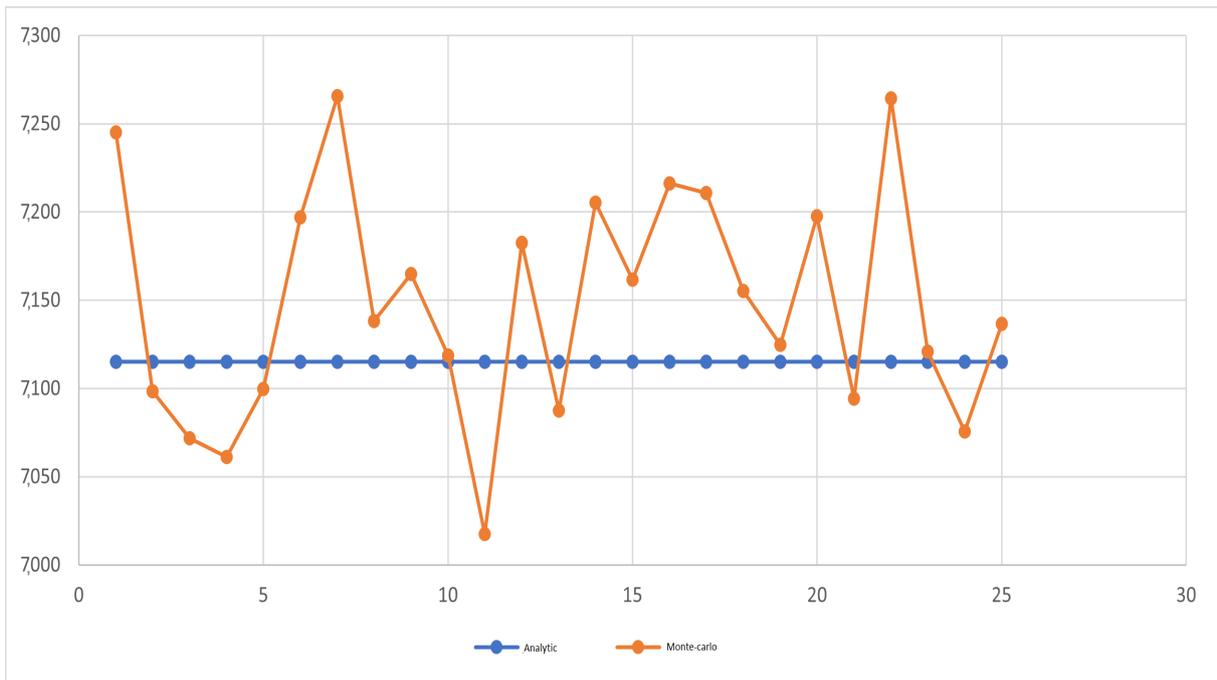


Figure 7. Comparison between analytical functions and Monte Carlo Case 2.

In general, the previous simulation cases show that the equations found present an optimal behavior, as long as it is ensured that the maximum energy to be dispatched $P_{e,\infty}$ is large enough to cover the demand, in this case, feeding Building 944. This is related to the uncertainty cost function for underestimation that depends on the value of $P_{e,\infty}$, as can be seen in Equation (12).

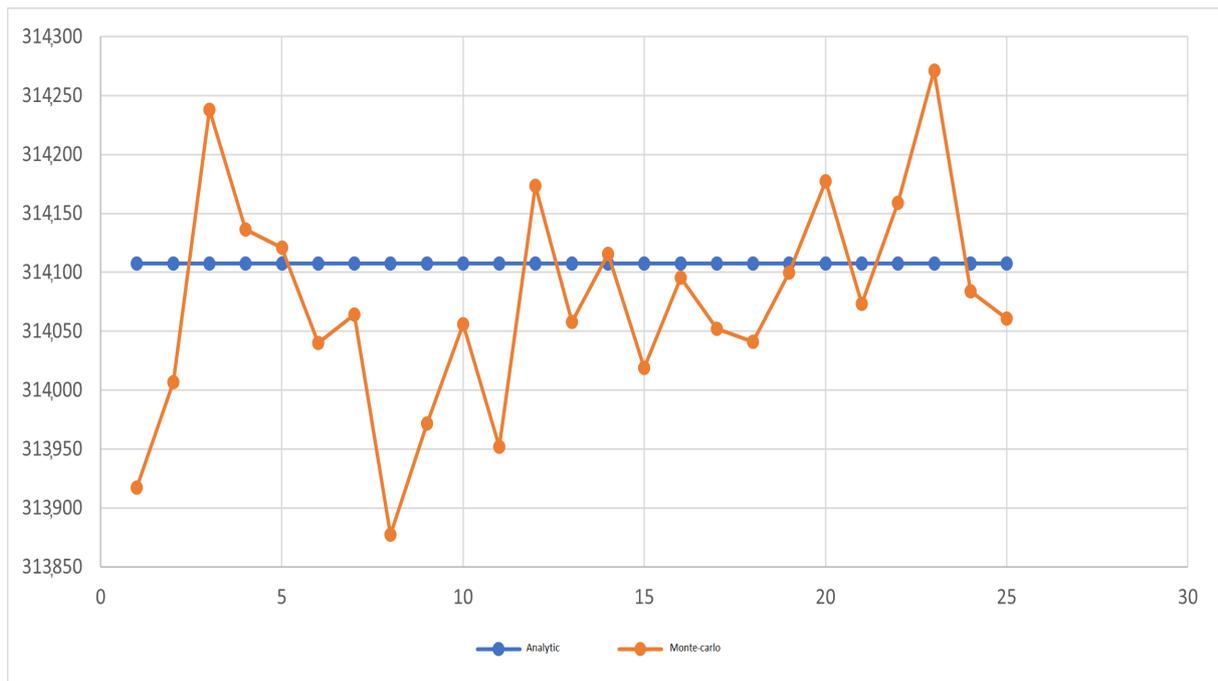


Figure 8. Comparison between analytical functions and Monte Carlo Case 3.

Table 5. Summary statistical values Cases 1–3.

Case	Average Error [%]	Mean	Variance	Std Dev.
1	0.071818	0.090601	0.005621	0.076522
2	0.564646	0.650946	0.354767	0.607905
3	0.016921	0.017571	0.000351	0.019133

8. Conclusions

It is possible to determine uncertainty cost functions for agents in the electricity sector with specific characteristics from a probability density function that describes their behavior and the analytical development of integrals of expected value. The probability density estimator represents a very useful tool for approximating probability density functions for future use in analytical procedures. In this paper, the relationship between electrical demand and room temperature is considered, which is susceptible to linearization based on the inclusion of exogenous variables such as cooling degree days (CDD) and heating degree days (HDD).

In this way, the phenomenon of autoregression has a considerable effect on the time series used in this study. Its inclusion in the models obtained from multiple linear regressions makes it possible to obtain models closer to the data recorded for the different cases.

Through the uncertainty cost functions found, it is possible to obtain the costs of an agent in the electricity sector, as long as statistical data on their behavior are available. Additionally, the functions found represent a good approximation to the behavior of the uncertainty costs of the analyzed controllable load, since they present errors of less than 1%, compared to the Monte-Carlo simulations.

There is an intrinsic dependence between the analytical functions found and the maximum value of demand that could be scheduled to the controllable load. The inclusion of “dummy” variables that represent conditions external to the relationship to be analyzed allows a model adjustment more faithful to the reality of the recorded data.

Author Contributions: Conceptualization, D.L., A.A.-S., and S.R.; methodology, D.L., S.R. and A.A.-S.; validation, D.L. and A.A.-S.; formal analysis, A.A.-S., S.R., and D.L.; investigation, D.L., A.A.-S. and S.R.; writing, review and editing, D.L., S.R. and A.A.-S.; funding acquisition and supervision, A.A.-S., and S.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by Universidad Nacional de Colombia, and by Khalifa University under Award No. kkjrc-2019-trans 2.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: This research was supported by Universidad Nacional de Colombia, and by Khalifa University under Award No. kkjrc-2019-trans 2.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

ERROR FUNCTION IDENTITIES

$$\int_a^b e^{-t^2} \cdot dt = \frac{\sqrt{\pi}}{2} (\operatorname{erf}(b) - \operatorname{erf}(a)) \quad (\text{A1})$$

$$\int_a^b e^{-(t-k)^2} \cdot dt = \frac{\sqrt{\pi}}{2} (\operatorname{erf}(b-k) - \operatorname{erf}(a-k)) \quad (\text{A2})$$

References

- Al-Sumaiti, A.S.; Hassan, M.A.; Rivera, S.; Salama, M.A.A.; El Moursi, M.; Alsumaiti, T. Stochastic PV Model for Power System Planning Applications. *IET Renew. Power Gener.* **2019**, *13*, 3168–3179. [[CrossRef](#)]
- The Nguyen, H.; Al-Sumaiti, A.S.; Vu, V.P.; Al-Durra, A.; Do, T.D. Optimal Power Tracking of PMSG Based Wind Energy Conversion Systems by Constrained Direct Control with Fast Convergence Rates. *Int. J. Electr. Power Energy Syst.* **2020**, *118*, 105807. [[CrossRef](#)]
- Al-Sumaiti, A.S.; Salama, M.M.A.; El Moursi, M.; Alsumaiti, T.; Marzband, M. Enabling Electricity Access: Revisiting Load Models for AC-Grid Operation—Part I. *IET Gener. Transm. Distrib.* **2019**, *13*, 2563–2571. [[CrossRef](#)]
- Al-Sumaiti, A.S.; Salama, M.M.A.; El Moursi, M.; Alsumaiti, T.; Marzband, M. Enabling electricity access: A comprehensive energy efficient approach mitigating climate/weather variability—Part II. *IET Gener. Transm. Distrib.* **2019**, *13*, 2572–2583. [[CrossRef](#)]
- Al-Sumaiti, A.S.; Ahmed, M.H.; Salama, M.M.A. Residential load management under stochastic weather condition in a developing Country. *Electr. Power Compon. Syst.* **2014**, *42*, 1452–1473. [[CrossRef](#)]
- Tookanlou, M.B.; Marzband, M.; Kyya, J.; Al-Sumaiti, A.; Al Hosani, K. Charging/Discharging strategy for electric vehicles based on bi-level programming approach: San Francisco case study. In Proceedings of the IEEE CPE-POWERENG 2020, Setubal, Portugal, 8–10 July 2020.
- Tookanlou, M.B.; Marzband, M.; Al-Sumaiti, A.; Mazza, A. Cost-benefit analysis for multiple agents considering an electric vehicle charging/discharging strategy and grid integration. In Proceedings of the 2020 IEEE 20th Mediterranean Electrotechnical Conference (MELECON), Palermo, Italy, 16–18 June 2020; pp. 19–24.
- Coria, G.E.; Sanchez, A.M.; Al-Sumaiti, A.S.; Rattá, G.A.; Rivera, S.R.; Romero, A.A. A Framework for Determining a Prediction-of-Use Tariff Aimed to Coordinate of Aggregators Plug-in Electric Vehicles. *Energies* **2019**, *12*, 4487. [[CrossRef](#)]
- Rodriguez, A.A.; Perdomo, L.; Al-Sumaiti, A.; Rivera, S. *Strategy for Charging Electric Vehicles in Office Buildings through Automatic Learning Algorithms*; Book Chapter for Upcoming Hardcover Edited Collection of Wiley-Scrivener Publisher Tentatively Entitled Smart Charging Solutions on Hybrid and Electric Vehicle; Sachan, S., Padmanaban, S., Eds.; Wiley: New York, NY, USA, 2020.
- Vargas, S.; Rodriguez, D.; Rivera, S. *Mathematical Formulation and Numerical Validation of Uncertainty Costs for Controllable Loads*; Artículo Revista Internacional de Métodos Numéricos para Cálculo y Diseño en Ingeniería; Scipedia: Sevilla, Spain, 2019.
- Surender, S.; Bijwe, P.; Abhyankar, A. Real-time economic dispatch considering renewable power generation variability and uncertainty over scheduling period. *IEEE Syst. J.* **2015**, *9*, 1440–1451. [[CrossRef](#)]
- Santos, F.; Arévalo, J.; Rivera, S. Uncertainty Cost Functions for Solar Photovoltaic Generation, Wind Energy, and Plug-In Electric Vehicles: Mathematical Expected Value and Verification by Monte Carlo Simulation. *Int. J. Power Energy Convers.* **2019**, in press.
- Molina, F.; Pérez, S.; Rivera, S. Uncertainty Cost Function Formulation in Small Hydropower Plants Inside a Microgrid. *Ing. USBMed* **2017**, *8*, 126.

14. Mehri, R.; Kalantar, M. Multi-objective Scheduling of Electric Vehicles Considering Wind and Demand Uncertainties. In Proceedings of the 2015 Smart Grid Conference (SGC), Tehran, Iran, 22–23 December 2015.
15. Bernal, J.A.; Neira, J.E.; Rivera, S. Mathematical Uncertainty Cost Functions for Controllable Photo-Voltaic Generators considering Uniform Distributions. *WSEAS Trans. Math.* **2019**, *18*, 137–142.
16. Pardo, A.; Meneu, V.; Valor, E. *Temperature and Seasonality Influences on Spanish Electricity Load*; Departamento de Economía Financiera y Matemática, Facultad de Economía, Universidad de Valencia: Valencia, Spain, 2002.
17. Moral-Carcedo, J.; Vicéns-Otero, J. *Modelling the Non-Linear Response of Spanish Electricity Demand to Temperature Variations*; Departamento Análisis Económico, Universidad Autónoma de Madrid: Madrid, Spain, 2005.
18. Henley, A.; Peirson, J. *Non-Linearities in Electricity Demand and Temperature: Parametric Versus Non Parametric Methods*; Oxford Bulletin of Economics and Statistics; Wiley: London, UK, 1997.
19. *Kernel Distribution*; MathWorks—Documentation; Matlab: Cambridge, MA, USA, 2021.
20. Weglarczyk, S. *Kernel Density Estimation and Its Application*; Cracow University of Technology, Institute of Water Management and Water Engineering: Warszawska, Poland, 2018.
21. Zhao, J.H.; Wen, F.; Dong, Z.Y.; Xue, Y.; Wong, K.P. Optimal dispatch of electricvehicles and wind power using enhanced particle swarm optimization. *IEEE Trans. Ind. Inform.* **2012**, *8*, 889–899. [[CrossRef](#)]
22. Chang, T. Investigation on frequency distribution of global radiation using different probability density functions. *Int. J. Appl. Sci. Eng.* **2010**, *8*, 99–107.
23. Montanari, R. Criteria for the economic planning of a low power hydroelectric plant. *Renew. Energy* **2003**, *28*, 2129–2145. [[CrossRef](#)]
24. Cabus, P. River flow prediction through rainfall–runoff modelling with a probability-distributed model (PDM) in Flanders, Belgium. *Agric. Water Manag.* **2008**, *95*, 859–868. [[CrossRef](#)]
25. Medina-Santiago, A.; Pano, A.; Gomez, J.; Jesus-Magaña, J.; Valdez-Ramos, M.; Sosa-Silva, E.; Falcon-Perez, F. Adaptive Model IoT for Monitoring in Data Centers. *IEEE Access* **2020**, *8*, 5622–5634. [[CrossRef](#)]
26. Chen, Y.C. *Lecture 6: Density Estimation: Histogram and Kernel Density Estimator*; STAT 425: Introduction to Nonparametric Statistics; University of Washington: Washington, DC, USA, 2018.
27. *Model Data Using the Distribution Fitter App*; MathWorks—Documentation; Matlab: Cambridge, MA, USA, 2021.
28. Cancelo, J.R.; Espasa, A. *Modelling and Forecasting Daily Series of Electricity Demand*; Departamento de Estadística, Investigaciones Económicas, Universidad Carlos III: Madrid, Spain, 1996.
29. Mujere, N. Flood Frequency Analysis Using the Gumbel Distribution. *Int. J. Comput. Sci. Eng. (IJCSE)* **2011**, *3*, 2774–2778.
30. Martínez, C.A.; Rivera, S. *Quadratic Modelling of Uncertainty Costs for Renewable Generation and Its Application on Economic Dispatch*; Revista Del Programa De Matemáticas, Facultad de Ciencias Básicas, Universidad del Atlántico: Barranquilla, Colombia, 2018; Volume 3, pp. 36–61.
31. Torres, J.; Rivera, S. Despacho de Energía óptimo en Múltiples Periodos Considerando la Incertidumbre de la Generación a Partir de Fuentes Renovables en un Modelo Reducido del Sistema de Potencia Colombiano. *Av. Investig. Ing.* **2018**, *15*, 48–58.
32. Ortiz, M.; Londoño, N. *Análisis del Mercado de Contratos del MEM de Colombia Durante el último Fenómeno de El Niño (2015–2016)*; Escuela de Economía y Finanzas, Departamento de Economía, Universidad EAFIT: Medellín, Colombia, 2017.
33. Abouabdellah, A.; Bannari, R.; El Kafazi, I. Modeling and forecasting energy demand. In Proceedings of the 2016 International Renewable and Sustainable Energy Conference (IRSEC), Marrakech, Morocco, 14–17 November 2017.