energies

MDPI

Article

Comparative Assessment of Uncertainty Representation Methods for Bilevel Stochastic Wind Power Investment

Dionysios Liaros, Athina P. Georgilaki and Pavlos S. Georgilakis *

School of Electrical and Computer Engineering, National Technical University of Athens (NTUA), 15780 Athens, Greece; dionisisliaros@gmail.com (D.L.); athinageorgilaki03@gmail.com (A.P.G.)
* Correspondence: pgeorg@power.ece.ntua.gr; Tel.: +30-210-772-4378

Abstract

Wind investors in network-constrained electricity markets must make siting and capacity decisions under uncertain demand and wind availability, while market prices and congestion respond to these decisions. This paper examines how alternative representations of these uncertainties affect optimal wind investment and system operation. A stochastic bilevel model is used, with upper-level discrete investment decisions and lower-level lossless DC market clearing. The model is reformulated as a Mathematical Program with Equilibrium Constraints (MPEC) and subsequently as a Mixed-Integer Linear Programming (MILP) problem. Two published scenario sets based on 2007 Iberian Peninsula demand and wind data are compared: 36 duration-curve operating conditions formed from four load blocks and combinations of three demand and three wind levels, and 48 correlated K-means centroids representing joint observations from three demand zones and two wind zones. A 3-bus verification case and eight IEEE 24-bus RTS cases examine investment budgets and spatial/network variations. In the directly comparable budget-constrained cases, 1000 MW is installed under either representation, but its allocation among four candidate buses changes from 300/300/200/200 MW to 100/300/300/300 MW, while the reported expected annual profit differs by 10.20%. Across the four reported 24-bus comparisons, weighted-average system locational marginal prices differ by 3.60–3.71 EUR/MWh. The scenario representations can therefore lead to different investment and network outcomes, although the comparisons involving externally imposed wind-availability factors must also be interpreted in light of their differing spatial assumptions.

Keywords: wind power investment; bilevel programming; stochastic programming; K-means clustering; mathematical program with equilibrium constraints (MPEC); mixed-integer linear programming (MILP)

Academic Editor: Henrik Pedersen

Received: 30 September 2026
Revised: 26 October 2026
Accepted: 3 November 2026
Published: 4 November 2026

Copyright: © 2026 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.

1. Introduction

As energy systems move toward climate neutrality, the penetration of renewable energy sources (RES) is increasing, driven by European decarbonization targets [1]. Achieving these climate-neutrality goals requires substantial wind and solar capacity and changes the structure and operating conditions of power systems [1,2]. Because wind and solar production depends on atmospheric conditions, greater RES penetration increases variability and can affect system adequacy as well as operational flexibility. Nicolosi and Tina [3], for example, show why correlations among meteorological varia-

bles matter when constructing adequacy-oriented generation and demand profiles. Investment in new wind capacity must therefore be evaluated within the broader development of the power system. Recent studies examine renewable-energy planning, resource siting and sizing, and network-constrained operation [4–6].

In liberalized electricity markets, investment in new generation capacity is increasingly driven by decentralized decisions of profit-oriented investors rather than by centralized planning [7]. For wind power projects, an investor should decide not only whether to build new capacity, but also where and how much capacity to install within the transmission network. The value of an investment depends on market revenues, which are determined by market-clearing outcomes and by the locational marginal price (LMP) at the connection bus. Local wind conditions determine the amount of wind energy that can be produced at each site, while transmission limits can restrict its delivery and alter local market conditions. Consequently, wind power investments should be evaluated using a framework that explicitly captures the interaction among the investor, the electricity market, and the transmission network [7].

A further challenge is the uncertainty in the operating conditions used to evaluate long-term wind power investments. In the present problem, the principal sources of uncertainty are electricity demand and wind power availability, both of which affect market outcomes and investment profitability [8,9]. These quantities should not automatically be treated as statistically independent, because the economic value of wind generation depends on how wind availability coincides with electricity demand. Ignoring this dependence may discard information contained in historical operating data and can therefore affect the resulting investment decision [8]. Accurate representation of both uncertain quantities and of their joint behavior is consequently important in a stochastic wind power investment framework.

Long-term wind-investment evaluations also face possible changes in future wind resources under climate change; such changes may alter expected energy production and investment returns [10]. The present study is a methodological comparison for a single target year using fixed historical scenario sets and does not project climate-driven changes in wind availability or future resource yield. The historical demand and wind data underlying both scenario sets are those of the Iberian Peninsula for the year 2007 [7,8].

Scenario-based methods provide a computationally tractable means of incorporating load and wind uncertainty into long-term investment models [9]. However, different scenario-construction methods retain different characteristics of the original data. Two relevant approaches are scenarios derived from load and wind duration curves and correlated scenarios obtained through K-means clustering [7,8]. The duration-curve approach introduces a block-level relationship between load and wind, but within each load block the selected demand and wind levels are combined independently; therefore, the full joint pattern of the underlying hourly observations is not retained. By contrast, K-means clustering is applied to multidimensional historical observations and can preserve a larger part of the joint and spatial dependence between demand and wind availability [8]. Because the two approaches supply different operating states to the same investment model, the choice of uncertainty representation can influence both investment and market outcomes.

The resulting methodological question is therefore not whether uncertainty should be represented, but whether different representations of the same uncertain quantities lead to materially different planning conclusions when every other component of the optimization problem is kept unchanged. A controlled comparison is particularly important in a network-constrained market setting, because changes in the retained load–wind dependence can affect nodal prices, congestion, curtailment, and ultimately

the profitability of candidate locations for wind power investments. This study addresses that question by comparing the two scenario representations within one common stochastic bilevel investment framework.

The relevant methodological foundations are provided by Baringo and Conejo [7], who developed the market-based wind-investment model, and by Baringo and Conejo [8], who investigated correlated demand–wind scenarios for investment decisions. Building on this literature, the contribution of the present study is a structured investigation of nine verification and application cases, with consistent reporting of discrete investment outcomes, expected profitability, LMPs, transmission congestion, and wind curtailment. The case design additionally clarifies the roles of investment-budget restrictions, spatial wind representation, and network topology in interpreting the results. Studies of representative operating periods [11–14] further highlight the importance of evaluating scenario aggregation through its effect on downstream planning decisions.

This article evaluates duration-curve and correlated K-means representations of electricity demand and wind availability within a common stochastic bilevel wind-investment framework. The investment and market-clearing formulations are shared by all cases, and the baseline and reduced-budget comparisons use matching network and investment parameters. The spatially differentiated cases additionally examine the effect of representing regional wind conditions through the respective scenario inputs. The three main contributions of this article are:

- A structured, reproducible comparison of two uncertainty-representation methods within a common stochastic bilevel wind-investment formulation, with directly matched baseline and reduced-budget network cases.
- A quantitative evaluation of how alternative joint and spatial load–wind representations affect discrete wind siting and reported profitability, LMPs, congestion, and wind power curtailment, explicitly distinguishing results attributable to scenario representation from comparisons that also vary spatial wind assumptions.
- A numerical assessment across one 3-bus verification and eight IEEE 24-bus cases, including constrained versus technically saturated investment budgets and a reported line-removal configuration; practical interpretation is provided for investors and system operators.

The remainder of this paper is organized as follows. Section 2 presents the stochastic bilevel wind power investment model and its equivalent MILP reformulation. Section 3 describes the two uncertainty-representation approaches and the controlled comparison methodology. Section 4 introduces the test systems, input data, and case-study configurations. Section 5 presents and discusses the numerical results, with emphasis on the differences attributable to the two scenario representations. Section 6 summarizes the main conclusions and their implications for wind power investors and system operators.

2. Stochastic Bilevel Wind Power Investment Model

2.1. Problem Description and Main Assumptions

This work formulates the wind power investment problem within an existing transmission network and an organized electricity market. The stochastic bilevel investment structure follows the framework introduced in [7] and serves as the common optimization model for the case studies. The upper-level problem represents a wind power investor, who selects the wind power capacity to install at candidate buses in order to maximize expected annual profit. The lower-level problem represents market clearing for each operating condition and determines conventional generation dispatch, transmission power flows, and LMPs. The comparison varies demand and wind scenario inputs while retaining the stated network and investment configuration for each case.

https://doi.org/10.3390/en19215297

The model is static and represents a single target year. Wind investment is restricted to predefined discrete capacity levels at each candidate bus and is subject to a maximum investment budget. Conventional generators are assumed to behave competitively by offering their marginal production costs, whereas wind generation is offered at zero price and is remunerated at the LMP of its connection bus. The transmission system is represented by a lossless DC power-flow model, so that network constraints and congestion can affect market-clearing outcomes and nodal prices. Demand is treated as fixed and inelastic in each operating state.

The upper-level investment variables are the binary capacity-level selections and the resulting installed wind power capacities (MW); the lower-level primal variables include conventional dispatch, dispatched wind and line flows (MW), and bus voltage angles (rad). The lower-level dual variables include nodal-balance multipliers/LMPs (EUR/MWh) and shadow prices associated with operating constraints. Power and energy are distinguished: power dispatch is in MW, its annual energy contribution is multiplied by scenario duration in hours (energy is in MWh), and prices are in EUR/MWh. Monetary figures are in EUR, with annualized investment costs in EUR/year. Appendix A lists parameter- and variable-specific units.

2.2. Bilevel Mathematical Formulation

The stochastic bilevel formulation consists of an upper-level wind-investment problem and a lower-level market-clearing problem for every operating condition. The interaction between the two levels allows the investment decision to account explicitly for the effect of network conditions and market outcomes on the profitability of wind generation.

At the upper level, the investor maximizes the expected annual profit, defined as the expected revenue obtained from selling wind energy in the electricity market minus the annualized investment cost:

$$max_{\Delta^{UL},\Delta^{LL}_{t,\omega},\forall t,\omega} \sum_{t\in\Omega^T} N_t^h \sum_{\omega\in\Psi_t^\omega} \alpha_t(\omega)\left[\sum_{n\in\Omega^N} \beta_{n,t}(\omega)P_{n,t}(\omega)\right] - \sum_{n\in\Omega^N} \tilde{c}_n^{inv} X_n \quad (1)$$

Here, X_n denotes the wind capacity installed at bus n, $P_{n,t}(\omega)$ is the wind power injected into the system under operating condition (t,ω), and $\beta_{n,t}(\omega)$ is the corresponding selling price. The contribution of each operating condition to the annual revenue is determined by its temporal duration N_t^h and scenario weight $\alpha_t(\omega)$. The indices t and ω denote the load block and scenario, respectively.

The actual wind power injected into the system cannot exceed the wind power available under the corresponding scenario and installed capacity:

$$P_{n,t}(\omega) \le k_{n,t}(\omega)X_n, \quad \forall\, n, \quad \forall\, t, \quad \forall\, \omega \quad (2)$$

where $k_{n,t}(\omega)$ represents the scenario-dependent wind availability factor, with $0 \le k_{n,t}(\omega) \le 1$. Therefore, the wind power dispatched in each operating condition ranges from zero to the installed wind power capacity, depending on wind availability.

Wind investment is represented through predefined discrete capacity levels at each candidate bus. The installed capacity and the associated investment-budget restriction are expressed as:

$$X_n = \sum_{l \in \Psi_n^L} u_n^l q_n^l, \quad \forall\, n \tag{3a}$$

$$\sum_{l \in \Psi_n^L} u_n^l = 1\,, \quad \forall n\,;\; u_n^l = \{0,1\}, \quad \forall\, n, \quad \forall\, l \tag{3b}$$

$$\sum_{n \in \Omega^N} c_n^{inv}\, X_n \;\leq\; c_{max}^{inv} \tag{3c}$$

Here, q_n^l denotes the capacity associated with investment level l, while u_n^l is a binary variable indicating whether that level is selected. Equation (3c) ensures that the total investment cost does not exceed the available budget.

The selling price of wind generation is determined endogenously through the lower-level market-clearing problem and is equal to the LMP of the corresponding bus:

$$\beta_{n,t}(\omega) = \lambda_{n,t}(\omega), \quad \forall n, t, \omega \tag{4}$$

For each operating condition (t, ω), the lower-level problem determines the market-clearing solution by minimizing the production cost of conventional generation:

$$min_{\Delta_{t\omega}^{LL}, \forall t, \omega} \sum_{i \in \Omega^G} \sum_{b \in \Omega_i} c_{ib}\; g_{ib,t}\;(\omega) \tag{5}$$

Since electricity demand is assumed to be fixed and inelastic within each operating condition, minimizing conventional generation cost is equivalent to maximizing social welfare for that state.

Market clearing requires the nodal power balance to be satisfied at every bus:

$$\begin{aligned} \sum_{i \in \Psi_n^G} \sum_{b \in \Omega_i} g_{ib,t}(\omega) - \sum_{k|o(k)=n} f_{k,t}(\omega) + \sum_{k|r(k)=n} f_{k,t}(\omega) + P_{n,t}(\omega) \\ = \sum_{j \in \Psi_n^D} d_{j,t}(\omega) : \lambda_{n,t}(\omega), \quad \forall\, n, t, \omega \end{aligned} \tag{6}$$

The dual variable associated with (6), $\lambda_{n,t}(\omega)$, represents the LMP at bus n. Consequently, the market price received by the wind investor in (1) depends endogenously on generation dispatch, demand, and network conditions.

The lower-level problem further includes the DC power-flow equations, transmission-capacity limits, conventional-generation limits, voltage-angle bounds, and the reference-bus condition:

$$f_{k,t}(\omega) = B_k\big(\,\delta_{o(k),t}(\omega) - \delta_{r(k),t}(\omega)\,\big) : \phi_{k,t}(\omega), \quad \forall\, k, t, \omega \tag{7a}$$

$$-f_k^{max} \;\leq\; f_{k,t}(\omega) \;\leq\; f_k^{max} : \phi_{k,t}^{min}(\omega)\,, \phi_{k,t}^{max}(\omega)\,, \quad \forall\, k, t, \omega \tag{7b}$$

$$0 \;\leq\; g_{ib,t}(\omega) \leq\; g_{ib}^{max} : \varphi_{ib,t}^{min}(\omega), \;\; \varphi_{ib,t}^{max}(\omega), \quad \forall\, i, b, t, \omega \tag{7c}$$

$$-\pi \leq \delta_{n,t}(\omega) \leq \pi : \xi_{n,t}^{\min}(\omega), \xi_{n,t}^{\max}(\omega), \quad \forall n \neq n^{\mathrm{ref}}, \forall t, \omega \tag{7d}$$

$$\delta_{n,t}(\omega) = 0 : \chi_{n,t}(\omega), \quad n = n^{\mathrm{ref}}, \forall t, \omega \tag{7e}$$

Equation (7a) represents transmission flows through the lossless DC power-flow approximation, while (7b) allows transmission congestion to be explicitly represented. Equation (7c) restricts conventional generation within its available capacity, whereas (7d) and (7e) define the voltage-angle bounds and establish a reference angle for the DC power-flow solution.

Equations (1)–(7) constitute the common bilevel mathematical structure employed throughout the analysis. Importantly, this structure remains unchanged under the two uncertainty-representation approaches investigated in this work. The quantities that vary between them are the scenario-dependent demand $d_{j,t}(\omega)$, wind availability $k_{n,t}(\omega)$, and the temporal weighting assigned to each operating condition. Section 3 presents the construction of these alternative input representations.

Settlement is performed ex post for each represented operating condition: the realized dispatched wind quantity is remunerated at that condition's endogenous nodal LMP, and these revenues are subsequently aggregated using the condition's annual-hour weight. There is no separate settlement at a single ex ante expected LMP. The DC power-flow approximation is lossless; consequently, neither marginal transmission losses nor loss factors are included. Available wind that is not dispatched is curtailed without a deemed-generation payment or explicit curtailment penalty. Such curtailment reduces the investor's market revenues through lower dispatched MWh; it can therefore affect site profitability endogenously.

2.3. Single-Level MILP Reformulation

The bilevel formulation in Section 2.2 is not directly compatible with a conventional single-level linear optimization solver because the lower-level market-clearing problem appears as an optimization problem embedded within the upper-level problem [15]. Since the lower-level problem is continuous and linear, it is convex, and its optimal solution can be characterized by the Karush–Kuhn–Tucker (KKT) conditions. Replacing the lower-level problem by these optimality conditions therefore yields an equivalent single-level Mathematical Program with Equilibrium Constraints (MPEC) [15].

The KKT system consists of primal feasibility, dual feasibility, stationarity, and complementarity conditions. Primal feasibility is already represented by the market-clearing constraints in (6) and (7). The dual variables associated with inequality constraints are non-negative, whereas those associated with equality constraints are unrestricted in sign. The stationarity conditions with respect to conventional generation, transmission flows, and voltage angles are given by:

$$c_{ib} - \lambda_{n(i),t}(\omega) + \varphi^{max}_{ib,t}(\omega) - \varphi^{min}_{ib,t}(\omega) = 0, \quad \forall\, i, b, t, \omega \tag{8a}$$

$$\lambda_{o(k),t}(\omega) - \lambda_{r(k),t}(\omega) - \phi_{k,t}(\omega) + \phi^{max}_{k,t}(\omega) - \phi^{min}_{k,t}(\omega) = 0, \quad \forall\, k, t, \omega \tag{8b}$$

$$-\sum_{k|o(k)=n} B_k\, \phi_{k,t}(\omega) + \sum_{k|r(k)=n} B_k\, \phi_{k,t}(\omega) + \xi^{\max}_{n,t}(\omega) - \xi^{\min}_{n,t}(\omega) = 0, \quad \forall n \neq n^{\mathrm{ref}}, \forall t, \omega \tag{8c}$$

$$-\sum_{k|o(k)=n} B_k\, \phi_{k,t}(\omega) + \sum_{k|r(k)=n} B_k\, \phi_{k,t}(\omega) - \chi_{n,t}(\omega) = 0, \quad n = n^{\mathrm{ref}}, \forall t, \omega \tag{8d}$$

The complementarity conditions link the dual variables of the inequality constraints with the corresponding primal slacks. In general, each pair can be expressed as:

$$0 \leq a \perp b \geq 0 \tag{9}$$

where a denotes the corresponding non-negative dual variable and b the slack of the associated inequality. These conditions are imposed on the transmission-flow limits, conventional-generation limits, and voltage-angle bounds. Together with primal and dual feasibility and the stationarity conditions in (8), they fully replace the lower-level optimization problem.

Although the KKT reformulation eliminates the bilevel structure, the resulting MPEC remains nonlinear because it contains two classes of bilinear terms. First, the upper-level objective contains the product of wind generation and the corresponding LMP.

Second, each complementarity condition in (9) imposes a zero product between a primal slack and its associated dual variable. These two classes of bilinear terms are treated separately using strong duality and a big-M reformulation, respectively.

Since the lower-level market-clearing problem is linear and convex, strong duality holds at its optimum. Using the equality between the primal and dual lower-level objective functions, and considering that $\beta_{n,t}(\omega) = \lambda_{n,t}(\omega)$, the bilinear wind-revenue term can be equivalently written as:

$$\begin{aligned}\sum_{n\in\Omega^N} \beta_{n,t}(\omega) P_{n,t}(\omega) = \sum_{n\in\Omega^N} \lambda_{n,t}(\omega) \sum_{j\in\Psi_n^D} d_{j,t}(\omega) \\ - \sum_{k\in\Omega^K} \left(\phi_{k,t}^{max}(\omega) + \phi_{k,t}^{min}(\omega) \right) f_k^{max} - \sum_{i\in\Omega^G} \sum_{b\in\Omega_i} \varphi_{ib,t}^{max}(\omega) g_{ib}^{max} \\ - \sum_{n\in\Omega^N\setminus\{n^{\mathrm{ref}}\}} \left(\xi_{n,t}^{\max}(\omega) + \xi_{n,t}^{\min}(\omega) \right) \pi - \sum_{i\in\Omega^G} \sum_{b\in\Omega_i} c_{ib}\, g_{ib,t}(\omega)\end{aligned} \tag{10}$$

Equation (10) replaces the original product of two decision variables by an expression that is linear in the variables of the reformulated problem. Substitution of (10) into the upper-level objective therefore removes the bilinear revenue term.

The remaining bilinearities arise from the complementarity conditions. For a generic complementarity pair $0 \le a \perp b \ge 0$, the Fortuny-Amat and McCarl reformulation introduces a binary variable u and a sufficiently large positive constant M [16], leading to:

$$a \le Mu \tag{11a}$$

$$b \le M(1-u) \tag{11b}$$

$$u \in \{0,1\} \tag{11c}$$

This formulation ensures that at most one member of each complementarity pair can be strictly positive. The same procedure is applied to the dual/slack pairs associated with transmission-flow limits, conventional-generation limits, and voltage-angle bounds. Separate big-M constants may be used for the different groups of constraints.

The big-M constants should be sufficiently large to preserve all feasible solutions, but not unnecessarily large, in order to avoid numerical difficulties and weak continuous relaxations.

For the nine cases, the implemented constants are M_ϕ = 1000 for transmission-flow complementarity, M_φ = 1000 for conventional-generation complementarity, and M_ξ = 10 for voltage-angle complementarity. These values satisfy the corresponding physical primal-slack bounds: $2f_k^{\max} \le 1000$ MW for line-flow limits, $g_{ib}^{\max} \le 660$ MW for generation blocks, and 2π = 6.2832 rad < 10 rad for angle bounds. The same constants are used in the associated dual constraints. The independently obtained Gurobi and CPLEX solutions agree in investment decisions and objective values across all nine cases (Section 5.5), providing a numerical consistency check for the implemented formulation. The primal-slack bounds serve as direct physical checks, whereas the dual-multiplier bounds are implementation parameters rather than analytical consequences of those physical limits.

After the strong-duality substitution and the explicit linearization of every complementarity pair, the stochastic bilevel model becomes a Mixed-Integer Linear Programming (MILP) problem. For completeness, the specific big-M constraints used in the final MILP are stated below rather than being represented only through the generic form in (11).

For all applicable indices, the dual variables associated with the upper and lower transmission-flow limits, conventional-generation limits, and voltage-angle limits are linearized as follows:

$$\phi_{k,t}^{max}(\omega) \le M_{\phi} u_{k,t}^{max}(\omega) \tag{12a}$$

$$\phi_{k,t}^{min}(\omega) \le M_{\phi} u_{k,t}^{min}(\omega) \tag{12b}$$

$$\varphi_{ib,t}^{max}(\omega) \le M_{\varphi} u_{ib,t}^{max}(\omega) \tag{12c}$$

$$\varphi_{ib,t}^{min}(\omega) \le M_{\varphi} u_{ib,t}^{min}(\omega) \tag{12d}$$

$$\xi_{n,t}^{max}(\omega) \le M_{\xi} u_{n,t}^{max}(\omega) \tag{12e}$$

$$\xi_{n,t}^{min}(\omega) \le M_{\xi} u_{n,t}^{min}(\omega) \tag{12f}$$

For the same index domains, the corresponding primal-slack constraints are:

$$f_k^{max} - f_{k,t}(\omega) \le M_{\phi} \left(1 - u_{k,t}^{max}(\omega)\right) \tag{13a}$$

$$f_{k,t}(\omega) + f_k^{max} \le M_{\phi} \left(1 - u_{k,t}^{min}(\omega)\right) \tag{13b}$$

$$g_{ib}^{max} - g_{ib,t}(\omega) \le M_{\varphi} \left(1 - u_{ib,t}^{max}(\omega)\right) \tag{13c}$$

$$g_{ib,t}(\omega) \le M_{\varphi} \left(1 - u_{ib,t}^{min}(\omega)\right) \tag{13d}$$

$$\pi - \delta_{n,t}(\omega) \le M_{\xi} \left(1 - u_{n,t}^{max}(\omega)\right) \tag{13e}$$

$$\pi + \delta_{n,t}(\omega) \le M_{\xi} \left(1 - u_{n,t}^{min}(\omega)\right) \tag{13f}$$

The auxiliary binary variables and the dual-feasibility domains required by the final MILP are:

$$u_{k,t}^{max}(\omega), u_{k,t}^{min}(\omega) \in \{0,1\} \tag{14a}$$

$$u_{ib,t}^{max}(\omega), u_{ib,t}^{min}(\omega) \in \{0,1\} \tag{14b}$$

$$u_{n,t}^{max}(\omega), u_{n,t}^{min}(\omega) \in \{0,1\} \tag{14c}$$

$$\phi_{k,t}^{max}(\omega), \phi_{k,t}^{min}(\omega) \ge 0 \tag{14d}$$

$$\varphi_{ib,t}^{max}(\omega), \varphi_{ib,t}^{min}(\omega) \ge 0 \tag{14e}$$

$$\xi_{n,t}^{max}(\omega), \xi_{n,t}^{min}(\omega) \ge 0 \tag{14f}$$

$$\lambda_{n,t}(\omega), \phi_{k,t}(\omega), \chi_{n,t}(\omega) \in \mathbb{R} \tag{14g}$$

Substituting the linear revenue identity in (10) into the upper-level objective (1) and combining it with the investment and wind-availability constraints in (2) and (4), the primal market-clearing constraints in (6) and (7), the stationarity conditions in (8), and the explicit linearization and domain constraints in (12)–(14) yields the final MILP formulation. Equations (5) and (9)–(11) are retained because they define the lower-level problem and document the reformulation steps, but they are not added as separate constraints to the final MILP.

The reformulation is identical for the two uncertainty representations examined in this study. Consequently, neither the investment model nor the market-clearing mecha-

nism is modified between the two approaches; the scenario-dependent demand, wind-availability, and temporal-weight parameters, together with the number of representative operating conditions, are determined by the adopted scenario set. This common optimization structure supports a controlled comparison of the two uncertainty representations while keeping the underlying investment and market-clearing model unchanged.

3. Uncertainty Representation and Scenario Construction

The stochastic investment model is evaluated over a finite set of representative operating conditions rather than over the complete hourly chronology of a year. This reduction is necessary to keep the bilevel/MILP formulation computationally manageable, but it also determines which combinations of electricity demand and wind availability are presented to the optimization problem. The two representations examined in this work differ only in the construction and weighting of these operating conditions; the mathematical model described in Section 2 remains unchanged. An overview of the two scenario-construction pathways and their common integration into the optimization framework is provided in Figure 1.

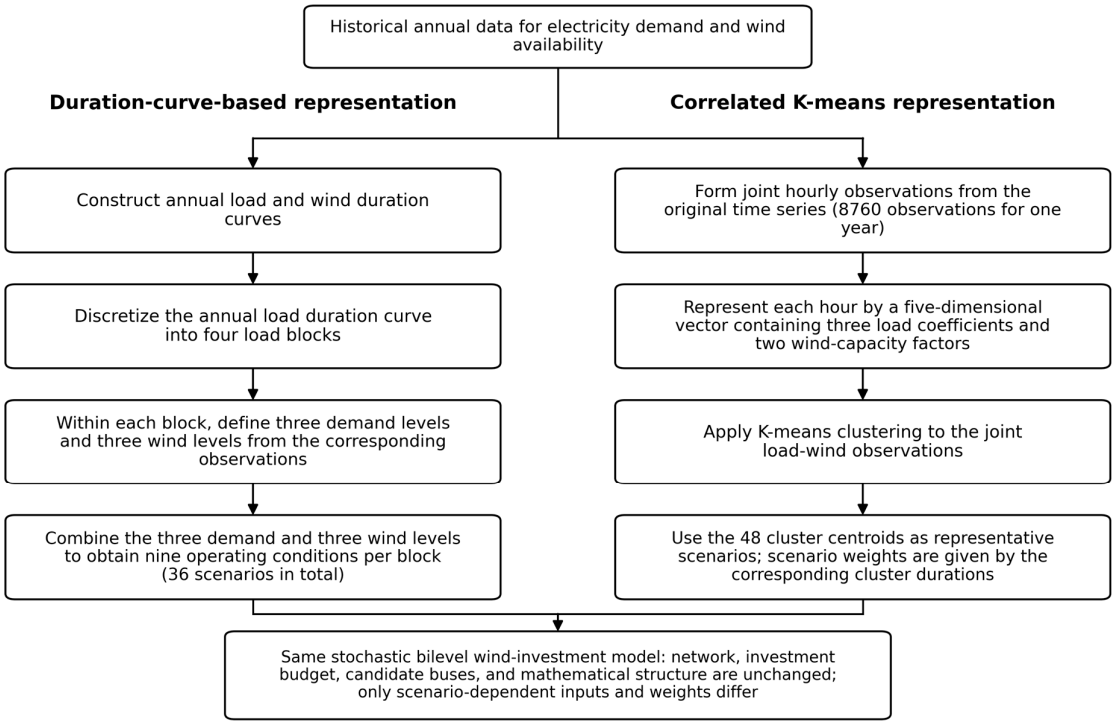


Figure 1. Schematic representation of the two uncertainty-representation approaches considered in this work and their integration into the common stochastic bilevel wind-investment model.

3.1. Duration-Curve-Based Scenario Representation

The first representation follows the scenario construction used in the reference wind-investment model [7]. Load duration curves provide a compact description of annual demand by ordering hourly load levels according to the number of hours for which they are equaled or exceeded [17]. In the adopted representation, the annual load duration curve is divided into four load blocks. Within each block, demand uncertainty is represented by three levels: the mean demand of the block and two additional levels set 5% below and 5% above that mean. This discretization retains the broad variation be-

tween high- and low-demand conditions while replacing the full hourly series by a limited number of representative values.

The adopted duration-curve inputs are the published 2007 Iberian electricity-market load and wind data used in [7]. The four load-block durations are 1200, 3600, 2400, and 1560 h, summing to 8760 h. With nine equally weighted states within each block, the corresponding hours per state are 133.33, 400.00, 266.67, and 173.33 h. These are representative combinations, not reconstructed original hourly pairs.

Wind availability is treated conditionally on the load blocks. For each block, the wind observations corresponding to the hours assigned to that load range are used to obtain three representative wind-availability levels. The three demand levels are then combined with the three wind levels, producing nine equally weighted scenarios per load block and 36 operating conditions in total [7]. Wind uncertainty is particularly important in market-based investment analysis because wind availability varies with meteorological conditions and is subject to substantial forecast uncertainty [18,19].

The duration-curve construction retains part of the dependence between demand and wind because the wind data used in each block originate from the hours associated with that demand range. However, the exact hourly load–wind pairs are not retained: once the representative levels have been obtained, the three demand levels and three wind levels are combined independently within the block. Spatial dependence is also represented only indirectly. This limitation is relevant in a transmission-constrained system, where geographically differentiated renewable availability and demand can modify line loading, nodal prices, and the value of candidate investment locations [7,8,20].

3.2. Correlated K-Means Scenario Representation

The second representation uses correlated scenarios obtained through K-means clustering, following the scenario set reported in [8]. Clustering partitions multidimensional observations into groups whose members are similar according to a selected distance measure [21]. K-means, originally introduced by MacQueen [22], is a centroid-based method that seeks a partition with low within-cluster dispersion. Related clustering approaches have been used in power-system studies to reduce large wind datasets and to construct representative operating conditions while retaining relevant statistical information [23,24].

For the correlated representation adopted here, the 8760 hourly observations of the year are treated as multidimensional operating states rather than as separate load and wind series. Each hourly observation contains three normalized load coefficients associated with three demand zones and two wind-availability factors associated with two wind zones [8]. The hourly multidimensional observations are therefore clustered jointly. After convergence, the centroid of each cluster defines one representative operating condition, while the number of original hours assigned to that cluster determines its temporal weight. Because the demand and wind components remain paired within the same multidimensional observations throughout the clustering process, this representation preserves a larger part of their empirical joint and spatial structure than the duration-curve construction.

The correlated scenario set [8] is based on 2007 Iberian Peninsula data associated with the Iberian electricity market and the Spanish transmission-system operator. The adopted operating states contain five dimensionless inputs: normalized/per-unit load coefficients for three demand zones and wind-availability factors for two wind zones. The 48 representative centroids and their source-reported cluster durations are used directly in the numerical comparison. Their five-dimensional construction retains joint demand–wind and spatial information within each centroid. As with any centroid-based

aggregation, rare original observations influence representative values through their assigned clusters rather than appearing necessarily as separate operating states.

The K-means approach minimizes within-cluster dispersion and was applied in the original source study [8] to obtain 48 representative multidimensional states. The choice of cluster count reflects a balance between scenario detail and computational tractability; statistical aids such as the gap statistic are discussed in [25], and the broader role of representative periods in energy-system optimization is addressed in [11–14]. The present comparison adopts the published 48-state data alongside the 36 duration-curve operating conditions to examine how the alternative reduced-state descriptions affect investment decisions.

The duration-curve approach assigns each of nine load–wind combinations a conditional weight of 1/9 within its load block. For load-block durations of 1200, 3600, 2400, and 1560 h, the annual contributions of individual states are 133.33, 400.00, 266.67, and 173.33 h, respectively. The correlated set uses cluster-membership durations reported in [8], which range from 4 to 529 h, giving substantially less uniform operating-state weights. For example, centroid 40 combines a nominal-zone demand coefficient of 0.7450 with a southern-zone wind-availability factor of 0.0798 (353 reported h); centroid 41 combines corresponding factors of 0.8495 and 0.0925 (245 reported h). Both capture joint higher-load/lower-wind operating conditions, while cluster centroids generally smooth within-cluster extremes. The case results use these source-supplied weights as published in [8].

3.3. Integration into the Investment Model and Controlled Comparison

The two scenario sets are incorporated into the same optimization model by changing only the scenario-dependent inputs. Specifically, the demand parameter, the wind-availability factor, and the temporal weight of each operating condition are populated from the corresponding duration-curve scenario or K-means centroid. Under the duration-curve representation, annual weighting is determined by the duration of each load block and the probability of the scenario within that block. Under the correlated representation, the weight of a scenario is the number of historical hours assigned to the corresponding cluster. In both cases, these weights enter the expected-revenue term of the upper-level objective, while demand and wind availability enter the market-clearing and wind-production constraints described in Section 2.

The experimental design uses the same investment model and market-clearing/MILP reformulation throughout. Cases 2/6 and 3/7 form the cleanest like-for-like uncertainty-representation comparisons because the deterministic network, wind-site limits, and budgets coincide. Cases 4/8 and 5/9 are informative extensions, but they should not be interpreted as isolating scenario construction alone: the duration-curve cases impose additional spatial wind-intensity multipliers, whereas their K-means counterparts use spatially differentiated coefficients embedded in the scenario inputs. Cases 5/9 additionally share the specified line 7-8 removal. Scenario count also differs (36 for the duration-curve versus 48 for the correlated K-means).

The comparison therefore considers both investment and operating outcomes. The principal indicators are the amount and location of installed wind capacity, the investor's expected annual profit, the resulting LMPs, transmission-line loading and congestion, and wind curtailment. The specific mapping of the scenario coefficients to the buses of each test system and the paired case-study configurations are presented in Section 4. Section 5 then evaluates whether the two uncertainty representations lead to materially different investment and market outcomes under otherwise identical conditions.

4. Test Systems and Case-Study Configuration

To evaluate both the mathematical implementation and the effect of uncertainty representation, nine case studies are considered. Case 1 uses a small 3-bus system as a verification problem, whereas Cases 2–9 use the IEEE 24-bus Reliability Test System (RTS). For the 24-bus system, four deterministic configurations are evaluated first with the duration-curve-based representation and then with the correlated K-means representation described in Section 3. Only the data and assumptions needed to interpret the numerical results are summarized below; standard network data that remain unchanged are referenced rather than reproduced in full.

4.1. Three-Bus Verification System

Case 1 is based on the 3-bus example in [7] and is used to verify the implementation and solution procedure before the model is applied to the larger 24-bus test system. The system, shown in Figure 2a, consists of three buses and three transmission lines, with one conventional generating unit connected to each bus. Bus 1 is the reference bus, wind power investment is allowed only at bus 2, and the only load is connected to bus 3, with a peak demand of 150 MW. All three lines have a reactance of 0.2 p.u. and a transmission limit of 80 MW, while the conventional-generator offer blocks follow the data reported in [7].

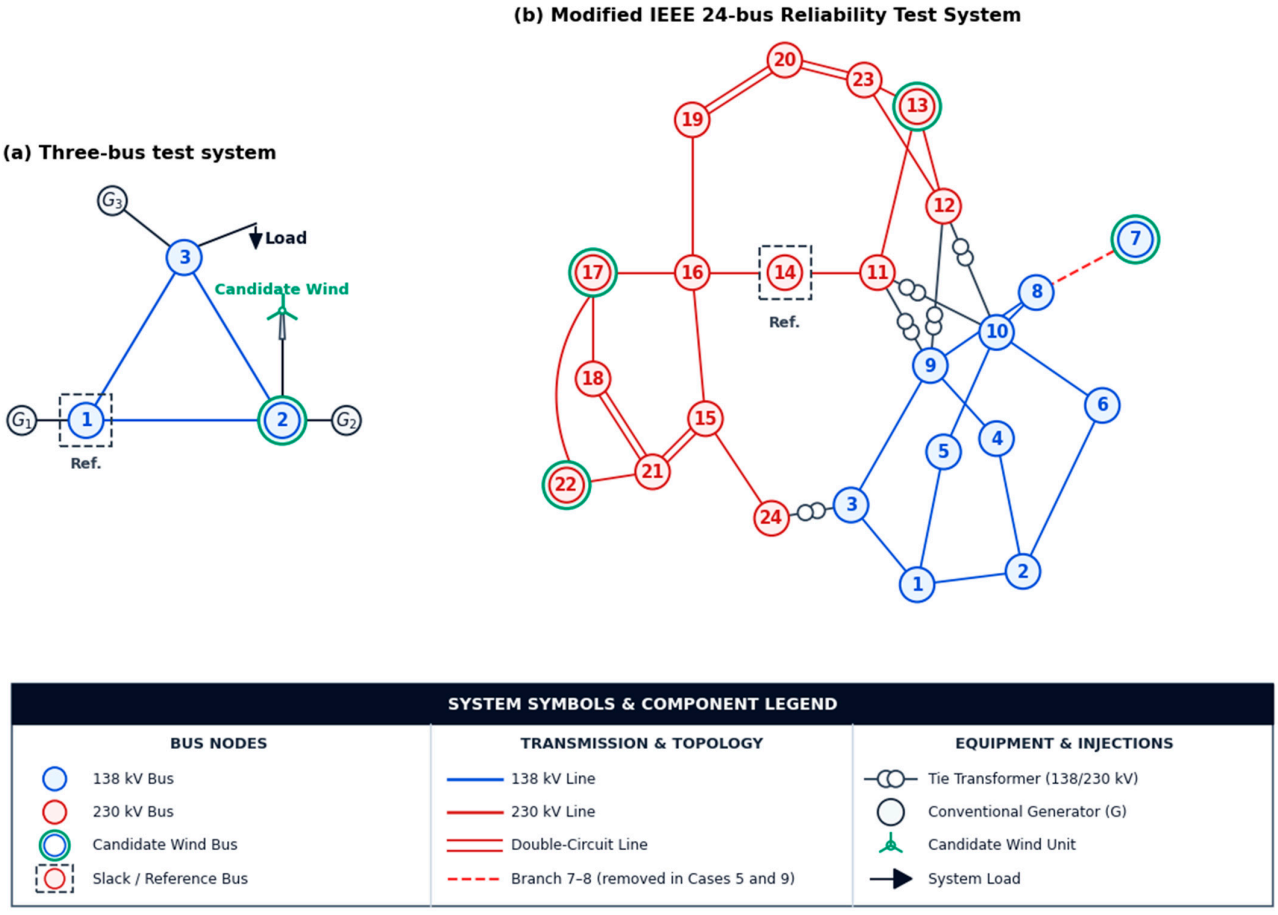


Figure 2. Test systems used in the numerical analysis: (**a**) 3-bus verification system; (**b**) modified IEEE 24-bus RTS. Candidate wind-investment buses are highlighted, and line 7–8 is indicated because it is removed in Cases 5 and 9.

Wind capacity at bus 2 is installed in discrete 30 MW blocks, up to a maximum of 150 MW. The investment cost is €1 million/MW, the recovery period is 15 years, and the capital recovery factor is 11.68%, with a total investment budget of €150 million [7]. Load and wind uncertainty are represented by the 36 duration-curve-based operating conditions described in Section 3.1. The scenario-specific demand coefficients are applied to the load at bus 3, whereas the wind-availability coefficients determine the available wind production at bus 2.

4.2. IEEE 24-Bus Reliability Test System

The main numerical analysis is performed on a 24-bus area of the IEEE RTS [26], using the modifications adopted in [7]. The system, shown in Figure 2b, contains 24 buses, 38 transmission lines, 32 conventional generating units, and 17 loads, with bus 14 selected as the reference bus. Relative to the original RTS data, the peak loads are increased by 50% and the conventional generation capacities are doubled, following [7]. Each conventional generating unit is represented by three offer blocks with increasing offer prices. The complete line and generator data are not repeated here because they are taken directly from the cited test-system and investment-model sources [7,26].

Wind power can be installed at buses 7, 13, 17, and 22 in discrete 50 MW blocks. The corresponding maximum capacities are 600 MW at bus 7 and 300 MW at each of buses 13, 17, and 22. The investment cost, recovery period, and capital recovery factor are the same as in the 3-bus case. For the duration-curve-based cases, uncertainty is represented by 36 operating conditions. For the correlated K-means cases, 48 representative scenarios are used. The latter retain the spatial structure defined in [8] through three demand zones and two wind zones; among the candidate investment buses, bus 7 belongs to the southern wind zone, whereas buses 13, 17, and 22 belong to the northern wind zone. Table 1 presents the main characteristics and wind-investment parameters of the two test systems.

Table 1. Main characteristics and wind-investment parameters of the two test systems.

Parameter	3-Bus System	IEEE 24-Bus RTS
Buses	3	24
Transmission lines	3	38
Conventional generating units	3	32
Loads	1	17
Reference bus	1	14
Candidate wind buses	2	7, 13, 17, 22
Wind-capacity block	30 MW	50 MW
Maximum wind capacity	150 MW at bus 2	600 MW at bus 7; 300 MW at buses 13, 17, 22
Investment cost	€1 million/MW	€1 million/MW
Recovery period/capital recovery factor	15 years/11.68%	15 years/11.68%
Investment budget	€150 million	€1 or €2 billion, depending on case
Scenario representation	36 duration-curve scenarios	36 duration-curve or 48 K-means scenarios

4.3. Case-Study Design

The 24-bus analysis is organized as four configurations under each uncertainty representation. Cases 2–5 use the duration-curve-based scenarios, whereas Cases 6–9 are their corresponding K-means cases. Within each branch, the deterministic system data remain unchanged except for the specific modification stated for the case. This paired

design allows changes in investment and operating outcomes to be interpreted against a common network and market-clearing model.

Case 2 is the baseline 24-bus configuration with an investment budget of €2 billion. Case 3 reduces the budget to €1 billion. Case 4 is based on Case 3 and introduces explicit spatial differentiation of wind intensity by applying a factor of 1.1 at buses 7 and 22 and 0.9 at buses 13 and 17. Case 5 retains these assumptions but removes the transmission line between buses 7 and 8. Cases 6–9 repeat the corresponding configurations using the 48 correlated K-means scenarios. In Case 8, no additional wind-intensity multipliers are imposed because spatial differentiation is already embedded in the correlated scenario set; the case is retained to preserve the one-to-one comparison with Case 4. Case 9 additionally removes line 7–8, as in Case 5. Table 2 presents a summary of the nine case studies.

Under the stated 50 MW investment blocks, the maximum feasible wind allocation at all four candidate sites is 600+300+300+300 = 1500 MW. At EUR 1 million/MW, a budget of EUR 1.5 billion is sufficient to make the total capital constraint nonbinding if all site maxima are installed. The already-reported EUR 1 billion and EUR 2 billion settings bracket this saturation level, but they do not identify the exact budget at which two methods first produce different optimal configurations; that transition requires additional intermediate-budget MILP solutions.

Table 2. Summary of the nine case studies.

Case	System	Scenario Representation	Budget	Defining Feature
1	3-bus	Duration-curve (36)	€150 million	Verification case
2	24-bus	Duration-curve (36)	€2 billion	Baseline
3	24-bus	Duration-curve (36)	€1 billion	Reduced budget
4	24-bus	Duration-curve (36)	€1 billion	Wind factors: 1.1 at buses 7, 22; 0.9 at buses 13, 17
5	24-bus	Duration-curve (36)	€1 billion	Case 4 with line 7–8 removed
6	24-bus	K-means (48)	€2 billion	Counterpart of Case 2
7	24-bus	K-means (48)	€1 billion	Counterpart of Case 3
8	24-bus	K-means (48)	€1 billion	Counterpart of Case 4; spatial differentiation embedded in scenarios
9	24-bus	K-means (48)	€1 billion	Counterpart of Case 5; line 7–8 removed

The resulting matched pairs are Cases 2 and 6, Cases 3 and 7, Cases 4 and 8, and Cases 5 and 9. Section 5 evaluates these cases in terms of installed wind capacity and location, the investor's expected annual profit, LMPs, transmission congestion, and wind curtailment.

5. Results and Discussion

The numerical results are discussed in terms of the investment decision, investor profitability, and the resulting market and network operation. The MILP formulations are implemented in Python 3.12.7 using Pyomo 6.9.5. The results reported below are obtained with Gurobi 13.0.1; CPLEX 22.1.2.1 is used as an independent implementation-consistency check, while CBC 2.10.13 is additionally used for the 3-bus verification case. All computational experiments were performed on a computer equipped with an AMD Ryzen 5 5500U processor with integrated Radeon graphics (2.1 GHz base frequency) and 16 GB of RAM.

For the paired comparisons on the IEEE 24-bus RTS, the Gurobi results are reported throughout. Independent CPLEX runs reproduce the same investment decision and objective-function value in all cases. Case 1 is first used to verify the implemented model, after which the four matched pairs on the 24-bus system are evaluated.

5.1. Model Verification

For the 3-bus verification system, the implemented formulation yields the same optimal solution under Gurobi, CPLEX, and CBC, selecting the maximum admissible wind investment of 150 MW at bus 2. The expected annual revenue is €20.262 million, the annualized investment cost is €17.520 million, and the resulting expected annual profit is €2.742 million. The optimal siting decision and the absence of line congestion are consistent with the benchmark 3-bus problem on which this verification case is based.

The resulting operating indicators are also consistent with the expected behavior of the verification case. The weighted-average LMP is 64.28 €/MWh at all three buses, reflecting the absence of binding transmission constraints during the year. The most heavily loaded line reaches 96.17% of its capacity, while annual wind curtailment is limited to 180.7 MWh, corresponding to 0.06% of the available wind energy. These results confirm that the reformulated model reproduces the expected economic and network behavior before it is applied to the larger IEEE 24-bus RTS.

5.2. Impact on Wind Investment and Profitability

Figure 3 compares the installed wind-capacity allocation by candidate bus and the corresponding expected annual profit for the four matched case pairs on the IEEE 24-bus RTS. The first matched pair, Cases 2 and 6, represents the baseline 24-bus configuration with a €2 billion investment budget. In both uncertainty representations, the budget is nonbinding and the investor reaches the technical wind-capacity limit at every candidate bus: 600 MW at bus 7 and 300 MW at each of buses 13, 17, and 22, for a total of 1500 MW. Thus, when sufficient capital is available to exhaust all candidate-site limits, the different scenario sets do not change the investment decision. The expected annual profit is also very similar, decreasing slightly from €79.722 million in Case 2 to €78.809 million in Case 6, a difference of −1.14%.

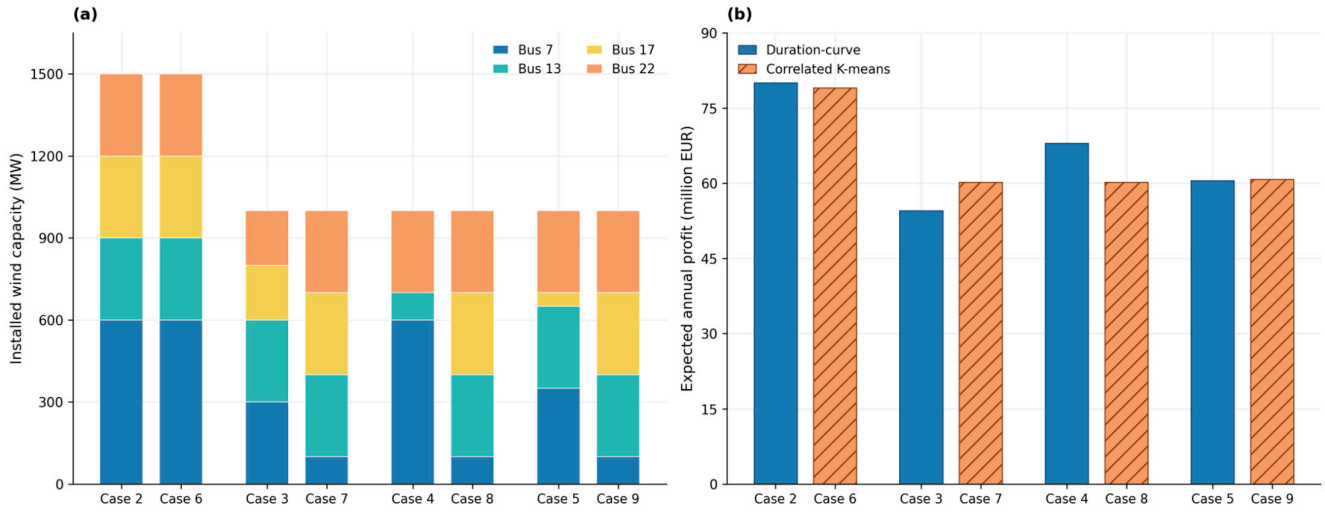


Figure 3. Investment and profitability outcomes in the four 24-bus case pairs: (**a**) installed wind capacity by candidate bus (MW); (**b**) expected annual profit (million EUR).

The effect of uncertainty representation becomes much clearer when the investment budget is binding. In Cases 3 and 7, both formulations install 1000 MW in total, but they allocate this capacity differently. The duration-curve case distributes the capacity as 300/300/200/200 MW across buses 7/13/17/22, whereas the correlated K-means case gives 100/300/300/300 MW. The latter solution places the maximum admissible capacity at the three northern candidate buses and limits the southern bus 7 to 100 MW. This redistri-

bution is accompanied by an increase in expected annual profit from €54.431 million to €59.983 million, i.e., by 10.20%.

The changed siting in Cases 3/7 corresponds to a 200 MW reduction at bus 7 and increases of 100 MW at each of buses 17 and 22, with bus 13 unchanged; the 10.20% profit difference refers to method-specific expected annual objective estimates, not to a realized gain from choosing one scenario method in an external market.

Cases 4 and 8 provide a complementary comparison. Case 4 introduces external wind-intensity multipliers into the duration-curve representation, favoring buses 7 and 22 and penalizing buses 13 and 17. The resulting investment is 600/100/0/300 MW. By contrast, Case 8 retains the correlated scenario data and reproduces the 100/300/300/300 MW allocation of Case 7. The expected annual profits are €67.925 million and €59.983 million, respectively, a difference of −11.69% when the K-means result is measured relative to the duration-curve case. This pair shows that the investment pattern implied by correlated joint scenarios differs markedly from the one obtained when spatial wind differentiation is imposed exogenously on top of a duration-curve representation.

The final pair, Cases 5 and 9, examines the same two representations after removal of line 7-8. The duration-curve case installs 350/300/50/300 MW, whereas the K-means case again gives 100/300/300/300 MW. Despite this substantial difference in siting, the expected annual profits are nearly identical: €60.229 million in Case 5 and €60.592 million in Case 9, a difference of only 0.60%. This pair is particularly important because it demonstrates that similar objective-function values do not necessarily imply similar investment decisions or similar network consequences.

These expected-profit values are outcomes under different reduced operating-state distributions, not directly comparable out-of-sample investment returns. The strictly matched deterministic configurations are Cases 2/6 and 3/7. Cases 4/8 and 5/9 additionally contrast externally imposed versus scenario-embedded spatial wind differences; their numerical contrasts should be described as combined representation/spatial-specification effects, not causal effects of K-means alone.

5.3. Impact on Market and Network Operation

Figure 4 compares the weighted-average system LMP, annual congestion hours, and wind curtailment for the same four matched case pairs. A systematic difference is observed in the weighted-average system LMP. For Cases 2/6, 3/7, 4/8, and 5/9, the duration-curve and K-means values are 80.60/76.89, 81.00/77.33, 80.93/77.33, and 80.91/77.23 €/MWh, respectively. The K-means representation therefore produces a lower weighted-average LMP in every matched pair, with differences between 3.60 and 3.71 €/MWh. This relatively stable offset is associated with the operating conditions retained by the correlated scenario set, including combinations of demand and wind output that differ from those implied by the duration-curve discretization.

The congestion results do not exhibit an equally monotonic pattern. For the reported solutions, annual congestion amounts to 1467 and 1081 h in Cases 2 and 6, 2000 and 1695 h in Cases 3 and 7, 400 and 1695 h in Cases 4 and 8, and 267 and 694 h in Cases 5 and 9. Consequently, the correlated representation reduces the total congestion hours in the first two matched pairs but increases them in the latter two. More importantly, the set of stressed transmission corridors changes with the scenario representation, indicating that the choice of scenarios influences not only the total amount of congestion but also its spatial pattern.

For investment and operations, these opposite signs matter: in the constrained pair 3/7, congestion drops from 2000 to 1695 annual hours, whereas in the spatially differentiated pair 4/8, it increases from 400 to 1695 h. A system operator should therefore assess

congestion location and exposure across several scenario specifications rather than regard lower average LMPs as evidence of uniformly easier network operation.

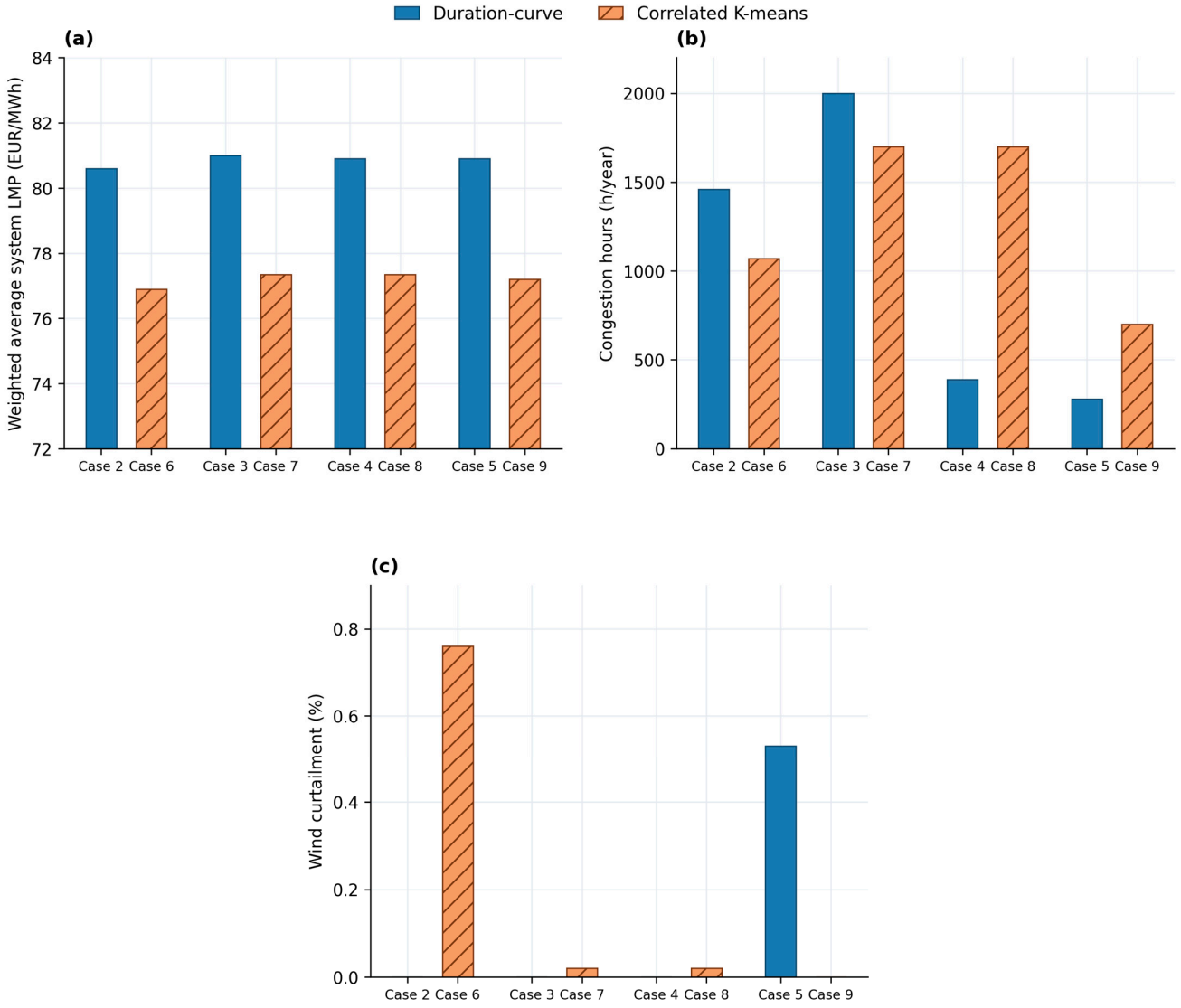


Figure 4. Market and network outcomes in the four 24-bus case pairs: (**a**) weighted-average system LMP (EUR/MWh); (**b**) annual congestion (h); (**c**) curtailed wind as a proportion of available generation (%).

Wind curtailment provides an additional distinction. In the duration-curve cases, curtailment is zero in Cases 2–4 and reaches 0.53% in Case 5 after bus 7 is isolated from the remainder of the network. Under the K-means representation, curtailment is 0.76% in Case 6, 0.02% in Cases 7 and 8, and zero in Case 9. The nonzero curtailment in Case 6 results from operating states with high simultaneous wind availability and limited ability to transfer all available wind production through the network. Conversely, the elimination of curtailment in Case 9 indicates that the correlated representation can shift wind investment and operating weights in a way that relieves the local restriction created by the line outage.

The isolated-bus cases also highlight the local effect of network topology on prices. In both Cases 5 and 9, the weighted-average LMP at bus 7 is 83.00 €/MWh because the

local conventional generator determines the marginal cost of serving the isolated load. At buses 13, 17, and 22, however, the corresponding weighted-average LMP is approximately 80.82 €/MWh in Case 5 and 76.93 €/MWh in Case 9. The lower values in the K-means case are consistent with the greater wind availability assigned to the northern buses under the correlated representation.

5.4. Synthesis of the Paired Comparison

Taken together, the four matched comparisons show that the effect of uncertainty representation depends on which constraint is active and on the operating state of the network. In the unconstrained-investment pair 2/6, both methods select the same capacity because all candidate sites reach their technical limits, yet the two scenario sets still produce different LMP, congestion, and curtailment outcomes. In the budget-constrained pair 3/7, the total investment remains fixed at 1000 MW but the siting changes substantially, indicating that the estimated value of the candidate buses is altered by the way correlation between load and wind is represented.

The budget experiments provide a useful sensitivity benchmark. At EUR 1 billion, the budget binds and Cases 3/7 select different allocations of the same 1000 MW total investment. At EUR 2 billion, Cases 2/6 reach all site-specific capacity limits and select the same 1500 MW configuration. Since the maximum eligible capacity is 600 + 300 + 300 + 300 = 1500 MW and the unit capital cost is EUR 1 million/MW, EUR 1.5 billion is sufficient to finance the full technical installation. Thus, the contrasting EUR 1 billion and EUR 2 billion configurations identify the economically relevant constrained and capacity-saturated regimes.

The comparison of Cases 4 and 8 further shows that a duration-curve model supplemented with externally imposed spatial wind multipliers does not reproduce the investment pattern obtained from the correlated scenario set. The two formulations identify different locations as most attractive and consequently estimate different annual profitability. The pair 5/9 provides an even stronger illustration of why the objective value alone is insufficient for evaluating uncertainty representations: the expected profit is almost unchanged, but the resulting siting pattern, congestion profile, and curtailment level differ materially.

These findings identify uncertainty representation as a substantive modeling choice rather than a neutral preprocessing step. A reduced scenario set can preserve aggregate demand and wind characteristics while modifying the joint operating states that influence the value of candidate wind locations, nodal prices, congestion, and wind power curtailment. The appropriate representation is therefore evaluated through the combination of investment, market, and network indicators, rather than a single ranking based only on the investment objective.

5.5. Implementation Consistency Check

The independent solver verification corroborates the reported computational solutions. CPLEX reproduces the investment decisions and objective-function values obtained with Gurobi in all nine case studies, and CBC reproduces the Case 1 optimum. These cross-solver results provide an implementation-consistency check for the reported cases.

6. Conclusions

This paper examined how electricity-demand and wind-availability scenario representations affect wind-investment decisions in a network-constrained electricity market. Duration-curve and correlated K-means operating states were evaluated within the same stochastic bilevel investment and market-clearing framework. The baseline and re-

duced-budget case pairs isolate the role of the alternative scenario inputs under matched deterministic assumptions, whereas the spatially differentiated cases demonstrate the combined consequences of scenario construction and regional wind representation. Across the investigated systems, investment allocations, estimated profitability, and transmission operating conditions respond to the adopted uncertainty representation.

The paired case studies lead to several practical conclusions. When the investment budget is sufficiently large for all candidate sites to reach their technical capacity limits, the two representations yield the same investment decision, although differences remain in LMPs, congestion, and wind curtailment. When the budget is binding, the allocation of wind capacity among candidate buses becomes sensitive to the retained load–wind dependence and spatial information. This sensitivity is particularly evident in the budget-constrained and network-modified cases, where materially different siting decisions are obtained. The comparison of Cases 5 and 9 further shows that nearly identical expected annual profits can correspond to different investment allocations, congestion patterns, and curtailment levels. Therefore, the objective-function value alone is not sufficient to assess the implications of an investment solution; the associated technical and market outcomes should also be examined.

From the investor perspective, the results indicate that the scenario representation used in long-term planning can change the estimated attractiveness of candidate buses and, consequently, the preferred siting of wind capacity. From the system-operator perspective, the same choice can alter the predicted LMPs, the location and duration of transmission congestion, and the amount of wind energy that must be curtailed. The correlated K-means representation produces lower weighted-average system LMPs in all four matched 24-bus comparisons, but its effect on congestion and curtailment is not monotonic. Hence, the results do not establish one uncertainty representation as uniformly superior. Rather, they demonstrate that scenario construction is a substantive modeling choice whose consequences should be evaluated using both economic and operational indicators.

The scope of the conclusions is defined by the single-target-year competitive electricity-market model with inelastic demand and lossless DC power flow. Storage, demand response, and N-1 security constraints are outside the modeled operating decisions; the line 7-8 removal is evaluated as a fixed network configuration rather than as a contingency-constrained market clearing. Because both scenario representations are processed by the same model, these omissions apply equally to both; nevertheless, co-located storage or flexible demand could shift energy across operating states and reduce the sensitivity of nodal prices and congestion to the retained load–wind structure, which might narrow the differences between the two representations, whereas security constraints would reduce transfer capability and could amplify them. Quantifying these effects requires an extended formulation. The 36-state and 48-state sets also differ in resolution, while spatial wind assumptions coincide directly in the baseline and reduced-budget pairs. These modeling distinctions guide the interpretation of the quantitative results and indicate where extensions involving operational flexibility and network security would be most informative.

Author Contributions: Conceptualization, D.L. and P.S.G.; methodology, D.L. and A.P.G.; software, D.L. and A.P.G.; validation, D.L., A.P.G. and P.S.G.; formal analysis, D.L. and A.P.G.; investigation, D.L. and A.P.G.; resources, P.S.G.; data curation, D.L.; writing—original draft preparation, D.L.; writing—review and editing, D.L., A.P.G. and P.S.G.; visualization, D.L.; supervision, P.S.G.; project administration, P.S.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

https://doi.org/10.3390/en19215297

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

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

Appendix A. Nomenclature

Tables A1–A3 summarize the sets and indices, parameters, and variables used in the mathematical formulation, respectively.

Appendix A.1. Sets and Indices

Table A1. Sets and indices used in the mathematical formulation.

Symbol	Definition
Ω^N	Set of system buses.
Ω^K	Set of transmission lines.
Ω^G	Set of conventional generating units.
Ω^T	Set of load blocks.
Ω_i	Set of production blocks of conventional generating unit i.
Ψ_n^D	Set of loads connected to bus n.
Ψ_n^G	Set of conventional generating units connected to bus n.
Ψ_n^L	Set of candidate wind-investment levels at bus n.
Ψ_t^ω	Set of scenarios associated with load block t.
Δ^{UL}	Set of upper-level decision variables.
$\Delta_{t,\omega}^{\mathrm{LL}}$	Set of lower-level decision variables for load block t and scenario ω.
n^{ref}	Reference-bus index.
n	Bus index.
$n(i)$	Bus to which conventional generating unit i is connected.
k	Transmission-line index.
i	Conventional generating-unit index.
b	Conventional-generation block index.
t	Load-block index.
ω	Scenario index.
l	Wind-investment-level index.

Appendix A.2. Parameters

Table A2. Parameters used in the mathematical formulation.

Symbol	Definition
B_k	Susceptance of transmission line k in the DC power-flow model.
c_{ib}	Offer price or marginal production cost of block b of conventional unit i [EUR/MWh].
$c_{\max}^{\mathrm{inv}}$	Maximum available wind-investment budget [EUR].
c_n^{inv}	Investment cost per unit of wind capacity at bus n [EUR/MW].
$\tilde{c}_n^{\mathrm{inv}}$	Annualized investment cost per unit of wind capacity at bus n [EUR/(MW year)].
$d_{j,t}(\omega)$	Demand of load j in load block t and scenario ω [MW].
$f_k^{\max}$	Maximum transmission capacity of line k [MW].
$g_{ib}^{\max}$	Maximum output of production block b of conventional unit i [MW].
$k_{n,t}(\omega)$	Scenario-dependent wind-availability factor at bus n, load block t, and scenario ω.
N_t^h	Number of hours represented by load block t [h].
$o(k)$	Sending bus of transmission line k.
$r(k)$	Receiving bus of transmission line k.
q_n^l	Wind capacity associated with investment level l at bus n [MW].
$\alpha_t(\omega)$	Probability or temporal weight of scenario ω in load block t.

M_{ϕ}	Big-M constant for transmission-flow complementarity constraints.
M_{φ}	Big-M constant for conventional-generation complementarity constraints.
M_{ξ}	Big-M constant for voltage-angle complementarity constraints.

Appendix A.3. Variables

Table A3. Variables used in the mathematical formulation.

Symbol	Definition
$f_{k,t}(\omega)$	Power flow on line k in load block t and scenario ω [MW].
$g_{ib,t}(\omega)$	Output of production block b of conventional unit i in load block t and scenario ω [MW].
$P_{n,t}(\omega)$	Wind power injected at bus n in load block t and scenario ω [MW].
u_n^l	Binary variable equal to 1 when investment level l is selected at bus n.
X_n	Installed wind capacity at bus n [MW].
$\beta_{n,t}(\omega)$	Wind selling price, equal to the LMP at bus n in load block t and scenario ω [EUR/MWh].
$\delta_{n,t}(\omega)$	Voltage phase angle at bus n in load block t and scenario ω [rad].
$u_{k,t}^{\max/\min}(\omega)$	Auxiliary binary variables for the upper/lower transmission-flow complementarity constraints.
$u_{ib,t}^{\max/\min}(\omega)$	Auxiliary binary variables for the upper/lower conventional-generation complementarity constraints.
$u_{n,t}^{\max/\min}(\omega)$	Auxiliary binary variables for the upper/lower voltage-angle complementarity constraints.
$\lambda_{n,t}(\omega)$	Dual variable of the nodal power-balance constraint; equal to the LMP at bus n [EUR/MWh].
$\phi_{k,t}(\omega)$	Dual variable of the DC power-flow equality for line k.
$\phi_{k,t}^{\max/\min}(\omega)$	Dual variables of the upper/lower transmission-flow limits.
$\varphi_{ib,t}^{\max/\min}(\omega)$	Dual variables of the upper/lower conventional-generation limits.
$\chi_{n,t}(\omega)$	Dual variable of the reference-bus angle equality.
$\xi_{n,t}^{\max/\min}(\omega)$	Dual variables of the upper/lower voltage-angle limits.

References

1. Directive (EU) 2023/2413 of the European Parliament and of the Council of 18 October 2023 amending Directive (EU) 2018/2001, Regulation (EU) 2018/1999 and Directive 98/70/EC as regards the promotion of energy from renewable sources, and repealing Council Directive (EU) 2015/652. Available online: https://eur-lex.europa.eu/eli/dir/2023/2413/oj/eng (accessed on 3 November 2026).
2. Georgilakis, P.S. *Economic and Reliable Operation of Modern Electric Power Systems*; Kallipos, Open Academic Editions: Athens, Greece, 2023. (In Greek)
3. Nicolosi, C.F.; Tina, G.M. Monte Carlo and bootstrap approaches for power system adequacy: A study on atmospheric variables correlations. *Int. J. Electr. Power Energy Syst.* **2025**, *169*, 110827.
4. Barrera-Singaña, C.; Comech, M.P.; Arcos, H. A Comprehensive Review on the Integration of Renewable Energy Through Advanced Planning and Optimization Techniques. *Energies* **2025**, *18*, 2961.
5. Xiao, J.; He, G.; Li, C. Collaborative Renewable Energy Resource Siting and Sizing Planning Method for Distribution and Sub-Transmission Networks. *Energies* **2025**, *18*, 5666.
6. Zhan, B.; Shan, Z.; Zhang, X.; Wang, K.; Yan, R.; Qiu, S.; Fan, Z.; Lin, Q. Topology-Constrained Flexibility Assessment of Adjustable Resources in the Regional Electricity Spot Market. *Energies* **2026**, *19*, 2501.
7. Baringo, L.; Conejo, A.J. Wind power investment within a market environment. *Appl. Energy* **2011**, *88*, 3239–3247.
8. Baringo, L.; Conejo, A.J. Correlated wind-power production and electric load scenarios for investment decisions. *Appl. Energy* **2013**, *101*, 475–482.
9. Conejo, A.J.; Carrion, M.; Morales, J.M. *Decision Making Under Uncertainty in Electricity Markets*; Springer: New York, NY, USA, 2010.
10. Nicolosi, C.F.; Tina, G.M. Impacts of Climate Change on the Wind Power Potential of an Offshore Wind Turbine: Three Case Studies from Sicily. In Proceedings of the 2025 AEIT International Annual Conference (AEIT), Amantea, Italy, 10–12 September 2025.

11. Poncelet, K.; Hoschle, H.; Delarue, E.; Virag, A.; D'haeseleer, W. Selecting representative days for capturing the implications of integrating intermittent renewables in generation expansion planning problems. *IEEE Trans. Power Syst.* **2017**, *32*, 1936–1948.
12. Hoffmann, M.; Kotzur, L.; Stolten, D.; Robinius, M. A review on time series aggregation methods for energy system models. *Energies* **2020**, *13*, 641.
13. Teichgraeber, H.; Brandt, A.R. Clustering methods to find representative periods for the optimization of energy systems: An initial framework and comparison. *Appl. Energy* **2019**, *239*, 1283–1293.
14. Teichgraeber, H.; Brandt, A.R. Time-series aggregation for the optimization of energy systems: Goals, challenges, approaches, and opportunities. *Renew. Sustain. Energy Rev.* **2022**, *157*, 111984.
15. Gabriel, S.A.; Conejo, A.J.; Fuller, J.D.; Hobbs, B.F.; Ruiz, C. *Complementarity Modeling in Energy Markets*; Springer: New York, NY, USA, 2013.
16. Fortuny-Amat, J.; McCarl, B. A representation and economic interpretation of a two-level programming problem. *J. Oper. Res. Soc.* **1981**, *32*, 783–792.
17. Stoft, S. *Power System Economics: Designing Markets for Electricity*; John Wiley & Sons-IEEE Press: New York, NY, USA, 2002.
18. Fabbri, A.; Roman, T.G.S.; Abbad, J.R.; Quezada, V.H.M. Assessment of the cost associated with wind generation prediction errors in a liberalized electricity market. *IEEE Trans. Power Syst.* **2005**, *20*, 1440–1446.
19. Pinson, P.; Madsen, H. Ensemble-based probabilistic forecasting at Horns Rev. *Wind Energy* **2009**, *12*, 137–155.
20. Papavasiliou, A.; Oren, S.S. Multiarea stochastic unit commitment for high wind penetration in a transmission constrained network. *Oper. Res.* **2013**, *61*, 578–592.
21. Hoppner, F.; Klawonn, F.; Kruse, R.; Runkler, T. *Fuzzy Cluster Analysis: Methods for Classification, Data Analysis, and Image Recognition*; John Wiley & Sons: Chichester, UK, 1999.
22. MacQueen, J. Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, CA, USA, 21 June–18 July 1965 and 27 December 1965–7 January 1966*; University of California Press: Berkeley, CA, USA, 1967; Volume 1, pp. 281–297.
23. Ramezani, M.; Singh, C.; Haghifam, M.-R. Role of clustering in the probabilistic evaluation of TTC in power systems including wind power generation. *IEEE Trans. Power Syst.* **2009**, *24*, 849–858.
24. Vallee, F.; Brunieau, G.; Pirlot, M.; Deblecker, O.; Lobry, J. Optimal wind clustering methodology for adequacy evaluation in system generation studies using nonsequential Monte Carlo simulation. *IEEE Trans. Power Syst.* **2011**, *26*, 2173–2184.
25. Tibshirani, R.; Walther, G.; Hastie, T. Estimating the number of clusters in a data set via the gap statistic. *J. R. Stat. Soc. Ser. B* **2001**, *63*, 411–423.
26. Grigg, C.; Wong, P.; Albrecht, P.; Allan, R.; Bhavaraju, M.; Billinton, R.; Chen, Q.; Fong, C.; Haddad, S.; Kuruganty, S.; et al. The IEEE Reliability Test System-1996: A report prepared by the Reliability Test System Task Force of the Application of Probability Methods Subcommittee. *IEEE Trans. Power Syst.* **1999**, *14*, 1010–1020.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.