This is a draft of a paper accepted for the IEEE Power and Energy Magazine vol. 23, no. 3, pp. 31-43, May-June 2025. Please cite the published version: doi: 10.1109/MPE.2025.3542381

Robustness

The Missing Ingredient in Generation Scheduling

By Alexandre Street, Alexandre Moreira, José M. Arroyo, Natalia Alguacil, and Luiz Barroso

Computational models are the foundational tools through which critical elements of electricity markets are represented, including energy and reserve schedules, price signals, and system security. They play a pivotal role in maximizing welfare by co-optimizing system resources and flexibilities while ensuring supply adequacy —cornerstones for reliable, cheap, and efficient power systems—. Therefore, while policymakers pass new laws and design innovative market guidelines, the success of these initiatives critically depends on the accuracy and capability of the models and data used to underpin them. In this context, among the most sensitive aspects to achieving these goals is the capability to integrate 1) robust decision-making modeling, 2) an accurate forward-looking policy framework, and 3) a precise description of relevant system characteristics capturing the main opportunity costs.

The schedule of generation resources to meet electricity consumption in the most economical way while considering the uncertainty in generation and load is essential in any power system. Generation scheduling is a well-known task that has become increasingly challenging in today's power systems, which are significantly different from those in the past. For example, the penetration of variable renewable resources has changed the generation mix globally and introduced many well-documented operational challenges such as the increased need for ancillary services, ramping and peak capacity, and backup generation, as well as reduced levels of inertia. In addition, the security criteria customarily adopted by system operators have proven to be insufficient to cope with the increasingly frequent extreme events involving the failure of multiple power system components such as those related to climate change. These novel features, among others, call for the adaptation of generation scheduling algorithms and tools to manage new and old uncertainties in order to adequately keep up delivering security of supply in the least-cost fashion.

The main topic of this article is robustness, which has emerged as a key ingredient of generation scheduling to attain a practical trade-off between the representation of uncertainties, solution quality, and computational tractability. For the purposes of the discussions, we assume the framework of a pool-based, security-constrained electricity market where generation offers and consumption bids are centrally managed by a system operator using a market-clearing procedure driven by social welfare maximization, as is the case of the power markets in the U.S. and Latin America.

1. Generation Scheduling

Generation scheduling involves the determination of the most economical and security-compliant operation of generation assets over a specific time span so that power balance is met without violating the limits characterizing the functioning of power system components. Security refers to the capability to withstand perturbations, e.g., demand fluctuations and component failures.

Generation scheduling was originally implemented by vertically integrated utilities operating as monopolies within their service territories. Under such non-competitive framework, the generation mix typically comprised conventional generators, namely fossil-fueled thermal power plants and, where available, hydroelectric units. Driven by the dynamics of the generation fleet, a look-ahead scheduling horizon spanning one day to one week with an hourly discretization was commonly used to capture the main system's opportunity costs due to relevant inter-temporal resource constraints. Moreover, consumers were considered as passive players thus characterized by inelastic demands, for which accurate short-term forecasting tools were available. As for security, system-wide reserve requirements were usually imposed to ensure a specific generation margin to address uncertainty in real time without explicitly considering its effect in the scheduling model. As a consequence, the goal of the so-called unit commitment problem was to minimize the total cost incurred by the use of the generation resources over the aforementioned look-ahead time span. The resulting cost minimization problem was subject to both generation-related constraints, e.g., production limits, ramping rates, and minimum up and down times, and system constraints, such as those associated with Kirchhoff's laws, the effect of the transmission network, and security.

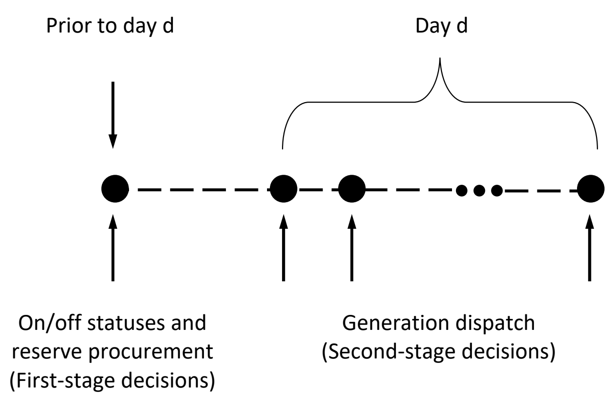

Figure 1. The two-stage decision framework for generation scheduling

As shown in Figure 1 for a day-ahead setting, the conventional unit commitment can be characterized as a two-stage decision-making problem, where first-stage decisions comprise the on/off statuses of generating units and the reserve procurement, whereas second-stage decisions correspond to the generation dispatch. Interestingly, generation resources bring important and complex inter-temporal constraints to the scheduling problem, which, in most cases, require the consideration of an extended look-ahead setting, beyond the scheduling horizon, to ensure proper characterization of the future system's opportunity costs and its correct consideration in the first-stage decisions. Note that future inter-temporal constraints and the uncertainty level to which the decisions affected by these constraints are subject constitute inseparable parts forming the relevant system's opportunity costs. Therefore, the look-ahead framework in generation scheduling policies should carefully observe the interactions of both salient features.

Depending on the system characteristics and the regulatory framework, different implementations of generation scheduling models can be found. For instance, Brazil was a latecomer in the utilization of short-term optimization-based scheduling tools, only implementing day-ahead scheduling based on a deterministic security-constrained unit commitment problem at the beginning of 2020. The "infinite" flexibility brought by a hydro-dominated system with large storage capacity postponed the implementation of these tools. The official short-term model now provides the unit commitment for thermoelectric generators and the energy dispatch for both thermal and hydroelectric units. Reserves are mainly scheduled on hydroelectric power plants according to a 5%-demand exogenous reserve requirement. A post-scheduling adjustment period before real time is then used to determine the unit commitment of hydroelectric units and refine reserves according to ad hoc procedures, where primary and secondary reserves are allocated according to 1%- and 4%-demand requirements, respectively. Notwithstanding, as is customary in industry applications, reserve deliverability is not endogenously considered.

Within a competitive context, unit commitment models are used by system operators to clear pool-based electricity markets, i.e., to determine the set of awarded generation offers and consumption bids. Such generation scheduling tools feature some distinctive modeling aspects compared to the classical non-competitive approaches, namely:

- ✓ The incorporation of the role played by consumers as active market agents;
- ✓ The use of a different optimization goal: the maximization of the social welfare associated with the offer and bid functions; and,
- ✓ The explicit characterization of reserves as trading commodities.

Note, however, that the unit-commitment-based market-clearing models also feature the two-stage structure depicted in Figure 1.

2. The Changing Paradigm of Generation Scheduling

The growing concerns about sustainability and climate change have a significant impact on the electric power sector around the world. Thus, existing practice for generation scheduling is challenged in the current restructured, increasingly decarbonized, and vulnerable power industry.

2.1. Increasingly Decarbonized Context

Along with the regulatory change to a market setting, the power industry is also immersed in the transition to a decarbonized context where fossil-fueled thermal power plants are replaced with renewables-based generation, particularly wind and solar power. The installation and use of such environmentally friendly resources have experienced a sharp growth worldwide, as illustrated in Figure 2 for Spain. As can be seen, the total installed capacity of wind and solar power generation has increased from an insignificant 3% in 2000 to around 47% in 2023. More importantly, the fraction of demand supplied by both renewable sources has risen from 2% up to 39% in such a relatively short time interval. Interestingly, wind energy is the leading technology but solar energy has been recently gaining momentum.

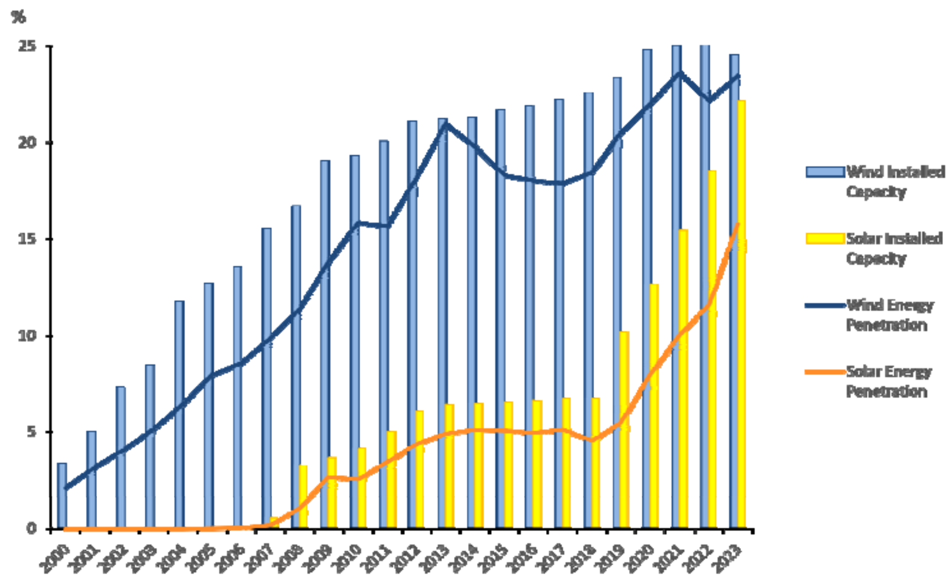

Figure 2. Evolution of the integration of wind and solar power generation in Spain (Source: Red Eléctrica de España)

Box 1. Renewables integration and the need for generation scheduling models considering uncertainty

It is easy to draw a simple example of two systems, A and B, with the same flat demand profile but different generation fleets, where the integration of wind generation would result in entirely different outcomes. For instance, system A's conventional generation is 50% from flexible hydroelectric units and 50% from a slow but cheap combined-cycle (CC) power plant. By contrast, system B's conventional generation is 50% from (inflexible) nuclear and 50% from a CC plant identical to that used in system A. Both systems also have a quick and expensive unit for backup. In this hypothetical yet archetypical example, system A could compensate for wind variability using the flexibility of hydroelectric power plants. Consequently, system demand would be supplied by either hydroelectric and CC units when no wind generation is available, or by a portfolio of hydro and wind generation complementing a reduced number of CC generating units when wind is abundant. So, in the case of system A, the total long-run cost and emissions would be reduced by the integration of variable renewable generation from wind. Real-time prices could also be reduced if only hydroelectric and wind power plants supply demand. As for system B, due to its inability to accommodate the sharp ramping rates imposed by the integrated wind generation, the whole schedule would be challenged. In this case, the slow CC units' ramping capacity, the wind generation's variability level, and the regulatory framework on renewables curtailment would determine the new schedule. Thus, one of the slower yet cheaper nuclear or CC units would typically be scheduled off to let the expensive backup generator operate as a flexible resource for compensation, which may substantially increase costs, price volatility, and emissions. Additionally, existing generators would significantly increase their cycling rate, which would raise maintenance costs, eventually leading to potentially higher price offers. The final long-run cost reduction or increase would depend on the market bid rules for renewables, current subsidies, and the curtailment policy, which is also not exempt from subsequent side effects worth avoiding.

Despite the appealing yet well-known dilemma raised by this illustrative example, two key elements frequently ignored in the discussion are the scheduling model and how it considers uncertainty. Both aspects are at the core of the concepts used in the example yet assumed as granted therein. Even in the case of system A, with hypothetically sufficient generation resources to cope with wind variability, if the scheduling model does not account for the uncertainty it is subjected to in the following hours and does not allocate reserves properly, it may overschedule hydroelectric generation and fall short of ramping capacity if the expected high wind generation does not materialize. Hence, the scheduling model and its ability to account for uncertainty are intrinsic and determinant parts of the discussions about fostering renewables integration and the broader energy transition agenda.

This is a draft of a paper accepted for the IEEE Power and Energy Magazine vol. 23, no. 3, pp. 31-43, May-June 2025. Please cite the published version: doi: 10.1109/MPE.2025.3542381

Box 2. Renewables integration, storage, and complexity of generation scheduling models

In some countries, the expansion of renewables is smoother than in others. This is certainly the case of Brazil, which enjoys most of the main ingredients needed to easily absorb variable resources, namely high shares of hydro generation, a fully interconnected system with 168,858 km of high-voltage transmission lines (as of December 2022), and an integrated system resource coordination centrally carried out by a single Independent System Operator. The consumption in 2022 was 69.7 avg GW, with 72.3% supplied by hydroelectric plants, 13.3% by wind farms, 2.0% by solar generators, and the remainder by thermal power plants, resulting in an 87.7% renewable energy supply. In 2024, this picture evolved to a consumption of 79.9 avg GW, with 61.5% supplied by hydroelectric plants, 15.1% by wind farms, and 10.5% by solar generators, amounting to an 87.1% renewable energy supply, which is almost equal to the 2022 percentage but with an increased participation of wind and solar power. This places Brazil as one of the most renewable power systems in the world. However, the high reliance on renewable yet variable and uncertain generation resources also brings highly complex challenges on the modeling side. For instance, the 2024 Mid-Term Electric Operation Planning report (PAR/PEL 2024) from the Brazilian National System Operator indicates that, on August 11, 2024, hydroelectric plants provided most of the ramping capacity, namely 28 GW, required to compensate for the net demand increase between 1 PM —featuring the least net demand ever—, when distributed generation accounted for 28.7% of the total supply, and 7 PM, when solar generation was no longer available. The same report foresees that the total required ramping capacity will increase from the current 36.1 GW to approximately 52.2 GW. Figure 3 depicts the typical generation profile (monthly and hourly averages) of a wind power plant in Northeastern Brazil. Note that the hourly profile presented in the right-hand-side chart corresponds to the monthly average depicted as an orange dot in the left-hand-side chart. In this region, where the total generation exceeds the total load in some days, a significant part of the system demand is expected to be supplied by either centralized or distributed intermittent renewables by 2030. Therefore, the Brazilian system operator must consider in its scheduling model two aspects: 1) the intermittency of hourly generation profiles of wind and solar power, which requires the typical transmission- and unit-commitment-aware coordination, and 2) the uncertainty in the total energy added to the system in different months and hours to withstand hydrological dry periods and the increasing intraday net demand variability. Thus, if, on the one hand, the high storage capacity allows for a smoother integration of intermittent renewables in the short term, on the other hand, the long-term storage management policy becomes more complex. As a consequence, more sophisticated scheduling models are needed so that 1) a multiscale uncertainty representation is considered, and 2) long-term resource needs are integrated into the short-term scheduling decisions. Notwithstanding, the installation of new controllable resources has not kept pace with the rapid increase in wind and solar participation in the Northeastern subsystem, primarily due to a lack of proper market incentives. This situation is currently posing significant operational challenges. As an example, in September 2024, the Brazilian system operator had to curtail 15.5% and 20.4% of wind and solar power generation, respectively. Therefore, the evolution of models must be accompanied by regulatory updates that provide appropriate incentives, remunerating agents for their services.

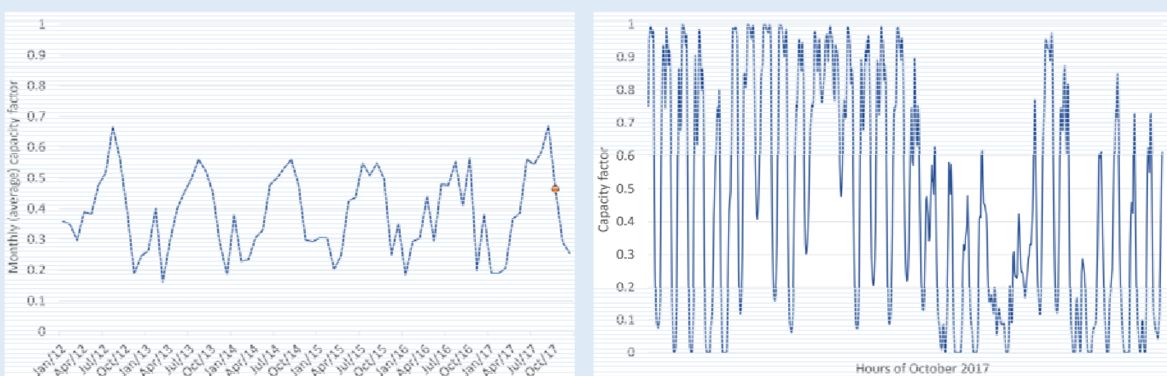

Figure 3. Monthly and hourly capacity factors of a typical wind power plant in Northeastern Brazil

Unfortunately, the production of wind and photovoltaic solar power plants is both variable and uncertain. Variability of renewables-based generation refers to the constant change over time, including the different seasonality cycles within days and years, which is strongly location dependent. Uncertainty is defined as the lack of information on the dynamics of the maximum available power from such variable generation assets, which gives rise to forecast errors in the various forecast horizons used in both short-term operation and mid- or long-term planning. Thus, the massive integration of variable and uncertain renewables in centralized and distributed forms has brought significant challenges for all agents, from the system operator to the conventional sources and consumers. The primary challenge materializes in real-time operation and day-ahead scheduling. At first sight, the more renewable generation we integrate into the network, the cleaner and cheaper the final generation profile used to supply demand should be. However, depending on the system's characteristics and the adopted generation scheduling model, the opposite can be true. As shown in Box 1, the existing generation fleet, the installed network, the location and the uncertainty profile of integrated renewables, as well as the choice of the scheduling model play a crucial role in deciding the generation dispatch and therefore in the final generation cost, electricity prices, and total emissions. It is worth pointing out that, even in the case of systems with high storage and transmission capacity, where the integration of intermittent renewables is smoother, new challenges on the modeling side appear as discussed in Box 2.

2.2. Increasingly Vulnerable Infrastructure

Power systems are also subject to new operating conditions due to internal and external factors. Among the former, it is worth highlighting that economic, environmental, and political reasons have prevented the transmission network from being expanded according to the increase rates of both consumption and generation. Consequently, the transmission network is being operated close to its static and dynamic limits, two practical examples being the Spain-France interconnection, which was congested over 60% of the time in 2024, and the major Brazilian blackout in August 2023, caused by dynamic limits being reached. In addition, congestion is driving significant hourly average nodal price differences in a number of transmission links within the U.S. As for the latter, power systems are increasingly being exposed to extreme yet rare weather-related events with potentially devastating impacts, eventually leading to interrelated simultaneous failures of multiple system components. The California heat waves in 2020 and 2022, the Texas 2021 winter storms, and the unprecedented 2024 flood crisis in the state of Rio Grande do Sul in Brazil, among others, illustrate significant climate-related multiple-contingency events that modern planning and operation practices need to address. In addition to the growing weather-related uncertainties, which affect renewable generation and system component availability, physical and cybernetic attacks also constitute relevant sources of uncertainty nowadays.

This unprecedented vulnerability represents an additional challenge for generation scheduling. Traditional deterministic n-1 and n-2 security standards rely on either 1) the imposition of pre-specified system-wide reserve requirements, or 2) the endogenous determination of reserve needs. The first scheme is computationally simpler and allows straightforwardly pricing reserves as the shadow prices of the corresponding constraints. However, in the current context with an ever-increasing number of congested transmission lines, the compliance with system-wide reserve requirements may no longer ensure security due to the potential inability to deploy reserves across the network. In addition, ex-ante reserve requirements may be hard to determine when events beyond those considered by standard security criteria come into play as the required engineering judgment may be unavailable. As for the second alternative, reserve deliverability is ensured as long as the effect of the transmission network is properly accounted for. However, the need for explicitly characterizing system operation under every operating condition through contingency-constrained models is likely to exceed current computing capabilities should more stringent n-K security criteria be adopted.

3. Robust Generation Scheduling

In the last decade, relevant discussions have been held on how to adapt existing power system infrastructure and practices to the above-described new landscape. The increased vulnerability of power systems to renewables variability and climate-related extreme events, such as those recently experienced across the globe, has triggered a wave of new research and developments in both stochastic programming (SP) and robust optimization (RO) models applied to generation scheduling. Key aspects addressed by the latter include: 1) the challenging task of devising more sophisticated security criteria based on the similarities between RO and current industry practices, which use expert-defined worst-case scenarios, 2) computational tractability, leveraging decomposition approaches and the potential two-way insights resulting from the concepts shared by RO and practical standards, and 3) the development of the more general distributionally robust optimization (DRO) models capable of generalizing SP and RO within a single framework that minimizes or constrains worst-case expected value-based functions.

3.1. Handling Uncertainty through Stochastic Optimization

Deterministic optimization approaches considering a single realization of uncertain parameters have been traditionally implemented to schedule generation. Between the end of the 1990s and the beginning of the 2000s, major system operators started to adopt mixed-integer linear programming (MILP) to solve their market-clearing problems. Over the last two decades, MILP methods have been fairly successful in supporting the decision making behind the actions of system operators. Unlike originally conceived power systems, where generation was firm and predictable and loads were inelastic, modern and future power systems will have to balance intermittent uncertain supply from renewables and potentially price-responsive demands. This new paradigm entails a massive presence of uncertainty and flexibility resources that cannot be properly characterized by purely deterministic approaches. Rather, the value of flexibility resources can only be more comprehensively captured in scheduling models when uncertainty is accounted for. Therefore, the operation of modern power systems aiming to properly schedule and value the opportunity costs of flexible resources needs to consider uncertainty-aware scheduling models.

One of the most popular frameworks to deal with uncertainty in optimization problems is SP, which provides decision makers with some key features that allow a systematic modeling of uncertainty, namely:

- ✓ The hierarchical decision-chain process where first-stage decisions are followed by recourse actions;
- ✓ The explicit incorporation of uncertainty via the consideration of probability distributions; and,
- ✓ The possibility of optimizing different metrics according to the risk aversion of the decision maker.

More importantly, by using SP, it is possible to quantify the benefit, or the value, of the stochastic solution. This value is intrinsically associated with the flexibility provided by the solutions from SP methods to adapt the first-stage decisions to a myriad of potential scenarios that may materialize.

Despite the very appealing characteristics of SP methods, there are still critical barriers that prevent their adoption for generation scheduling. One of these problems is their reliance on probabilistic forecasts, which need to account for both temporal and spatial dependencies for a high-dimensional time series describing demand and renewables uncertainties. In this context, the practical application of SP in short-term power system operation would involve the formulation of a Sample Average Approximation (SAA) of the problem that gives rise to large-scale MILP models, where the underlying probability distribution is described (with some degree of fidelity) by a set of sampled scenarios. In order to alleviate the computational burden of the resulting scenario-dependent problems, one typical approach is to strategically select a reduced number of scenarios, which has the usual effect of compromising the approximation of the stochastic process behind uncertainty. In addition, the use of scenario reduction yields sample-dependent solutions, which significantly challenges the transparency of the whole process and the reproducibility of the results. This fact would be of major concern for agents who would have their offers/bids awarded or not according to an eventual SAA-based market clearing whose prices are also sample dependent. State-of-the-art research on importance sampling and decomposition techniques such as adaptive partitioning has significantly improved the capability to solve SAA-based models and mitigated some of the previously raised issues. However, the non-trivial steps needed to achieve good approximations, the scenario-dependence shortcoming, and the difficulty in explaining solutions that come from stochastic models still could not produce a simple enough-to-implement process and the necessary momentum to convince system operators to test these models in practice.

3.2. Handling Uncertainty through Robust Optimization

Differently from SP, RO provides a framework where it is not necessary to have full knowledge of the underlying process behind uncertainty realization. Therefore, RO constitutes an alternative simpler way to account for uncertainty that may be difficult to represent in a probabilistic manner. Indeed, the description of uncertainty through a probabilistic (stochastic) framework in the decision-making process is a choice itself. As will be further explained, RO offers a balance between simplicity of the uncertainty modeling and tractability of the resulting optimization problem.

The first RO approach proposed in the 1970s modeled uncertainty by means of box constraints defined by the lower and upper bounds for the corresponding random variables against which the decisions would be immunized. Such an approach offered the benefits of decreased computational burden and simplicity in the form of linear constraints at the expense of potentially leading to over-conservative solutions. Efforts to alleviate this issue included modeling uncertainty via nonlinear ellipsoidal constraints, linear budget-constrained polyhedral uncertainty sets, and data-driven polyhedral uncertainty sets, which allowed the user to specify the conservativeness level by intuitive parameters and brought the possibility to use real data to approximate uncertainty sets to real processes. Ellipsoidal uncertainty sets enable a direct link between the conservativeness parameter and the confidence level of chance constraints based on joint Gaussian distribution confidence intervals. Consequently, such uncertainty sets are often used to model correlated demand and renewable generation. As a downside, ellipsoidal uncertainty sets result in nonlinear second-order cone constraints, which preclude the use of powerful off-the-shelf MILP solvers in favor of less efficient mixed-integer nonlinear programming solvers. On the other hand, the budget-constrained approach allows the interpretation of its conservativeness parameter as the number of uncertainty factors that would simultaneously deviate from their nominal values to their worst-case values. This interpretation resembles classical n-K security criteria used to address generator and line contingencies where up to K components can deviate from available to unavailable in the worst-case scenario. Alternatively, data-driven polyhedral uncertainty sets aim at improving the description of correlated uncertainty by considering as vertices of the uncertainty set data either observed within a recent time window or selected by state-of-the-art machine-learning algorithms. Thus, the true complex spatial-temporal dependencies are accounted for. Additionally, this approach also has the advantage of a budget-constrained polyhedral uncertainty set with a polynomial number of vertices, in contrast to the combinatorial number of vertices in a box-constrained uncertainty set. Figure 4 illustrates the evolution of the uncertainty sets from the box-constrained to the more recent data-driven. Subsequent advancements resulted in the development of the adjustable robust optimization or two-stage class of problems, which introduced within the RO framework the concepts of first-stage and recourse decisions customarily used in SP and consistent with industry practices. Such advancements inspired several pioneering works in the technical literature on power systems.

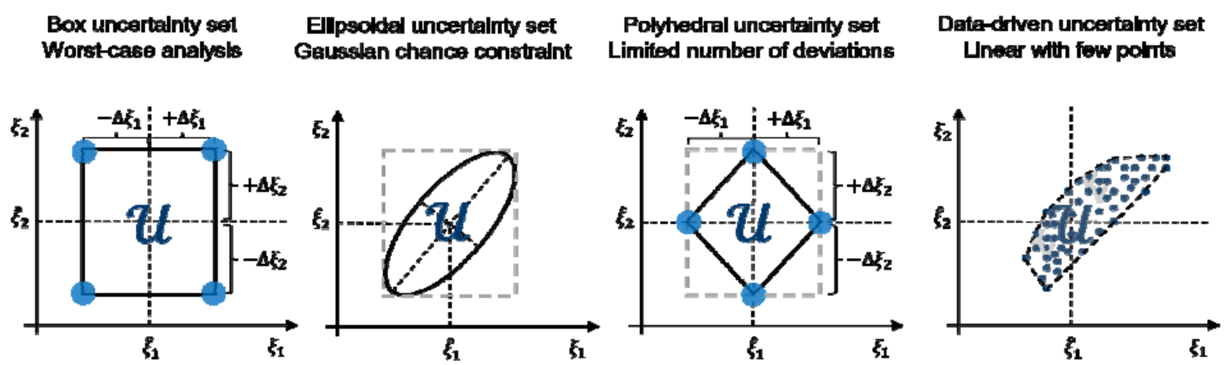

Figure 4. Different types of uncertainty set typically used in robust optimization (ξ represents an uncertainty factor whereas $\hat{\xi}$ and $\Delta\xi$ stand for its nominal value and variability, respectively)

More recently, as a further improvement, DRO has been proposed as a compelling alternative to improve the uncertainty description in RO and decrease the gap with SP models in terms of both conservativeness and usage of relevant data-driven information, which are the two main limitations featured by RO. This is done by leveraging incomplete (or imprecise) but relevant available information of the underlying stochastic process without requiring its full specification. The DRO framework generalizes the stochastic and robust frameworks by means of the minimization of worst-case expected values over ambiguity sets, i.e., over sets of possible probability distributions. Different ambiguity set designs have been proposed in the literature. One approach relies on the set of all probability distributions with a given support and specified first and second moments (mean and variance), leading to ambiguity sets based on moment constraints. Another method adopts a more data-driven approach, generating ambiguity sets comprising all probability distributions that deviate from an empirical distribution —derived from data— by no more than a specified robustness parameter under a chosen probability distance measure. This data-driven approach leverages statistical convergence properties to construct ambiguity sets grounded in real-world data and relies on the Wasserstein distance as one of the most popular metrics due to its relevant properties. Interestingly, DRO models are often used to represent chance constraints, as the probabilities being selected in the ambiguity set can be used to constrain scheduling decisions to reliable operating points. Additionally, DRO models can also be employed to capture certain instances of RO. For example, DRO models reduce to particular cases of RO if we consider 1) the specific support for the random variables that recover the desired RO uncertainty set, and 2) an ambiguity set with no additional constraint beyond those needed to define a probability distribution, i.e., probability densities should be non-negative and sum up to one. This ends up in worst-case distributions that collapse all the probability mass in the worst-case scenario of the uncertainty set, thereby reproducing RO models.

Finally, DRO has recently been used to efficiently model a computationally challenging, and thereby rarely addressed, class of decision-making problems under uncertainty including decision-dependent probabilities. A relevant practical example is planning infrastructure hardening for earthquakes and other natural hazards, where investments in certain assets alter their failure probabilities. Additional practical instances include line undergrounding and equipment maintenance in electrical distribution networks, among others. Therefore, DRO represents a powerful framework that remains far from fully explored and constitutes a relevant research area with promising practical impacts. Figure 5 illustrates the main types of uncertainty characterization in two-stage decision-making problems under uncertainty.

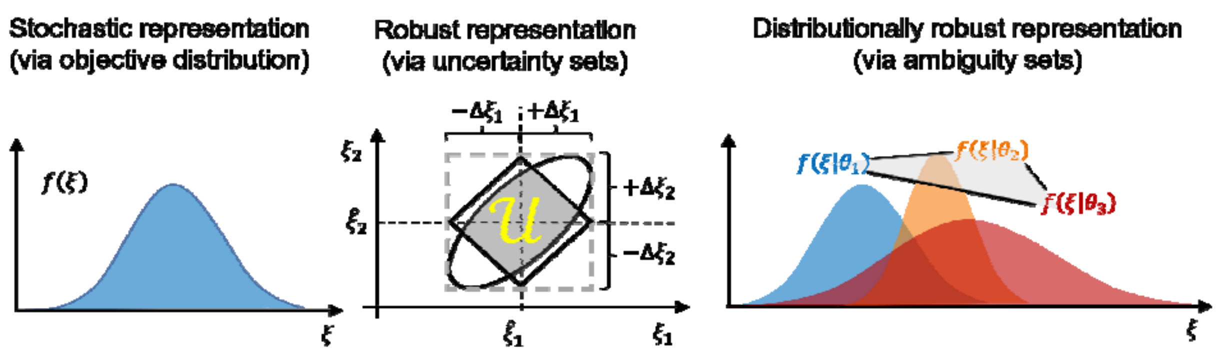

Figure 5. The main uncertainty description approaches for stochastic (single probability distribution), robust (uncertainty sets), and distributionally robust (ambiguity sets –sets of probability distributions–) optimization

In general, robust models are built to behave as risk-averse models, i.e., to protect the first-stage solution against adverse conditions or high costs, which is typically addressed through risk measures (e.g., conditional value-at-risk or chance constraints) in stochastic optimization. To that end, the conservativeness level (term used as a counterpart of risk aversion in the absence of a probabilistic setting) is accounted for in RO through the size and shape of the uncertainty set. Additionally, it is worth noting that SP models intrinsically rely on the value of integrals (measures) over the whole set of sampled scenarios characterizing the uncertainties. By contrast, RO relies on feasibility constraints individually enforced for every scenario in the selected uncertainty set. Hence, one relevant advantage of RO over the SP framework is that, for many classes of optimization problems, the feasibility requirement for all scenarios is equivalent to requiring feasibility for just a very reduced set of the worst-case scenarios, also known as umbrella set of scenarios. This salient feature of robust models grants a great computational advantage over SAA-based SP models. In Table 1, we present a summary of the main modeling characteristics of the three uncertainty representation approaches.

Table 1. Uncertainty representation approaches in optimization models for decision making under uncertainty

Approach	Uncertainty description	Optimization process
SP	Probabilistic framework (multivariate probability density functions)	Based on expected value-based measures, but implemented in practice by extensive enumeration of scenarios in the optimization model (SAA).
RO	Uncertainty set (multivariate sets of points generally defined by bounds for the variables)	Based on feasibility for all scenarios in the uncertainty set, but implemented in practice by a reduced umbrella set of worst-case scenarios that must be discovered through iterative methods.
DRO	Ambiguity sets or sets of distributions (multivariate worst-case distributions defined over uncertainty sets specifying only limited information about the distributions in the ambiguity set, e.g., first and second moments)	Based on worst-case expected value-based measures, but implemented in practice by fewer scenarios than in SAA-based SP models. These fewer scenarios are discovered through iterative methods to compose the worst-case expectations. Solution methods are closely related to RO.

3.3. Generation Scheduling Models and Robust Optimization

As mentioned above, the n-K security criterion is a deterministic standard largely employed in power system operation and planning activities, typically for values of K equal to 1 and even sometimes for K equal to 2. However, what is not widely known about this security criterion is that it is intrinsically rooted in RO concepts. The n-K criterion requires the system to guard against any contingency state involving the loss of up to K components, which is equivalent to ensuring that the system withstands the worst-case contingency with an uncertainty budget of up to K elements deviating from their nominal state (in service), precisely as done in RO. Therefore, as soon as this link was made in the literature on power system operation, the classical yet computationally intractable contingency-constrained models, which explicitly consider all n-K contingency scenarios to endogenously define reserve needs with deliverability guarantees, could be recast as RO models and more efficiently solved through decomposition algorithms. So, although the computational benefits of RO solution methods relying on worst-case structures were only recently developed (in the last decade) by the academic community, the actual utilization of RO concepts implicitly embedded in the n-1 and n-2 security criteria is much older (it dates back to the 1980s) and has been widely considered in power systems standards and industry's activities through many different ad hoc implementations. This fact shows that power systems worldwide largely rely on RO concepts. Moreover, this fact also brings evidence that introducing RO solution methods into industry practices has the potential to significantly improve the security and economics of reserve allocation, sizing, and deliverability, thereby improving the quality of the service to end consumers.

A salient feature of robust solutions is that the protection provided against worst-case scenarios within a given uncertainty set often extends to other adverse scenarios outside the uncertainty set, an insight that can explain, enrich, and promote new power system theories and practices. For instance, n−2 reliable schedules can also withstand contingencies involving three or four simultaneous failures, provided that these do not constitute the worst-case scenarios under the n−3 and n−4 criteria. Notwithstanding, this relevant feature of robust solutions extends beyond the traditional n−K criteria, including any other uncertainty source, which can be considered or combined with contingencies. Likewise, the RO framework is general enough to allow the formulation of models that describe the generation scheduling problem under multiple sources of uncertainty, such as renewables injection, fuel availability, and equipment performance, among others, as well as multiple simultaneous security criteria with compound security levels. Due to the "here and now" (e.g., day-ahead) and "wait and see" (e.g., real-time) structure of generation scheduling, the class of two-stage (or the more sophisticated multistage) robust models is particularly tailored to characterize this decision-making process. More specifically, considering as first-stage decisions the on/off statuses of generating units, the reserve procurement, and the nominal dispatch, and as second-stage decisions those related to real-time operation, the problem can be equivalently formulated as either a monolithic enumerated model or a trilevel optimization model.

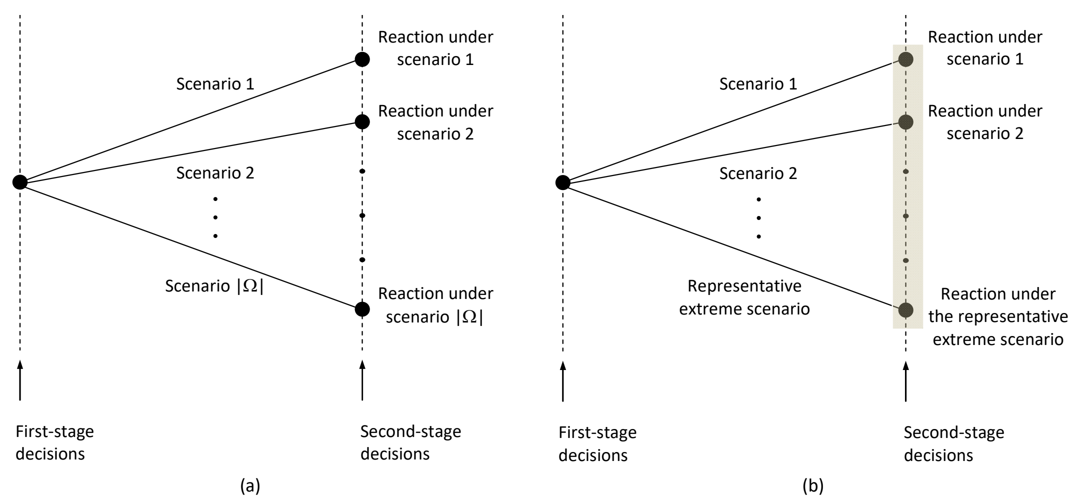

Figure 6. The two-stage decision frameworks for robust generation scheduling: (a) Under a finite set of scenarios, (b) Under an infinite set of scenarios

The former is deemed more intuitive as it is presented as a single-level optimization problem. In this format, constraints affected by the uncertainties are explicitly and exhaustively enumerated for all scenarios in the uncertainty set, generating a huge constraint set if the n-K criterion is modeled, and an infinite number of constraints in the case of renewables and load uncertainties. Figure 6 depicts the corresponding two-stage decision-making frameworks by focusing on the differences between first- and second-stage decisions and how such decisions are related to the scenarios used to characterize uncertainty. Note that both decision-making frameworks can be viewed as natural extensions of that represented in Figure 1. In the monolithic enumerated model, the characteristics of the uncertainty set are not explicitly defined; only a repetition of uncertainty-affected constraints is considered for a general set representing all possible scenarios. Notwithstanding, a relaxed version of this model, wherein only a few scenarios are considered, plays the role of master problem in most of the decomposition methods used to tackle this problem, as will be further discussed.

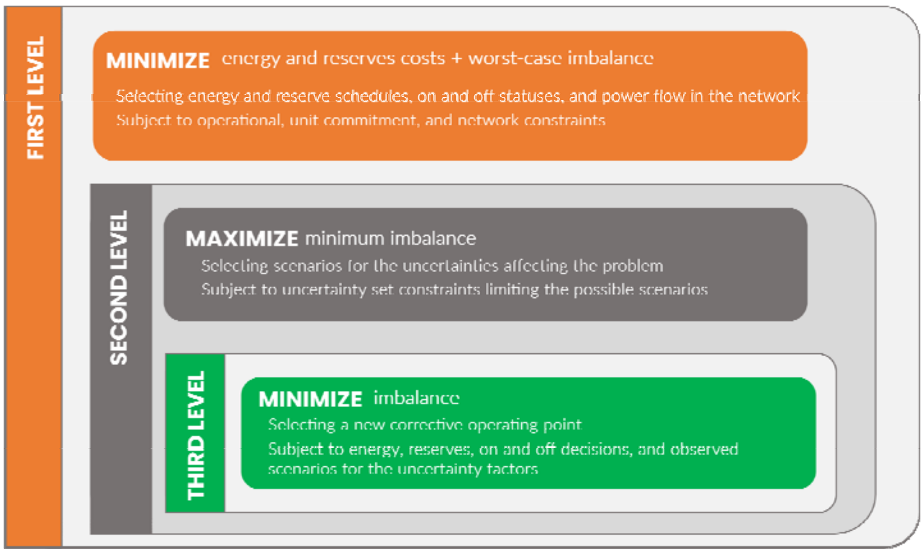

Figure 7. Hierarchically nested structure of two-stage robust generation scheduling

The trilevel format, on the other hand, involves three nested optimization levels. While first-stage decisions are determined in the first level, in the second level, the worst-case scenario is identified according to the third-level best reaction and associated cost. Thus, the worst-case scenario identification is a bilevel model itself, which considers, for each feasible scenario (defined by the uncertainty set constraints), the best real-time reaction (e.g., redispatch to minimize the imbalance cost) within the first-stage scheduled energy and reserves. Therefore, in this setting, the second level can be viewed as an oracle that, for each schedule proposed by the first level, returns the worst-case scenario considering the best reaction in terms of real-time operation. It is worth highlighting that the second level is flexible enough to consider scenarios associated with multiple sources of uncertainty, such as renewable generation, demand, and outages. Figure 7 depicts the hierarchically nested structure of two-stage robust generation scheduling. It is relevant to emphasize that this structure is not used in practice but serves as a modeling tool to let the modeler control the level of information considered in each variable and constraint, i.e., which operational decisions belong to the first stage (first level) and the second stage (third level). Additionally, it also allows us to mathematically describe the uncertainty set, its characteristics, such as dependencies among variables, how they affect the constraints, and its conservativeness level. Thus, the trilevel format is the preferred scheme for the modelers and those that will discuss the conceptual details of the scheduling problem, whereas a relaxed version of the monolithic enumerated format is the way through which the computational and algorithmic team will actually solve the problem. It is important to say, however, that the two lowermost levels also play a key role when solving the problem, as will be described next.

The set of three nested optimization problems illustrated in Figure 7 is typically solved by means of iterative procedures that approximate the second-stage recourse function within the first-stage problem. The first step is to formulate a master problem, which is basically a relaxation of the aforementioned monolithic enumerated model. This master problem accounts for the first-stage cost through a lower approximation of the recourse function output, as not all possible scenarios are considered. Figure 8 shows a popular manner of implementing this approximation, namely the so-called column-and-constraint generation (CCG) technique, which is particularly tailored to the resulting instances of trilevel programming.

Figure 8. CCG technique

This technique involves the iterative identification of scenarios whose corresponding constraints are included in the master problem, a relaxed version of the monolithic formulation initialized with relevant easy-to-obtain specific scenarios. More specifically, at each iteration, once a first-stage solution is obtained, an oracle identifies which of the scenarios considered in the uncertainty set is the most violated one under such a solution. Finite convergence to optimality is guaranteed if both the master problem and the oracle are solved to optimality at each iteration. As a main takeaway, robust models, although seemingly complex, are easier to solve than they appear. Note also that the trilevel formulation offers a theoretical framework that brings relevant insights, whereas practical solution methods for two-stage multiperiod problems –the class of problems where most unit commitment and dispatch models lie in– rely on simple iterative procedures as described above and illustrated in Figure 8.

There are different ways to formulate and solve the oracle. For instance, the two lowermost levels (second and third levels) are typically converted to a single-level MILP equivalent through the application of either Karush-Kuhn-Tucker optimality conditions or duality theory to the third level and state-of-the-art linearization techniques for the resulting bilinear products of variables. In this case, the oracle would be a MILP problem that would search into the space of scenarios through off-the-shelf MILP solvers for the one with the worst-case reaction cost. Nonetheless, in specific cases, enumerative or more sophisticated machine-learning oracles can be built. Thus, once the oracle identifies the most violated scenario, its respective block of constraints is included within the master problem, and the next iteration starts. This process is repeated until the second-stage function is sufficiently well approximated in the master problem.

Box 3. Appealing features of robust generation scheduling

- ✓ **Uncertainty awareness** – Generation scheduling is intrinsically a decision-making process under uncertainty. As widely discussed before, uncertainty plays a key role in look-ahead-based policies and opportunity cost assessment. So, it is widely accepted, almost by the very definition of the scheduling process, that generation scheduling must consider first-stage solutions that are aware of the lack of perfect information, thereby guarding against possible adverse conditions. This fact has been widely demonstrated in the academic literature yet has not been completely internalized in industry practices, which still mainly rely on deterministic approaches. Additionally, it is relevant to emphasize that uncertainty-aware pricing schemes can be readily derived from RO models as per recent academic publications on the subject.
- ✓ **Reproducibility and compliance** – In electricity markets, uniqueness and non-discrimination are crucial when it comes to determining prices, producers' revenues, and consumers' payments. Thus, in short-term generation scheduling, there is no room for sample-dependent solutions, such as those based on the SAA approach. Traditional SAA-based stochastic models fail in compliance and reproducibility requirements by allowing different outcomes, with different beneficiaries, depending on the sample considered. By contrast, robust scheduling models are deterministic, in the sense that the system operator will not rely on a solution that depends on randomly selected samples.
- ✓ **Computational tractability** – Compared to the benchmark modeling approach based on classical SAA stochastic models, RO scheduling models relying on the efficient implementation of decomposition algorithms can leverage the previous day's discovered umbrella set of scenarios to quickly update uncertainty sets and find optimal solutions. Additionally, the classical approach of implicitly representing all network constraints via cutting-plane methods, if implemented based on strong (worst-case) cuts, conceptually matches CCG algorithms. Therefore, energy and reserve scheduling for large-scale power systems (with a few thousand buses) endogenously defining n-1 or n-2 compliant reserves with ensured deliverability across the network is not science fiction for current computational capabilities. If, on top of the aforementioned ideas, we add readily available techniques such as 1) high-performance and parallel cloud computing, 2) state-of-the-art machine-learning methods to keep track of the umbrella set of scenarios, and 3) data-driven polynomial-time-based uncertainty sets to describe renewables and load variability, it is fair to say that all the technology needed for a real-world scheduling of energy and reserves is available.
- ✓ **Flexibility and easy interpretability** – Robust scheduling models are flexible to account for a wide range of uncertainty descriptions and also provide interesting interpretations of their uncertainty-aware solutions and prices. For instance, the n-K criterion provides an easy-to-convey explanation and interpretation for how system components' availability uncertainty is considered. Additionally, stress tests constitute relevant tools used in industry practices to test solutions under adverse conditions when uncertainties exhibit complex dependencies of difficult characterization. In these tests, a given solution is submitted and tested against real adverse conditions selected from historical data. The flexibility of robust models allows for solutions that can endogenously account for this approach through data-driven uncertainty sets (DDUS). The solution of robust models based on DDUS can be understood as the least-cost solution that passes the stress test. Figure 9 illustrates the process of building DDUS.

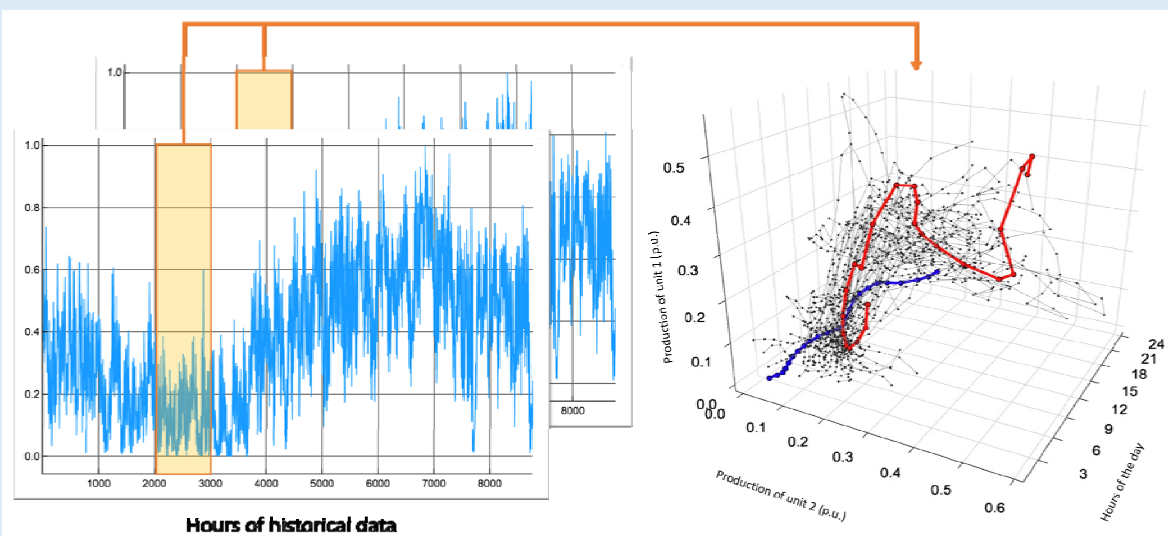

Figure 9. Data-driven uncertainty sets. On the left-hand side, two time series associated with different wind power plants and a time window of approximately one month of data for a given time period. On the right-hand side, the set of 2-unit × 24-hour paths within the same time window. The DDUS is built based on the convex combination of these 48-dimensional points to preserve spatial and temporal dependencies. In blue is the path with the least amount of total (both units) wind power generation within 24 hours, and in red is the path summing up the maximum total generation

Different variants of the CCG approach have been reported and many clever ideas have been proposed to enhance its computational efficiency. Additionally, a robust generation scheduling problem accounting for multistage real-time operation under wind power output uncertainty can be formulated as a trilevel problem in which only the first level takes into account time-coupling ramping constraints. As a result, second-stage problems can be decoupled time-wise and solved in parallel, which greatly improves the computational performance of a methodology that can be used for electricity market-clearing purposes. Furthermore, the use of decision rules to account for the non-anticipative aspect of the decision-making process can also generate multistage robust policies through two-stage models.

Overall, robust generation scheduling models gather most of the desired properties that both industry and academics value, as summarized in Box 3. From 2011, when RO-based scheduling problems were first published in IEEE journals, until now, only a few practical initiatives have used RO. The picture is changing as RO ideas are being disseminated. For instance, the New England ISO in the U.S. is using RO in the context of the "Do Not Exceed (DNE) Dispatch Project", which aims to define limits for the utilization of hydro and renewable intermittent resources. Further details can be found in the rules and procedures document CROP.34013 at https://www.iso-ne.com/participate/support/participant-readiness-outlook/do-not-exceed-dispatch#technical-documentation.

4. Conclusions

This article provides four main takeaways for the power industry. First, practical day-ahead generation scheduling models are typically instances of two-stage decision-making problems, where the second-stage decisions play the role of look-ahead assessment to account for the system's opportunity costs. Secondly, uncertainty is a key part of look-ahead assessments in generation scheduling. In this context, we highlight that uncertainty-aware scheduling decisions are crucial for addressing relevant challenges brought by the massive integration of renewables. Thirdly, robustness has been customarily albeit possibly inadvertently considered as a driving factor in power system operation as exemplified by the wide adoption in industry practice of a worst-case setting in the form of the well-known deterministic n-1 and n-2 security criteria. Therefore, broadening the incorporation of robustness to handle the requirements of generation scheduling in the current restructured, increasingly decarbonized, and vulnerable power industry is an effective and practical alternative. Finally, RO can also be viewed as an alternative and simpler way to account for uncertainty-aware solutions in modern scheduling models without the need to fully describe uncertainty in a complex probabilistic manner within the optimization problem. In this sense, RO provides an interesting computationally tractable alternative to stochastic approaches that enjoys relevant features justifying further consideration from system operators.

For Further Reading

- D. Bertsimas, E. Litvinov, X. A. Sun, J. Zhao, and T. Zheng, "Adaptive robust optimization for the security constrained unit commitment problem," *IEEE Trans. Power Syst.*, vol. 28, no. 1, pp. 52–63, Feb. 2013.
- B. Carlson, Y. Chen, M. Hong, R. Jones, K. Larson, X. Ma, P. Nieuwesteeg, H. Song, K. Sperry, M. Tackett, D. Taylor, J. Wan, and E. Zak, "MISO unlocks billions in savings through the application

of operations research for energy and ancillary services markets," *Interfaces*, vol. 42, no. 1, pp. 58–73, Jan.-Feb. 2012.

- Y. Chen, P. Gribik, and J. Gardner, "Incorporating post zonal reserve deployment transmission constraints into energy and ancillary service co-optimization," *IEEE Trans. Power Syst*., vol. 29, no. 2, pp. 537–549, Mar. 2014.
- N. G. Cobos, J. M. Arroyo, N. Alguacil, and A. Street, "Network-constrained unit commitment under significant wind penetration: A multistage robust approach with non-fixed recourse," *Appl. Energy*, vol. 232, pp. 489–503, Dec. 2018.
- B. F. Hobbs, M. H. Rothkopf, R. P. O'Neill, and H.-P. Chao, Eds., *The Next Generation of Electric Power Unit Commitment Models*. Boston, MA: Kluwer Academic Publishers, 2001.
- A. Street, F. Oliveira, and J. M. Arroyo, "Contingency-constrained unit commitment with n-K security criterion: A robust optimization approach," *IEEE Trans. Power Syst.*, vol. 26, no. 3, pp. 1581–1590, Aug. 2011.

Acknowledgements

The opinions expressed here are the authors' own and do not necessarily reflect the views of the institutions they work with.

This work was supported in part by grants PID2021-126566OB-I00 and PID2021-122579OB-I00, funded by MICIU/AEI/10.13039/501100011033 and ERDF/EU; by grant SBPLY/21/180501/000154, funded by the Junta de Comunidades de Castilla-La Mancha, Spain, the Spanish Ministry of Finance and Civil Service, the European Union Funds, and the ERDF; by grant 2022-GRIN-34074, funded by the Universidad de Castilla-La Mancha, Spain, under the UCLM Research Group Program, and by the European Commission, under the ERDF; by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Brazil; and by Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ), Brazil.

Biographies

Alexandre Street is with Pontifícia Universidade Católica do Rio de Janeiro – PUC-Rio, Brazil.

Alexandre Moreira is with Lawrence Berkeley National Laboratory, USA.

José M. Arroyo and Natalia Alguacil are with Universidad de Castilla-La Mancha, Spain.

Luiz Barroso is with PSR Energy Consulting and Analytics, Brazil, and Instituto de Investigación Tecnológica, Universidad Pontifica Comillas, Spain.

Citation

A. Street, A. Moreira, J. M. Arroyo, N. Alguacil and L. Barroso, "Robustness: The Missing Ingredient in Generation Scheduling," in IEEE Power and Energy Magazine, vol. 23, no. 3, pp. 31-43, May-June 2025.