Risk-Aware Data-Driven Cross-Market Trading for Renewable Virtual Power Plants: A Survey

Sathish Shankar[1],*, Karunakaran Patchamuthu[2]

[1]University of Technology and Applied Sciences, Department of Engineering and Technology, Shinas, Sultanate of Oman.

[2]Bahrain Polytechnic, Structural and Environmental Engineering, Bahrain.

E-Mail: satish.shankar@utas.edu.om

Abstract- Renewable-rich virtual power plants (VPPs) are now regularly trading energy, reserve, balancing, and carbon-linked products in coupled markets. Their choices are hard to make as prices and renewable generation vary together, while flexible demand and settlement penalties change concurrently and can potentially fall outside the data used to train the model. This survey summarizes studies on cross-market trading that are based on data and consider risk factors from recent literatures. It categorizes approaches as forecast-then-optimize, stochastic and distributionally robust optimization (DRO), strategic game models, reinforcement learning, and hybrid safety-shielded learning. Based on the synthesis, stochastic and robust models are implementation-ready, whereas deep reinforcement learning (DRL) is supported only by simulation. The main research gap is settlement-consistent risk calibration that accounts for market regimes. An evaluation ladder and research agenda are suggested from a deployment perspective.

Keywords- Virtual power plant; cross-market trading; data-driven optimization; distributionally robust optimization; reinforcement learning; settlement risk; market regime shift.

I. INTRODUCTION

Virtual power plants (VPPs) aggregate the power generation and storage capabilities of distributed solar and wind power, as well as batteries, electric vehicles (EVs), controllable generators, and flexible loads, into a single market participant. Aggregation is no longer merely a scheduling tool. It represents a trading entity in wholesale markets for short-term flexibility and energy. A renewable-rich VPP can include day-ahead energy offers, intraday/real-time offers, reserves for ancillary services, carbon management, and renewable certificate management. There are general articles on VPP planning and operation [1] and on optimization techniques [2], but with the recent evolution of the trading problem, with the increased integration of products, gate closuress and settlement stages, it has evolved. This topic is topical due to three developments. The first case concerns data availability: smart meters, market feeds, weather services and asset telemetry are sufficient to come up with estimates of the uncertainty of joint price, renewables and flexibility. Second, a high penetration of renewables leads to steeper ramps, negative prices, scarcity events and larger exposure to imbalances. Third, the aggregators are getting into markets with a different settlement period for reserve, balancing, flexibility and low-carbon solutions. So, an apparently attractive bid in the day-ahead market could use storage headroom required for reserve delivery or have a real-time imbalance or alter a carbon obligation.

Single-market optimization ignores these opportunity costs, and a purely predictive approach will not necessarily result in a physically feasible portfolio. The review gap is specific. A bidding review classifies resources and optimization tools recently carried out [3] and a spot market bidding review with deep-reinforcement-learning was carried out in 2026 [4]. Both studies are important, yet neither frames the literature on the intersection of uncertainty due to data, cross-

Journal of Intelligent Engineering and Informatics, Volume 1, Issue 1, Pages 14-27
Corresponding author. E-mail address: satish.shankar@utas.edu.om
Received 9 July 2026; **revised** 10 August 2026; **accepted** 19 August 2026; **published** 31 August 2026.

product settlement, physical feasibility, regime shift, and evidence realism. Furthermore, in the existing reviews, mathematical optimization and learning are discussed separately. Separation conceals an essential design choice as to what data are used for – forecasting inputs, creating scenarios or ambiguity sets, learning a bidding policy or adapting a constrained optimizer? All these choices are not commonly compared along the settlement chain and deployment evidence review axes by an identified review, according to this research.

Retained studies were those that made financial commitments in the form of market decisions for an aggregated VPP and disclosed a repeatable method to optimize, learn or evaluate the VPP. Thirty journal papers were eventually chosen, and each reference was verified against its digital object identifier (DOI) metadata. There is no chronological arrangement of the main taxonomy. It differentiates between forecast-then-optimize and multistage stochastic models, robust, chance-constrained and distributionally robust models, price-maker, bi-level and game-theoretic models, single-agent, multi-agent, and federated reinforcement learning, as well as hybrid approaches in which learned distributions or policies are complemented by explicit feasibility protection. The same questions are asked of each family and the responses used as a basis for comparison: Which markets are coupled? What is the process by which uncertainty is learned? At what cost and where are the physical and market rules applied? What are the keys to successful performance? Is there any way for an operator to explain and overturn a decision? This analysis is different from a typical review, as it presents the road from observation to settlement.

This path sheds light on how a mistake in one module gets to later markets and helps to judge deployment. It is not correct that a model is considered practical simply if it gives a quick solution. It is correct to say that a model is considered practical if it uses available information, respects the timing of the gates, respects headroom, settles deviations correctly, and resists a shift not encountered during tuning. The resulting structure will be of interest to researchers developing algorithms for autonomous bidding, to VPP operators deciding on choice of decision support and to regulators to determine if an autonomous bidding process can be monitored and safely interrupted. Settlement-consistent risk calibration under market regime shifts is the central unresolved challenge. Historical data do not necessarily include future increases or decreases in the price ceiling, weather-related events, entry by participants, a change in the imbalance tariff, and a coronavirus-related renewable-energy shortfall. Consequently, a model can be statistically correct and economically unsound at the same time.

This survey adds an exact scope, a cross-market taxonomy, a unified mathematical view, a 26-study evidence matrix, and a deployment ladder that separates a historical replay with a closed-loop simulation and field operation. Future directions include safe offline learning, a common settlement benchmark, adaptive ambiguity sets, uncertainty calibration, privacy-preserving coordination and auditable human control. As a whole these elements link the abstract problem, the method families, the evidence, and the research agenda.

II. SURVEY SCOPE AND REVIEW METHOD

A. RESEARCH QUESTIONS AND EXACT BOUNDARIES

Three questions are asked in the review. First, how are data-driven VPP trading methods translated into orders based on forecasts, samples and market feedback? Secondly, how are they applied to coupling of energy, balancing, reserve, carbon and certificate settlements? Third, which reported methods are applicable to a controlled implementation, and which are still demonstrated within a simulator? These questions ensure that the discussion is driven towards decisions that result in financial and delivery obligations. It includes a VPP when at least two classes of distributed resources are merged and a portfolio is submitted, scheduled or priced on their behalf by a coordinating entity. Flexible buildings, demand response, renewable generation, storage and controllable generation can be included in the resource set. A study that simultaneously determines a minimum of two market settlement stages/products, e.g. day-ahead and intraday energy, energy and reserve, electricity and carbon, is classified as a cross-market study. Single-market papers are kept only if they provide information or risk approach that is relevant and cross-sections. The term data-driven is used strictly. It encompasses learned forecasts, empirical scenarios, sample-based approximation or Wasserstein ambiguity sets, value-function approximation, supervised bid construction, and reinforcement-learning policies. Use of deterministic optimization with a single-point forecast is not stand-alone learning but forecast-then-optimize. Risk-aware refers to a measure or constraint on downside risk based on situations or chance constraints, robust sets, distributional ambiguity, conditional value-at-risk (CVaR), regret, reliability limits, or safety projection. The survey does not include peer-to-peer local trading without a VPP aggregator, unit commitment without aggregation, purely physical voltage or frequency control or retail tariff design without wholesale exposure and network market clearing that does not resolve a VPP order. These literatures provide a useful way of ensuring that a broad smart-grid review is not interpreted as a trading survey.

B. LITERATURE SEARCH AND VALIDATION

Market stages, products, uncertainty methods and learning methods were combined with the VPP term in query strings. Through citation tracing, studies of older foundations in bidding and recent papers that did not mention data-driven in their titles but indirectly referred to their learning from observations either distributions, policies or bidding packages were identified. An aggregated VPP or a sufficiently similar hybrid plant, an explicit market bid, offer, dispatch commitment, or clearing response were required to be considered inclusion; the presence of an operating region that said treatment of uncertainty, data, and/or strategic behavior were an additional inclusion criterion; and a sufficient methodological detail was required to classify the study. Four review papers [1]–[4] and 26 primary journal studies [5]–[30] constitute the final set. Each primary study is coded, however, in five categories: decision family, market coupling, whether there is an uncertainty or risk mechanism, constraint location, and level of evidence. This coding facilitates comparisons of results between different methods using varying test systems and parameters of profit. It also uncovers a hidden failure to evaluate apparently more evolved algorithms with less realistic market rules, when compared with simpler optimized models.

C. POSITION RELATIVE TO EXISTING REVIEWS

The actual differences are detailed in Table 1. Broad reviews [1], [2] provide fundamentals of VPP resources, control and scheduling, optimization. The risk review [3] focuses on downside protection of different resources types, while the DRL review [4] focuses on learned bidding in spot-market settings. The present survey connects those strands via cross-market settlement and an evidence-realism test. This means it is not something new that bidding or risk or DRL has not been reviewed, this is the first identified synthesis that focuses on their joint deployment problem.

TABLE 1. Coverage of existing reviews and the exact gap addressed here.

Review	Main coverage	Data-driven decision scope	Cross-market settlement and evidence gap
Ullah et al. [1]	Planning, operation, scheduling, renewable uncertainty	Optimization methods broadly	Does not organize trading methods by coupled settlements or deployment evidence
Nafkha-Tayari et al. [2]	VPP concepts, services, control, optimization, practical cases	Broad operational coverage	Trading, risk, and learning are not unified through one settlement model
Xiao [3]	Risk-averse VPP bidding by resource type	Stochastic, robust, DRO, game, and AI methods	Risk focus; limited common comparison of data use, regime shift, and evidence levels
Mo et al. [4]	DRL bidding in electricity spot and competitive markets	Single- and multi-agent DRL	Does not synthesize all optimization families across carbon/certificate and product settlements
This survey	Data-driven, risk-aware cross-market VPP trading	Forecasts, scenarios, DRO, games, RL, federated and hybrid methods	Settlement-centered taxonomy, regime-shift challenge, evidence ladder, and implementation test

D. CODING AND SYNTHESIS RULES

The first of the two passes was used to gather information regarding VPP resources, VPP role in market, VPP products, steps of settlement, source of uncertainty, goal, and the method to solve it. The second pass reported whether constraint checking had been performed within the decision model, after the decision making was performed, or just by means of the reward penalties alone. Evidence was then put on the multi-level ladder introduced in Section V: a price-maker DRL paper is attributed to both strategic and learning families, a Wasserstein model is attributed to robust, and can also be cross-market. The pooled percentage improvement is not provided. A numerical meta-analysis would give false comparability using different currencies, time horizons, resource portfolios, baselines and penalty rules for profit, cost and reliability. Instead, the synthesis makes comparisons between processes and the strength of the test. Any claimed improvement is not considered within-study evidence if the paper is utilizing the same data, the same settlements, and the same budget for risk. Instead, there are currently more weights for out-of-sample testing than in-sample optimization, and likewise chronological replay compared to randomly sampled scenarios. The review consequently distinguishes between bibliographic validity and deployment credibility.

III. CROSS-MARKET TRADING MODEL AND NOVEL CHALLENGE

A. TRADING TIMELINE AND RESOURCE COUPLING

The survey model is organized in Figure 1. The four input categories are market prices and rules, renewable and asset states, customer flexibility, and low-carbon instruments. These inputs can be used with any method family. A common risk/feasibility shield is then consulted for the proposed portfolio and orders are submitted to the day-ahead, intraday, real-time, reserve, carbon, or renewable-certificate markets. Settlement outcomes include profit, imbalance, violations, and evidence of regime change, which are used in the subsequent update. An editable source diagram is also included to allow labels and connectors to be edited in PowerPoint. Timing is important because information is revealed between the gates. The day-ahead order selection is based on uncertain prices, renewable production, demand response, and availability of equipment. Intraday trades have less liquidity and shorter decision times but correct earlier forecasts. Reserve awards require capacity that is not committed elsewhere and can lead to obligations for activation. Real-time deviations are settled using single- or dual-price mechanisms. Although carbon and certificates may be settled over extended time periods, the shadow value impacts the preferred dispatch. A good model should not separate these or add market profits that have not been optimized to the model.

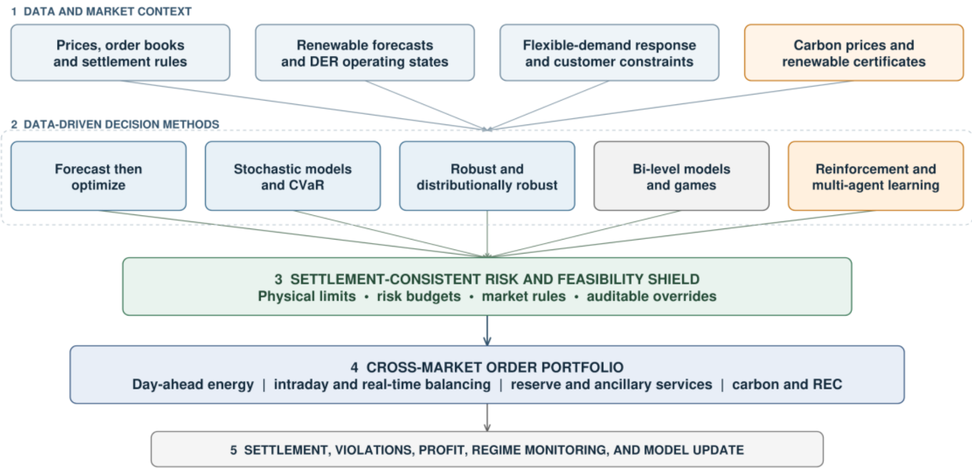

Figure 1. Taxonomy and decision flow for risk-aware data-driven cross-market VPP trading

B. COMMON MATHEMATICAL STRUCTURE

Let *m* index market products and *t* index decision intervals. The action vector a has quantities plus price-quantity blocks, if a price-maker. The prices, renewable output, demand response, activations, outages and penalties are stored in the random vector x_i. The small portfolio profit is written below. These operating, degradation, and imbalance terms prevent gross market revenue from being confused with net value.

$$\Pi(\mathbf{a},\xi)=\sum_{t\square\mathrm{T}}\left(\sum_{m\square\mathrm{M}}\lambda_{m,t}\,q_{m,t}-C_t^{\mathrm{op}}-C_t^{\mathrm{deg}}-C_t^{\mathrm{imb}}\right) \tag{1}$$

Market volumes must be supported by actual delivery. A convenient distinction between net grid injection and the day-ahead and intraday amounts is established: Any mismatch is settled in real time. Reserve activation can be added to the same balance. This poses a common mismatch of the learning studies: a reward can penalize imbalance without precluding an infeasible storage or demand action.

$$p_t^{\mathrm{grid}}=p_t^{\mathrm{ren}}+p_t^{\mathrm{dg}}+p_t^{\mathrm{dis}}-p_t^{\mathrm{ch}}-d_t+\Delta d_t,\quad \delta_t=p_t^{\mathrm{grid}}-q_t^{\mathrm{DA}}-q_t^{\mathrm{ID}} \tag{2}$$

Storage introduces intertemporal linkages. The state of energy needs to account for charge and discharge efficiency, power limits, terminal requirements and degradation. However, the intertemporal constraints on demand are similar to those of flexible demands for rebound, satisfaction and cumulative service. These are the primary reasons for using at least a fast policy, in conjunction with, or as trained by, a feasibility mechanism.

$$e_{t+1}=e_t+\eta^{\mathrm{ch}}p_t^{\mathrm{ch}}\Delta t-\frac{p_t^{\mathrm{dis}}}{\eta^{\mathrm{dis}}}\Delta t,\quad \underline{e}\le e_t\le\overline{e} \tag{3}$$

Risk measures are used to account for risk awareness, typically by adjusting expected profit. CVaR regulates the mean loss in the worst tail beyond confidence level alpha and the coefficient beta indicates the level of risk aversion. Instead, chance constraints can be used to bound the probability of delivery failure. Both require a distribution or scenarios and may be poorly calibrated after a market regime changes.

$$\max_{\mathbf{a}\in\mathrm{A}}\ \mathrm{E_P}[\Pi]-\beta\mathrm{CVaR}_\alpha(-\Pi),\quad \mathrm{CVaR}_\alpha(L)=\min_\zeta\left(\zeta+\frac{\mathrm{E}[(L-\zeta)_+]}{1-\alpha}\right) \tag{4}$$

Distributionally robust optimization (DRO) protects against uncertainty in the estimated distribution itself. The ambiguity set P_N(epsilon) is built from N observations and a radius epsilon. Small radii behave like empirical stochastic optimization; large radii become conservative. The practical question is not only how to solve the inner worst-case problem, but how to select and update epsilon using out-of-sample settlement losses.

$$\max_{\mathbf{a}\in\mathrm{A}}\ \min_{\mathrm{Q}\in\mathrm{P}_N(\varepsilon)}\ \mathrm{E_Q}[\Pi(\mathbf{a},\xi)] \tag{5}$$

A learning policy can propose continuous or discrete orders from a state containing forecasts, awards, asset states, and time. For deployment, the proposal should be projected onto the state-dependent feasible action set before submission.

$$\mathbf{a}_t^{\mathrm{safe}}=\arg\min_{\mathbf{a}\in\mathrm{A}(s_t)}\|\mathbf{a}-\pi_\theta(s_t)\|_2^2 \tag{6}$$

This safety projection preserves rapid inference while keeping market and physical obligations explicit. It also gives an operator a record of when and why the learned action was changed.

C. INNOVATIVE CHALLENGE

This novel challenge tested by this survey is settlement-consistent risk calibration in the presence of a changing joint market distribution. This could lead to a shortage in energy supply due to the weather and occur at the same time as a reserve activation, poor performance of renewable energy generation, lack of customer responsiveness and a high imbalance price.

Training market-by-market models underestimates this joint tail. As an alternative, an extremely worst-case robust model could be more flexible and risk losing normal-market revenues. A regime-change detection procedure should increase or alter the size of the uncertainty set, not make any cross-market attestation commitments more implausible, and adopt a less conservative policy when evidence suggests that doing so is appropriate. The criterion for the update should be based on financial and operational outcomes: realized tail losses, shortfalls in delivery, calibration error, and constraint interventions. Just predicting the right weather isn't sufficient. This challenge motivates the hybrid architecture depicted in Figure 1 and the future research agenda.

III. CLASSIFICATION OF EXISTING APPROACHES

Table 2 summarizes the five method families before the detailed synthesis. The categories are defined by how data enter the decision and where feasibility and risk are enforced; one study can therefore inform more than one family.

TABLE 2. Method-based taxonomy for cross-market VPP trading.

Method family	Decision logic	Risk and constraints	Practical profile	Representative studies
Forecast then optimize	Point/probabilistic forecasts feed mixed-integer linear programming (MILP) or multistage recourse	Explicit physical constraints; scenario-based risk	Transparent and solver-ready; vulnerable to forecast and scenario misspecification	[5]–[7], [11], [12], [14], [17], [20],
Robust / DRO	Optimize against a set of outcomes or distributions	Worst case, chance constraints, CVaR, Wasserstein or sample ambiguity	Strong downside control; calibration and conservatism are central	[18], [19], [21], [24], [25]
Strategic / game	Embed clearing, rivals, or leader-follower response	Market power and strategic uncertainty; physical constraints remain explicit	Useful for price makers; depends on private or simulated rival behavior	[8], [9], [15], [26]
Learning policy	Learn state-to-order or state-to-dispatch mapping	Reward penalties, constrained reinforcement learning (RL), multi-agent or federated updates	Fast inference and adaptive recourse; mostly simulation-based and hard to audit	[10], [13], [16], [22], [23], [26], [27]
Hybrid shielded learning	Optimizer builds base schedule; policy adapts; shield repairs actions	Calibrated distribution plus explicit feasible action set	Best deployment pathway; needs joint training, settlement replay, and intervention reporting	Synthesis of [19], [22]–[25]

A. FORECASTING AND STOCHASTIC OPTIMIZATION MODELS

The forecast-then-optimize approach first forecasts prices, renewable output, load, and flexibility and then solves an explicit scheduling/bidding problem. Its appeal lies in transparency: every order is linked to a forecast, objective period and a constraint. In previous cross-product models it was already demonstrated that the potential of energy, spinning reserve and reactive-power opportunities should be priced concurrently [5]. The resource models were extended to demand-response bidding [6], intraday exchange [7], electric-vehicle aggregation [12] and to renewable portfolios [14], [17] but in none of these cases did the separation of estimation and decision change. Stochastic programming uses several forecast paths as scenarios. When gate closuress and information revelation is formulated as a scenario tree or as a Markov decision process, it is most powerful. Wozabal and Rameseder [11] developed a multistage program for modeling the Spanish day-ahead market and six intraday auctions; they then presented their simulation results in comparison with the deterministic, restricted-market models and the industry benchmark. That's more informative than an in-sample profit gain since it's looking at different market stages and testing recourse during them. Risk-constrained stochastic bidding also models generator outages [18] and multi-objective stochastic variants that trade operating cost against uncertainty regret are addressed in [21]. The biggest restriction is designing the scenarios. If independence across sampling is applied, the cross-market dependence can be broken, if the joint tree is large, it is computationally challenging. Point forecasts are blind with respect to distributional uncertainty while fixed scenario weights treat the distribution of whatever it is the forecast is for as identical to the mechanics observed during the estimation period. These models are thus ready for implementation when market rules are stationary, portfolio size is medium-sized and rolling forecasts are well calibrated. If rare events and policy changes are the preponderant drivers of losses, they require adaptive scenarios or ambiguity sets.

B. ROBUST OPTIMIZATION MODELS

A decision is made robust by optimization against all scenarios in a specified set. It provides good indication of the delivery guarantee, but may be overly conservative if the set is based only on engineering judgment. Risk-constrained stochastic models [18] and risk-based multi-objective models [21] mitigate this situation and reveal a tunable reliability-cost trade-off. Sample robust optimization [24] treats the observed samples as they are, and the semi-anticipative multistage DRO [25] imposes some limitations in the way future information can influence current decisions. Wasserstein DRO is particularly pertinent to data-rich VPPs where it provides an ambiguity ball to enclose the empirical distribution. This instrument is employed in [19] for dealing with chance constrained electricity-carbon trading. Such studies shift data from the realm of forecast accuracy to the realm of uncertainty itself. Generally compatible with existing solvers, DRO provides an interpretable buffer against sampling errors. The problem with it is its lack of calibration. An ambiguity radius is typically chosen based not upon an out-of-sample coverage goal, but rather by sensitivity analysis. Other papers refer to expected profit with varying uncertainty budgets, and a higher number might just be a result of less protection. Reporting nominal profit, tail loss, violation frequency and the empirical coverage of the uncertainty set are all essential to a proper comparison.

C. HIERARCHICAL DECISION MODELS

The VPP is assumed to be a price-taker. This makes sense for a little aggregator, but not so much for a large aggregation or market where there isn't that much reserve. Bi-level models put market clearing below the VPP's offer decision. Strategic bidding and multiple objectives are integrated by Shafiekhani et al. [8] and leader-follower relationships between the VPP and other stakeholders are captured under the Stackelberg model in [9]. Carbon-electricity coupling [15] alters generation mix and bid structure as the marginal cost imposed on high-carbon resources. Although game models are more directly indicative of strategic interaction, they require competitor models, competitor network information and clearing assumptions that an aggregator may not hold. Of course, math programs sometimes have equilibrium constraints which could be slow or sensitive to linearization. The multi-agent twin-delayed deep deterministic policy gradient (MATD3) price-maker model [26] is trained to adapt competitively but without the necessity to solve the lower-level market at each online decision, however the learned equilibrium cannot be easily separated from the simulator or opponents' policies employed during training. Strategic models can be applied in cases of market-design studies and large VPPs where the VPP clearing rules can be observed. They are not as prepared for standard deployment when costs of the competitors are kept private or when competitors respond. Strategic misspecification should, therefore, also be a part of robustness; not just renewable-generation and price-forecast errors. This is more convincing from repeated back tests with varying opponent policies, than simply converging on one simulated equilibrium.

D. DISTRIBUTED LEARNING APPROACHES

Learning-based trading eases part of the explicit optimization burden. In [10], a data-driven VPP bidding package and a Vickrey–Clarke–Groves (VCG) auction is tested on an IEEE 30 bus system. In [13], deep reinforcement learning is used to learn a mapping from the state of a system to a dispatchable command. Multi-agent learning was used in distinguishing between market stages and participants: day-ahead and real-time bidding in [16] and partially observed multi-agent methods in [23] coordinate network-constrained VPPs. Community VPP agents are able to learn autonomous decisions for ancillary services [27]. Federated learning addresses the problem of data ownership. Multiple VPPs are able to train a common policy, while the strong federated approach in [22] offers some protection against disturbed updates. However, while this is helpful when coordinating EVs and customers, federation alone does not ensure adherence to the market or statistical robustness. Clients may use varying mixes of resources, vary their tariffs, and have varying price regimes; an averaged model could work out badly for some minority VPPs. Reinforcement learning is attractive in situations where bids have to be generated within seconds, and the state-action space is large enough that re-optimization is not possible. It can also discover hard-to-encode patterns of recourse. Exploration in a live market is, however, expensive, degradation or settlement may not be explained by the reward, and it is hard to audit neural policies. The vast majority of the DRL results surveyed are closed-loop simulations. To achieve credible live orders, safe offline training, conservative policy improvement, uncertainty-aware value estimates, and an optimized action shield must be used.

E. CROSS-MARKET AND HYBRID DECISION ARCHITECTURES

Cross-market models differ in the components they connect. Many also incorporate some form of physical sharing between energy and reserve, with various publications exploring this option [5], [20] while others introduce an intraday recourse [7], [11] or flexible demand [6], [12], [15], [19] or electric vehicles [12] or carbon costs [15] or joint electric/carbon risk [19]. The adaptive robust model from 2026, for example, explicitly addresses multi-market uncertainty in its style by directly modeling it with a nonparametric distribution that is learned. These choices are mapped by all the primary studies in Table 3. Hybrid architecture is the most promising one. Forecasting: estimating conditional distributions/regime indicators. Stochastic/DRO optimization builds a coherent base portfolio according to customers' financial requirements. A learned policy provides rapid recourse or improves a residual decision. The safety layer projects the action onto network, resource, reserve, and settlement constraints. This division employs both methods where they are each best: learning for adaptation and optimization for obligations, as well as statistical calibration for risk. A hybrid approach is not merely the process of stringing different modules together with no integration testing. The error in the forecast has to be passed on to the order and settlement; the reward earned by the RL has to correspond to the optimization goal and finally the safety projection also has to make it to training so that there are no systematic interventions. Thus, a common digital settlement engine is more useful than any other standalone profit comparison.

TABLE 3. Comparative evidence matrix for the 26 primary studies.

Ref./year	Method	Market coupling	Data or risk treatment	Evidence and readiness
[5], 2016	Arbitrage optimization	Energy + spinning reserve + reactive power	Scenario/cost tradeoffs	Numerical market case; implementation-oriented formulation
[6], 2018	Technical VPP bidding with DR	Energy + commercial demand response	Uncertain customer response	Simulation/case study; explicit program rules
[7], 2018	Stochastic programming	Day-ahead + intraday DR exchange	Scenarios for market and response uncertainty	Market-rule case study; no field trial
[8], 2019	Bi-level multi-objective	Strategic energy market bidding	Price-maker and objective tradeoffs	Numerical clearing case; model-dependent
[9], 2019	Stackelberg game	Dispatch + market bidding	Leader-follower response	Simulation; requires credible follower model
[10], 2020	Data-driven bid packages + VCG	Real-time auction	Observed capability/cost packages	IEEE 30-bus simulation; network detail, no live market
[11], 2020	Multistage stochastic / SDDP	Spanish day-ahead + six intraday auctions	Historical prices and wind uncertainty	Chronological out-of-sample replay + industry benchmark; strongest Level 2 evidence
[12], 2020	Stochastic VPP bidding	Energy market + large EV fleet	EV availability and renewable uncertainty	Fleet simulation; explicit operating constraints
[13], 2020	Deep reinforcement learning	Economic dispatch / Internet of Energy	Learned policy from environment interaction	Closed-loop simulation; fast inference
[14],	Renewable VPP	Day-ahead energy	Wind, solar, and flexible	Numerical case;

2022	optimization		load uncertainty	transparent schedule
[15], 2022	Integrated market bidding	Electricity + carbon	Carbon price and renewable uncertainty	Numerical coupled-market case
[16], 2022	Multi-agent DRL	Day-ahead + real-time	Sequential policy learning	Market-data simulation; no protected field operation
[17], 2023	Renewable VPP optimization	Day-ahead energy	Renewable and market uncertainty	Numerical case; single primary market
[18], 2023	Two-stage stochastic + risk constraint	Day-ahead bidding	Distributed-generation outage and tail risk	Scenario stress test; solver-ready
[19], 2023	Wasserstein DRO chance constraints	Electricity + carbon	Distributional ambiguity and chance limits	Out-of-sample-style numerical analysis; calibration still model-specific
[20], 2023	Combined-market optimization	Electricity + ancillary services	Dynamic demand-response price and satisfaction	Numerical integrated-market case
[21], 2023	Risk-based multi-objective	Day-ahead energy	Regret/cost tradeoff for renewable energy sources (RES), load, and price	Sensitivity and numerical case
[22], 2023	Robust federated DRL	Multiple VPPs with EVs	Private distributed updates + disturbance filter	Federated simulation; privacy/scalability evidence, not market settlement proof
[23], 2023	Network-constrained multi-agent DRL	Multiple VPP coordination	Partially observed Markov game	Closed-loop network simulation; no live trading
[24], 2024	Sample robust optimization	VPP bidding	Direct sample-based protection	Out-of-sample numerical tests; promising for shadow trial
[25], 2024	Multistage semi-anticipative DRO	Sequential VPP bidding	Distributional ambiguity with information limits	Numerical multistage test; complex but auditable
[26], 2024	Multi-agent TD3	Price-maker day-ahead market	Learned strategic interaction	Agent-based market simulation; opponent dependence
[27], 2024	Multi-agent DRL	Community VPP + ancillary services	Autonomous decentralized policies	Community VPP simulation; no market pilot

V. EVALUATION PROTOCOLS AND EVIDENCE REALISM

A. EVALUATION DIMENSIONS

Table 4 groups the measures needed for a fair evaluation. Expected profit alone is not enough, since methods that take a bigger tail risk and/or more delivery failures can make more money. As a minimum it is important that studies argue about net profit after degradation, imbalance charge, some measure of downside risk, physical and market violation rates, renewable curtailment and decision time. Training techniques should also be sample efficient, have stable policies across sequences of seeds, and frequently evolve some change to an action. A measure of forecast quality must be based on a probabilistic scoring rule, not just on the mean absolute error. The average price error indicated by the presented values, and which is quite low, still cannot capture scarcity intervals that make the profits. Joint calibration of cross-market decisions is required – the combination of observed and model frequencies of prices, renewable output, responses, and activation should agree. The ambiguity set should attain the coverage stated, on a rolling holdout set for DRO.

TABLE 4. Recommended evaluation dimensions and measures.

Dimension	Recommended measure	Reason for inclusion
Financial value	Net profit; regret; transaction, degradation, and imbalance cost	Prevents gross revenue from hiding physical or settlement costs
Downside risk	CVaR; worst-decile loss; loss probability; drawdown	Shows whether a profit gain comes from greater tail exposure
Reliability	Energy shortfall; reserve non-delivery; constraint-violation rate	Links orders to actual delivery obligations
Uncertainty quality	continuous ranked probability score (CRPS)/pinball loss; coverage; joint calibration; regime detection delay	Tests distributions used by stochastic, DRO, and learning methods
Resource impact	Curtailment; battery throughput/degradation; comfort and rebound	Checks whether flexibility is economically and socially sustainable
Computation	End-to-end latency; memory; optimality gap; failed solves	Determines whether decisions fit each market gate
Learning stability	Seeds; sample efficiency; out-of-distribution return; safety interventions	Separates a stable policy from a favorable training run

B. EVIDENCE REALISM

Evidence can be organized on a five-level ladder. Level 0 could be a toy or full synthetic market. At Level 1, a standard network or detailed asset simulator is added, however, prices and rival behavior are generated in-house. At Level 2, historical market data are replayed through chronological, train-validation-test splits, alongside full settlement rules. Level 3 leverages hardware-in-the-loop or live timing and telemetry with a utility-grade digital twin. Level 4 consists of shadow bidding or a controlled market pilot that compares proposed orders with operator decisions without exposing customers to uncontrolled risk. The majority of papers reviewed are either Level 0 or 1. Live market behavior is not tested in the IEEE 30-bus study presented in [10] where network interactions are tested. On the one hand, some studies about multi-agent and federated DRL are done [16] [22] [23] [26] [27] but the opponents, prices and transition rules are simulated. The Spanish multistage study [11] is the closest to level 2, because it employs a real series of day-ahead and intraday auctions, an out-of-sample evaluation and an industry benchmark. Even that evidence is not live bidding any more, it is a replay. Constraints, market products and solver terminate are explicit in optimization models and these are typically more prone to implementation. The models that are plausible candidates for a shadow trial are the models [19], [24], [25] after they are calibrated for ambiguity parameters. While DRL is operationally fast, it should be emphasized that it is still software-based simulation until it passes historical-operation replay, adversarial stress test and protected shadow phase. This distinction is explicit in Table 5.

TABLE 5. Evidence-realism ladder for VPP trading studies.

Level	Evidence	What it establishes	Current literature profile
0	Toy/synthetic prices and assets	Algorithmic feasibility only	Common for early RL and strategic games
1	Standard grid or detailed simulator	Closed-loop constraints and interactions	Common for VCG, multi-agent, networked, and federated studies
2	Chronological historical market replay	Out-of-sample settlement performance	Limited; [11] is the clearest example
3	Hardware-in-the-loop or utility digital twin	Telemetry, latency, and interface behavior	Rare or absent in the reviewed trading set
4	Shadow bidding or controlled market pilot	Operational value without uncontrolled autonomous risk	No publicly auditable multi-market example identified

C. A REPRODUCIBLE BENCHMARK PROTOCOL

A helpful benchmark should provide a market calendar, gate closures, order format, settlement equations, asset parameters and chronological data splits. After the training period, hyperparameter validation is necessary, as well as a test period that includes ordinary and stressed states. It is not adequate to randomly shuffle hourly observations, as this contains information about the season and market-state. The same forecasts or the same raw data, the same portfolio and the same transaction and penalty rules should be imposed on each method. The required baselines include the deterministic forecast-then-optimize, stochastic optimization, a calibrated DRO policy and a simple safe policy. When reporting results, the data should be over several years or rolling windows with error bars. The value of forecasting, risk treatment, strategic modeling, policy learning and the safety shield should be separated in ablations. Runtime should include not only a call to the final solver or neural inference call, but also the scenario generation, model updates, and feasibility repair.

VI. CRITICAL ANALYSIS

Stochastic and distributionally robust optimization are perhaps presently the most viable options for implementation. Their encoding directly includes gate closures, storage dynamics, reserve headroom, imbalance settlement and their decisions are traceable; mature solvers are able to provide feasibility guarantees. Especially compelling is the Spanish multistage study [11] with the chronological market profiling of the market and its comparison to an industry method. The weakness of it is the computational growth when products, resources and scenarios increase in numbers. Robust and DRO models [19], [24], [25] provide greater safeguarding of sparse data and joint uncertainty. They can be used for shadow operation where calibrated ambiguity radii are defined on unobserved settlements. Otherwise, it is a reliability statement and a tuning option with conservatism. For price-making VPPs with believable information about market clearing and competitors, bi-level models and game models can be employed with realism, but otherwise strategic details can lead to false precision. Deep and multi-agent reinforcement learning [13], [16], [22], [23], [26], [27] are capable of making very fast decisions online and are good at coping with sequential feedback. These studies, however, are primarily simulations. Their results depend on the reward completeness, opponent behavior, training seed, and exploration policy. Under an unseen price rule or correlated outage, a high simulated profit is not sufficient for meeting safe delivery. Federated learning does not ensure feasibility; it makes them more private. The best practical design is hybrid, using probabilistic forecasts in a stochastic base schedule or DRO setup, learning providing quick recourse, and a mathematical safety layer guaranteeing physical and settlement constraints. In the literature, it is not possible to make direct comparisons of profit figures, due to different market time horizons, penalties, portfolios of resources, or risk budgets. As such, the amount of out-of-sample tail loss, violation rate, runtime, frequency of intervention, and regime performance, not only percentage profit improvement should be used in deployment claims. Today, the most studied methods are still pre-deployment evidence, and no study has been reviewed that has a publicly auditable live multi-market VPP trial. Priorities should be determined by that evidence gap.

VII. OPEN CHALLENGES AND FUTURE DIRECTIONS

The first challenge is settlement-consistent regime shift risk. Future models also need to account for the interplay among the variables: prices, renewable errors, activation of reserves, demand response, outages, and penalties and recalibrate the risk if there is a change in this joint dependence. Second, there should be no reinforcement learning that does not pass through network, asset and market constraints, and only reinforces learned actions on a logged operator behavior that is known to be safe. Third, ambiguity-set radii and CVaR weights should have data-based selection rules rather than manually performed sensitivity studies based on observed tail coverage. Fourth, the field should have an open cross-market benchmark with chronological energy, balancing, reserve, carbon, weather, and asset data paired with precise rules for gate and settlement. It should include negative-price episodes, scarcity, communication failures and policy-changes. Fifth, a fundamental requirement of a market simulator is to include strategic learning and liquidity, and leaving out these complicating factors would mislead on the difficulty of strategic learning. Sixth, privacy-preserving coordination should include measurement of performance degradation for heterogeneous VPPs and resilience to poisoned updates. Last, but not least, operator explainability must be supported. A useful explanation should identify the forecast, constraint, risk budget, and market opportunity that influenced an order and document all safety interventions. Some sort of shadow bidding, hardware in the loop timing tests, and staged approvals by the regulatory authorities should be done prior to live control. These steps make algorithmic performance accountable and ensure that customers are protected and system reliability is assured during operation.

VIII. CONCLUSION

The transition from single-market scheduling to coupled energy decisions, balancing reserve, carbon and certificate decisions based on data-driven VPP trading is in progress. Mature stochastic programs exist, as do capable distributionally robust models, strategic bi-level formulations and reinforcement-learning policies now available in the literature. However, as algorithms become more complicated, evidence quality does not necessarily improve. Explicit stochastic and robust optimization is closest to controlled implementation; most deep and multi-agent learning results are still simulator dependent. This survey brings a settlement-centric approach, a five-family taxonomy, a common risk formulation, a comparative evidence matrix and a deployment ladder. Its finding is that regime-shift risks should be calibrated against actual cross-market settlements. Common chronological benchmarks, safe hybrid policies, auditable feasibility shields, and shadow-market validation are prerequisites to autonomous live bidding at scale.

REFERENCES

[1] Z. Ullah, G. Mokryani, F. Campean, and Y. F. Hu, "Comprehensive review of VPPs planning, operation and scheduling considering the uncertainties related to renewable energy sources," IET Energy Syst. Integr., vol. 1, no. 3, pp. 147-157, 2019, doi: 10.1049/iet-esi.2018.0041.

[2] W. Nafkha-Tayari, S. Ben Elghali, E. Heydarian-Forushani, and M. Benbouzid, "Virtual power plants optimization issue: A comprehensive review on methods, solutions, and prospects," Energies, vol. 15, no. 10, Art. no. 3607, 2022, doi: 10.3390/en15103607.

[3] D. Xiao, "A review on risk-averse bidding strategies for virtual power plants with uncertainties: Resources, technologies, and future pathways," Technologies, vol. 13, no. 11, Art. no. 488, 2025, doi: 10.3390/technologies13110488.

[4] P. Mo, K. Li, Y. Lu, Y. Wen, and K. Huang, "Deep reinforcement learning-based bidding strategies for virtual power plants in electricity spot markets: A review," IEEE Access, vol. 14, pp. 50628-50643, 2026, doi: 10.1109/ACCESS.2026.3678911.

[5] H. Nezamabadi and M. S. Nazar, "Arbitrage strategy of virtual power plants in energy, spinning reserve and reactive power markets," IET Gener. Transm. Distrib., vol. 10, no. 3, pp. 750-763, 2016, doi: 10.1049/iet-gtd.2015.0402.

[6] N. Pourghaderi, M. Fotuhi-Firuzabad, M. Moeini-Aghtaie, and M. Kabirifar, "Commercial demand response programs in bidding of a technical virtual power plant," IEEE Trans. Ind. Informat., vol. 14, no. 11, pp. 5100-5111, 2018, doi: 10.1109/TII.2018.2828039.

[7] H. T. Nguyen, L. B. Le, and Z. Wang, "A bidding strategy for virtual power plants with the intraday demand response exchange market using the stochastic programming," IEEE Trans. Ind. Appl., vol. 54, no. 4, pp. 3044-3055, 2018, doi: 10.1109/TIA.2018.2828379.

[8] M. Shafiekhani, A. Badri, M. Shafie-khah, and J. P. S. Catalao, "Strategic bidding of virtual power plant in energy markets: A bi-level multi-objective approach," Int. J. Electr. Power Energy Syst., vol. 113, pp. 208-219, 2019, doi: 10.1016/j.ijepes.2019.05.023.

[9] H. Wu, X. Liu, B. Ye, and B. Xu, "Optimal dispatch and bidding strategy of a virtual power plant based on a Stackelberg game," IET Gener. Transm. Distrib., vol. 14, no. 4, pp. 552-563, 2020, doi: 10.1049/iet-gtd.2019.0493.

[10] X. Chen, W. Pei, W. Deng, and H. Xiao, "Data-driven virtual power plant bidding package model and its application to virtual VCG auction-based real-time power market," IET Smart Grid, vol. 3, no. 5, pp. 614-625, 2020, doi: 10.1049/iet-stg.2020.0038.

[11] D. Wozabal and G. Rameseder, "Optimal bidding of a virtual power plant on the Spanish day-ahead and intraday market for electricity," Eur. J. Oper. Res., vol. 280, no. 2, pp. 639-655, 2020, doi: 10.1016/j.ejor.2019.07.022.

[12] D. Yang, S. He, M. Wang, and H. Pandzic, "Bidding strategy for virtual power plant considering the large-scale integrations of electric vehicles," IEEE Trans. Ind. Appl., vol. 56, no. 5, pp. 5890-5900, 2020, doi: 10.1109/TIA.2020.2993532.

[13] L. Lin, X. Guan, Y. Peng, N. Wang, S. Maharjan, and T. Ohtsuki, "Deep reinforcement learning for economic dispatch of virtual power plant in Internet of Energy," IEEE Internet Things J., vol. 7, no. 7, pp. 6288-6301, 2020, doi: 10.1109/JIOT.2020.2966232.

[14] M. Shafiekhani, A. Ahmadi, O. Homaee, M. Shafie-khah, and J. P. S. Catalao, "Optimal bidding strategy of a renewable-based virtual power plant including wind and solar units and dispatchable loads," Energy, vol. 239, Art. no. 122379, 2022, doi: 10.1016/j.energy.2021.122379.

[15] X. Liu, "Research on bidding strategy of virtual power plant considering carbon-electricity integrated market mechanism," Int. J. Electr. Power Energy Syst., vol. 137, Art. no. 107891, 2022, doi: 10.1016/j.ijepes.2021.107891.

[16] T. Ochoa, E. Gil, A. Angulo, and C. Valle, "Multi-agent deep reinforcement learning for efficient multi-timescale bidding of a hybrid power plant in day-ahead and real-time markets," Appl. Energy, vol. 317, Art. no. 119067, 2022, doi: 10.1016/j.apenergy.2022.119067.

[17] C. Yang, X. Du, D. Xu, J. Tang, X. Lin, K. Xie, and W. Li, "Optimal bidding strategy of renewable-based virtual power plant in the day-ahead market," Int. J. Electr. Power Energy Syst., vol. 144, Art. no. 108557, 2023, doi: 10.1016/j.ijepes.2022.108557.

[18] F. Ghasemi-Olanlari, M. Moradi-Sepahvand, and T. Amraee, "Two-stage risk-constrained stochastic optimal bidding strategy of virtual power plant considering distributed generation outage," IET Gener. Transm. Distrib., vol. 17, no. 8, pp. 1884-1901, 2023, doi: 10.1049/gtd2.12826.

[19] Q. Fan and D. Liu, "A Wasserstein-distance-based distributionally robust chance constrained bidding model for virtual power plant considering electricity-carbon trading," IET Renew. Power Gener., vol. 18, no. 3, pp. 545-557, 2024, doi: 10.1049/rpg2.12806.

[20] S. Mei, Q. Tan, Y. Liu, A. Trivedi, and D. Srinivasan, "Optimal bidding strategy for virtual power plant participating in combined electricity and ancillary services market considering dynamic demand response price and integrated consumption satisfaction," Energy, vol. 284, Art. no. 128592, 2023, doi: 10.1016/j.energy.2023.128592.

[21] A. Ghanuni, R. Sharifi, and H. F. Farahani, "A risk-based multi-objective energy scheduling and bidding strategy for a technical virtual power plant," Electr. Power Syst. Res., vol. 220, Art. no. 109344, 2023, doi: 10.1016/j.epsr.2023.109344.

[22] B. Feng, Z. Liu, G. Huang, and C. Guo, "Robust federated deep reinforcement learning for optimal control in multiple virtual power plants with electric vehicles," Appl. Energy, vol. 349, Art. no. 121615, 2023, doi: 10.1016/j.apenergy.2023.121615.

[23] X. Liu, S. Li, and J. Zhu, "Optimal coordination for multiple network-constrained VPPs via multi-agent deep reinforcement learning," IEEE Trans. Smart Grid, vol. 14, no. 4, pp. 3016-3031, 2023, doi: 10.1109/TSG.2022.3225814.

[24] S. Kim and D. G. Choi, "A sample robust optimal bidding model for a virtual power plant," Eur. J. Oper. Res., vol. 316, no. 3, pp. 1101-1113, 2024, doi: 10.1016/j.ejor.2024.03.001.

[25] W. Li, "The virtual power plant bidding strategy model based on multi-stage semi-anticipativity distributionally robust optimization," Electr. Power Syst. Res., vol. 237, Art. no. 111015, 2024, doi: 10.1016/j.epsr.2024.111015.

[26] Y. Jiang, J. Dong, and H. Huang, "Optimal bidding strategy for the price-maker virtual power plant in the day-ahead market based on multi-agent twin delayed deep deterministic policy gradient algorithm," Energy, vol. 306, Art. no. 132388, 2024, doi: 10.1016/j.energy.2024.132388.

[27] X. Li, F. Luo, and C. Li, "Multi-agent deep reinforcement learning-based autonomous decision-making framework for community virtual power plants," Appl. Energy, vol. 360, Art. no. 122813, 2024, doi: 10.1016/j.apenergy.2024.122813.