IEEE INTERNET OF THINGS JOURNAL, VOL. X, NO. X, XXXX

# Scalable and Robust Energy Routing Optimization in Stochastic Vehicular Energy Network

Yao Tang, *Student Member, IEEE*, Wei Liu, *Senior Member, IEEE*, Yunhe Hou, *Fellow, IEEE*, K.T. Chau, *Fellow, IEEE*

1 ***Abstract*—A vehicular energy network (VEN) enables energy transfer by leveraging electric vehicles as mobile carriers through wireless exchange across large geographic areas. This article presents a scalable and robust framework for energy routing in stochastic VENs with the objective of minimizing transmission loss. The problem is formulated as a graph generalized flow optimization, solvable to global optimality via linear programming. To ensure scalability, a flow-guided graph reduction method is proposed, which preserves critical supply-demand connectivity by prioritizing high-impact routes based on vehicular flow patterns. Building upon this, a route-guided time-expanded graph construction strategy is developed to avoid exhaustive temporal replication by generating only time-relevant nodes and arcs along active routes. To address long-horizon stochasticity, a long short-term memory-based model predictive control framework is designed, which captures both randomness and uncertainty via data-driven forecasting and residual-aware robust correction under a rolling-horizon decomposition. The proposed methods are validated on 100 real-world U.S. datasets, demonstrating significant gains in computational efficiency, scalability, and solution robustness across both time-invariant and time-varying VENs.**

***Index Terms*—Electric vehicle, energy routing, graph theory, generalized flow optimization, vehicular energy network.**

## I. Introduction

THE proliferation of Internet of Things (IoT) has accelerated the transition toward intelligent energy solutions [1, 2]. While its architecture promotes bi-directional energy flow and higher renewable integration, practical constraints, such as transmission bottlenecks and deployment limitations, often hinder large-scale implementation [3]. These challenges are particularly evident when transporting renewable energy from remote generation sites to urban load centres [4].

Vehicular energy network (VEN) is a system that enables energy transport across a wide geographical area through electric vehicles (EVs) as mobile energy carriers [5, 6]. Physically, the main components are: i) EVs. The transferable energy is stored in the form of an energy packet, typically a small unit such as 1 kWh, which occupies only a small portion of the EV's battery capacity [7]. It is assumed to be reserved exclusively and not consumed for driving purposes. ii) Road junctions [8]. It is equipped with the wireless power transfer (WPT) device and local energy storage, which support flexible and asynchronous energy transfer. iii) IoT-based control components, supported by advancements in wireless communication [9], enabling real-time data collection, system monitoring, and coordinated energy routing by integrated communication modules, sensing devices, and smart scheduling platform. In the transmission process, EVs wirelessly receive surplus energy at supply junctions, typically located in rural areas with excess renewable generation. Without altering their planned routes, an IoT-based scheduling system determines the optimal time, location, and amount of energy to charge or discharge based on system-level objectives. The energy can optionally be transferred through intermediate relay nodes and is ultimately delivered to demand junctions, which are usually located in urban areas with high energy consumption. As the process requires no driver intervention or additional stops, VEN enables efficient and user-friendly energy delivery. Compared with conventional vehicle-to-grid (V2G) systems [10, 11], VEN shares several similar components, such as EVs and communication systems. Both systems are also concerned with the integration and scheduling of renewable energy resources. However, VEN differs significantly in several core dimensions, including energy transfer paradigm, infrastructure dependency, primary function, operational focus, and EVs' role, as summarized in Table I. This comparison highlights VEN's unique role in renewable energy distribution and grid flexibility.

TABLE I
Comparison Between Conventional V2G System and VEN

| Dimension | Conventional V2G system | VEN |
|---|---|---|
| Energy transfer paradigm | Stationary energy exchange | Mobile energy transport |
| Infrastructure | Local infrastructures | Distributed wireless infrastructure |
| Function | Regional grid support services | Redistribution of renewable energy |
| Operation | Local energy scheduling | Spatiotemporal energy routing |
| EVs' role | Stationary energy storages | Dynamic energy carriers |

Furthermore, the concept of VEN aligns closely with the broader vision of the energy internet (EI), which emphasizes

Manuscript received X, X; revised X, X; accepted X, X. This work was supported by two grants (Project No. 17206222 and T23-701/20-R) from the Hong Kong Research Grants Council, Hong Kong Special Administrative Region, China. (*Corresponding author: K.T. Chau.*)

Yao Tang and Yunhe Hou are with Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong, China. (e-mails: tangyao@eee.hku.hk; yhhou@eee.hku.hk).

Wei Liu and K.T. Chau are with Research Centre for Electric Vehicles and Department of Electrical and Electronic Engineering, The Hong Kong Polytechnic University, Hong Kong, China. (wei.liu@polyu.edu.hk; k.t.chau@polyu.edu.hk).

distributed energy coordination, digital control, and flexible transmission beyond fixed infrastructures [12, 13]. By introducing mobility, VEN complements EI by offering dynamic and spatially adaptive energy routing. This routing capability is made technically feasible by the integration of IoT technologies [14]. By standardized interfaces, each EV and energy routing junction connects to the VEN, allowing for scalable and low-complexity integration. Real-time data, such as vehicle location, battery status, travel path, and local energy demand, can be continuously collected and transmitted through the IoT network to support smart routing decisions [15]. Beyond the conceptual design, recent technological and empirical developments further reinforce the feasibility of VEN. Key enablers include the rapid penetration of EVs [16], advances in dynamic WPT [17, 18], and the increasing availability of compatible infrastructure within urban and regional transportation networks [19]. Empirical evidence also substantiates VEN's practical potential. For example, one UK-based experimental study demonstrated that over 6 MWh of energy could be delivered within 5 hours with only 0.1% EV penetration and 0.1 kWh energy packets [5]. Therefore, these findings affirm that VEN is not only theoretically promising but also technically viable for future energy systems.

Traditional power grids often suffer from significant energy losses, up to 30%, due to long-distance transmission and repeated AC/DC conversions. In contrast, VEN leverages the high charging and discharging efficiency of EV batteries, reducing overall losses to approximately 10% [20, 21]. This efficiency stems from the fact that energy is carried passively during vehicle transit and is only subject to conversion losses at wireless charge-discharge junctions. However, this benefit also introduces a new operational challenge: since losses primarily occur during energy exchange events, the number and location of such interactions directly influence the system's overall efficiency. As a result, energy routing, the process of determining when, where, and how energy is transferred between EVs and infrastructure, becomes a core function in VEN operation. Optimizing energy routing to minimize transfer cycles is therefore essential to unlocking the full potential of VEN as a low-loss, sustainable energy delivery network.

Some advancements have highlighted the critical role of minimizing transmission losses in energy routing. In the context of the EI, graph-based methods have been widely adopted to optimize power flow. For instance, [22] proposes a graph-theoretic routing algorithm within an energy local area network, using weighted shortest-path strategies to achieve efficient source-load matching. [23] extends this direction to multi-source multi-load scenarios in residential microgrids, accounting for congestion and bidirectional flow. In [24], a carbon-emission-aware routing strategy is introduced for regional EIs, achieving a Pareto balance between power loss and environmental impact. While these studies validate the feasibility of graph-driven energy routing, they are largely confined to static, infrastructure-based systems. In contrast, VENs involve energy transfer via moving EVs, where routing decisions must account for real-time mobility, spatial constraints, and temporal dynamics, rendering conventional models insufficient.

To address the unique characteristics of VENs, several studies have extended graph-based optimization frameworks to this context. [25] introduced the concept of a mobile electric grid, functionally similar to VEN, and modeled energy routing via bipartite graphs under traffic congestion constraints. While it demonstrates the feasibility of EV-based energy transfer, the formulation focuses on minimizing hop counts rather than transmission loss and lacks support for supply-demand balancing or time-varying dynamics due to its static graph structure. A more foundational contribution is presented in [26]. The VEN routing problem is formulated as a generalized flow optimization solvable via linear programming (LP). This work constructs all energy paths between a single source-destination pair and highlights the global solvability of the routing task. However, it is limited by its exponential complexity due to full path enumeration and only addresses static scenarios, making it unsuitable for scalable or time-varying applications. To incorporate time-varying behavior, [27] adopts a time-expanded network to jointly optimize energy routing and dynamic storage allocation under time-varying traffic conditions. Although this approach improves delivery capacity in dynamic environments, it emphasizes maximizing energy throughput and does not explicitly model transmission losses. These works underscore the potential of graph-based approaches in VEN routing. Yet, they face challenges such as limited scalability, static formulations, or the absence of energy loss modeling.

A recent study [28] builds directly upon prior works [25-27] and addresses several key limitations. A generalized flow optimization model is formulated that supports multi-source, multi-destination energy transfer with scenario-specific graphs. However, several limitations remain:

- Limited efficiency. Full traversal of all routes and nodes incurs high computational cost in dense networks, while time expansion further introduces redundant structures across temporal layers.
- Insufficient scalability. Long-horizon scenarios cause graph complexity to grow rapidly with time, limiting scalability in large networks.
- Stochasticity-agnostic. The model assumes deterministic inputs, neglecting stochastic variations in supply, demand, and traffic flow.

Efficiency issue can be addressed at the preprocessing stage by graph reduction techniques [29], which aim to simplify network structures while preserving essential properties. Classical approaches include coarsening, condensation, and sparsification [30]. However, in VENs, coarsening is inapplicable because it merges nodes and edges into super-nodes, while each node and edge carries explicit physical meaning that cannot be arbitrarily merged. Condensation is also unsuitable because it replaces the original network with synthetic surrogate graphs that lack physical interpretability. Therefore, sparsification is more appropriate choice as it simplifies networks by selectively removing less critical structures. Nevertheless, conventional methods do not exploit VEN-specific information, hence a task-driven approach is required to enable more effective graph reduction. Next, at solving stage, decomposition-based methods can be utilized for better scalability. A representative example is column

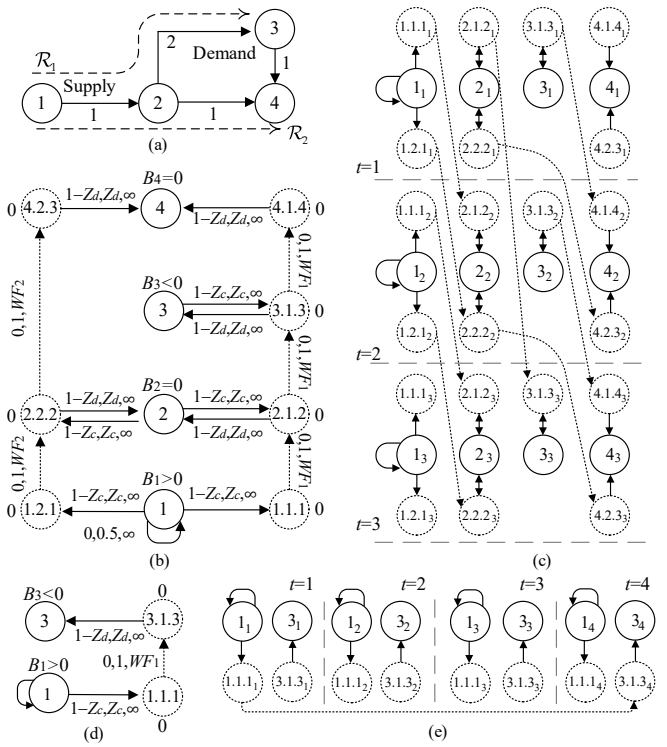


Fig. 1. VENs models. (a) $\mathcal{D}$ in VEN. (b) Time-invariant baseline model. (c) Time-varying baseline model. (d) Time-invariant model with proposed graph reduction. (e) Time-varying model with proposed graph reduction.

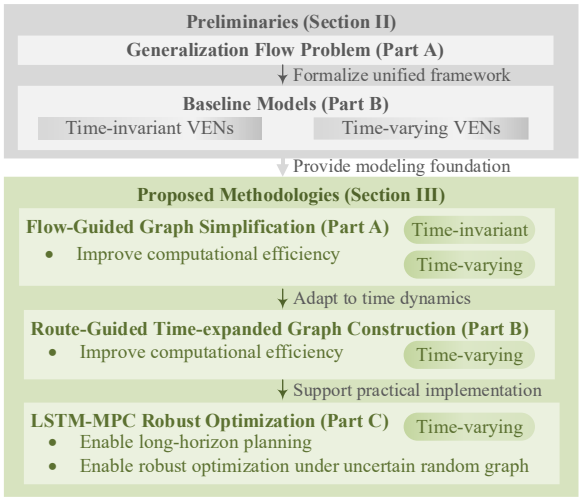


Fig. 2. Overview of proposed work.

generation [31], which iteratively expands a restricted master problem, but its finite-horizon assumption prevents it from handling the unbounded temporal growth of time-expanded VENs. Hence, another typical method is the rolling-horizon strategy [32], which executes immediate decisions and advances sequentially to ensure tractability under continuous evolution. Its sliding-window formulation naturally partitions time, enabling the use of current information to forecast subsequent states and thereby handle both randomness and uncertainty [33]. Compared with scenario-tree approaches [34], it avoids exponential state-space growth. Compared with distribution-based methods [35], it does not rely on precise probabilistic modeling. However, prediction-based schemes cannot guarantee accurate forecasts, so robust enhancement is required to ensure resilience under disruptive events.

Therefore, based on the above identified gaps, the main contributions of our work are as follows:

- A flow-guided graph reduction method is developed, which leverages vehicular flow patterns and energy transfer pathways to iteratively construct a reduced subgraph that preserves critical supply-demand connectivity with better scalability.
- A route-guided time-expanded graph construction approach is introduced to reduce temporal redundancy. It selectively generates spatiotemporally relevant nodes and arcs along active vehicular routes, ensuring alignment with operational dynamics.
- A robust optimization framework of the long short-term memory-based model predictive control (LSTM-MPC) is designed, which integrates learning-based forecasting with rolling-horizon robust optimization. By predicting stochastic inputs and applying residual-aware correction, the method captures both randomness and uncertainty, enabling resilient energy routing over long-time horizons.

Section II presents preliminaries, and Sections III introduce the proposed methodology. Section IV conducts case studies. A discussion and conclusion are drawn in Section V and Section VI, respectively.

## II. Preliminaries

### A. Generalized Flow Problem

In VENs, energy routing can be modeled as a generalized flow problem [36] on a directed graph $\mathcal{G}=(\mathcal{N},\mathcal{A})$, where $\mathcal{N}$ represent nodes, and $\mathcal{A}$ indicates arcs between nodes. Each arc $(i,j)\in\mathcal{A}$ is associated with a transfer efficiency $M_{ij}$, cost $C_{ij}$, and upper capacity $U_{ij}$. Let $x_{ij}$ denote the amount of energy routed from $i$ to $j$. The goal is to minimize the total transmission loss:

$$\min \sum_{(i,j)\in\mathcal{A}} C_{ij}x_{ij} \tag{1}$$

subject to

$$\sum_{j:(i,j)\in\mathcal{A}} x_{ij} - \sum_{j:(j,i)\in\mathcal{A}} M_{ji}x_{ji} = B_i, \quad \forall i\in\mathcal{N} \tag{2}$$

$$0 \le x_{ij} \le U_{ij}, \quad \forall (i,j)\in\mathcal{A} \tag{3}$$

where $B_i$ denotes net supply (positive) or demand (negative) at node $i$. This formulation provides a globally optimal solution for energy routing in VENs, which can be solved based on LP.

### B. Baseline Method

The study in [28] serves as the foundation for our model, and the core formulations are explained below.

#### 1) VEN Specification

In VEN, the road network is represented as $\mathcal{D}=(\mathcal{V},\mathcal{E})$, where $\mathcal{V}$ represents road junctions, and $\mathcal{E}$ indicates edges between junctions. EVs travel through vehicular routes $\mathcal{R}$, which are represented as ordered sequences of junctions. Time is discretized into slots $T_0$, and vehicular flow $F_n^{\mathcal{R}}$ on each route is defined per time slot. Each EV can carry up to $W$ units of energy, with wireless charging and discharging at junctions governed by efficiency factors $Z_c$ and $Z_d$, respectively. In Fig. 1(a), the road network $\mathcal{D}$ is illustrated using circles and arrows

**Algorithm 1:** Basic Time-Invariant VEN Model

**Input:** Set of routes $\mathcal{R}$, set of vehicles $\mathcal{V}$ with supplies $B_i$
**Output:** VEN graph $\mathcal{G} = (\mathcal{N}, \mathcal{A})$

Add all nodes in $\mathcal{V}$ to the network with their corresponding supplies $B_i$;
**foreach** *route* $\mathcal{R}_n \in R$ **do**
  **for** $m = 1$ **to** $|\mathcal{R}_n|$ **do**
    Add node $J^n_m.n.m$ with zero supply;
    **if** $m \neq |\mathcal{R}_n|$ **then**
      Add directed arc $(J^n_m, J^n_m.n.m)$ with cost $1 - Z_c$, multiplier $Z_c$, infinite capacity;
    **if** $m \neq 1$ **then**
      Add directed arc $(J^n_m.n.m, J^n_m)$ with cost $1 - Z_d$, multiplier $Z_d$, infinite capacity;
      Add directed arc $(J^n_{m-1}.n.m-1, J^n_m.n.m)$ with cost 0, multiplier 1, capacity $WF_n$;
**foreach** *node i with positive nonzero supply* **do**
  Add directed arc $(i, i)$ with cost 0, multiplier 0.5, infinite capacity;

**Algorithm 2:** Basic Time-Varying VEN Model

**Input:** Time horizon $T_0$, set of routes $\mathcal{R}$, supplies $B_i$ of junction nodes $\mathcal{V}$
**Output:** Time-varying VEN graph $\mathcal{G}^{T_0} = (\mathcal{N}^{T_0}, \mathcal{A}^{T_0})$

**for** $t = 1$ **to** $T_0$ **do**
  Perform Algorithm 1 (without line 11, 12), and take $B_i$ as $B^t_i T$;
  Add the resulting nodes with their supplies and the resulting arcs with their costs, multipliers, and capacities to $\mathcal{G}^{T_0}$;
  Attach subscript $t$ to the nodes added to $\mathcal{G}^{T_0}$ in the above line;
**for** $n = 1$ **to** $|\mathcal{R}|$ **do**
  **for** $m = 1$ **to** $|\mathcal{R}_n| - 1$ **do**
    Let $L$ be the travel time between $J^n_m$ and $J^n_{m+1}$;
    **for** $t = 1$ **to** $T_0 - L$ **do**
      Add arc $(J^n_m.n.m_t, J^n_{m+1}.n.(m+1)_{t+L})$ with cost 0, multiplier 1, and capacity $WF^t_n T$;

**Algorithm 3:** Flow-Guided Graph Reduction

**Input:** Time steps $T_0$, number of extend steps $n_{\text{trans}}$
**Output:** Simplified graph $\mathcal{G}_{\text{simplified}}$

**Initialize:** Define initial $\mathcal{V}_{\text{sel}}$ and $\mathcal{R}_{\text{sel}}$ ; // by Eq. (4)-(6)
Fix demand node set $\mathcal{V}_{\text{dem}}$ as static sinks;
**for** $t = 1$ **to** $T_0$ **do**
  **for** $e = 1$ **to** $n_{trans}$ **do**
    Update $\mathcal{V}_{\text{sel}}$
    Select candidate routes ; // by Eq. (7)
    Compute node-level importance weights ; // by Eq. (8)-(9)
    Identify intermediate relays ; // by Eq. (10)-(11)
    Calculate $\mathcal{V}_{\text{sel}}$ and $\mathcal{R}_{\text{sel}}$ ; // by Eq. (12)-(13)
    Update $\mathcal{R}_{\text{sel}}$
Trim $\mathcal{R}_{\text{sel}}$ by $\mathcal{V}_{\text{sel}}$ ; // by Eq. (14)-(15)
Output final simplified graph $\mathcal{G}_{\text{simplified}}$ based on $\mathcal{V}_{\text{sel}}$, $\mathcal{R}^{\text{trim}}_{\text{sel}}$;

to represent 4 junctions and 5 directional edges, with edge weights indicating travel time. Two vehicular routes are shown. Route $\mathcal{R}_1$ travels from node 1 to node 3, and Route $\mathcal{R}_2$ proceeds from node 1 to node 4. This example will serve as the basis for explaining the VEN model.

*2) Basic Time-Invariant VEN Model*

To enable energy routing over a VEN in the time-invariant case, a generalized flow-based graph $\mathcal{G}$=($\mathcal{N}$,$\mathcal{A}$) is constructed based on the physical road network $\mathcal{D}$=($\mathcal{V}$,$\mathcal{E}$) and vehicular routes $\mathcal{R}$. The objective is to solve a multi-source multi-destination energy routing problem that minimizes total transmission loss with related constraints. The node $\mathcal{N}$ includes: 1) Junction nodes $J^n_m$. These nodes directly represent the $m$-th physical road junction visited by route $\mathcal{R}_n$. Actual supply $B$>0 and demand $B$<0 are assigned. 2) Artificial nodes $J^n_m.n.m$. They are created for each occurrence of $m$-th junction along $n$-th vehicular route.

The arc $\mathcal{A}$ consist of three types: 1) Transport arcs between artificial nodes ($J^n_{m-1}.n.m-1$, $J^n_m.n.m$). They model the lossless energy transfer onboard EVs, with cost $C$=0, multiplier $M$=1, and capacity $U = WF_n$, where $F_n$ is the vehicular flow and $W$ is the energy pocket per vehicle, representing the physical upper limit of transferable energy. 2) Transfer arcs connecting artificial nodes to junction nodes ($J^n_m.n.m$, $J^n_m$ ). These arcs capture the wireless charging and discharging processes. Their cost and multipliers are set to reflect the energy losses during routing. For charging, $M$=$Z_c$ and $C$=(1− $Z_c$), while for discharging, $M$=$Z_d$ and $C$=(1−$Z_d$). The capacity $U$=∞; 3) Self-loops at supply junctions ($J^n_m$, $J^n_m$). These arcs carry no cost, and they are only constructed for supply junctions, allowing surplus energy to remain at the origin without being routed, hence $M$=1 and $C$=0. Fig. 1(b) shows the baseline VEN model in a time-invariant scenario. Solid nodes are junctions; dashed nodes are artificial nodes. Only arcs between junctions and artificial nodes incur energy loss, while transport along artificial nodes is lossless. This structure enables efficient modeling of EV energy loss in VENs and directly supports multi-source and multi-destination routing via LP. The detailed steps of the basic time-invariant VEN model is descripted in Algorithm 1 [28].

*3) Baseline Model for Time-Varying VENs*

To handle time-varying supply $B^t_i$ and vehicle flow $F^t_n$, the graph $\mathcal{G}$ is transformed into a time-expanded graph $\mathcal{G}^{T_0}$, where each time slot $t$ corresponds to a replica of $\mathcal{G}$ with the number of total time slot $T_0$. In the time-expanded model in Fig. 1(c), the supply of each junction is set as $B^t_i T$, where $T$ is the time slot duration. All junctions and artificial nodes and associated transfer arcs are replicated similarly. Specially, the transport arc is created from the artificial node ($J^n_m.n.m_t$, $J^n_{m+1}.n.(m+1)_{t+L}$) with cost $C$=0, multiplier $M$=1, and capacity $WF^t_n T$, where $L$ is the travel delay. This transforms temporal dynamics into spatial structure, allowing the use of time-invariant flow optimization techniques. The detailed steps of the basic time-varying VEN model is descripted in Algorithm 2 [28].

The above basic model serves as the foundation of our study, upon which Algorithms 3 and 4 are subsequently developed, without affecting the basic LP formulation for generalized flow problem.

## III. Proposed Methodology

To clarify the overall logic of the proposed approach, Fig. 2 presents an overview of the framework. Building on the Section II as the modeling foundation, three key components are developed: (A) Flow-guided graph reduction, to reduce structural complexity under both time-invariant and time-varying settings; (B) Route-guided time-expanded graph construction, to efficiently model temporal dynamics while avoiding redundant expansion; (C) LSTM-MPC robust optimization, to support long-horizon planning under an uncertain random graph for practical implementation.

### A. Flow-Guided Graph Reduction

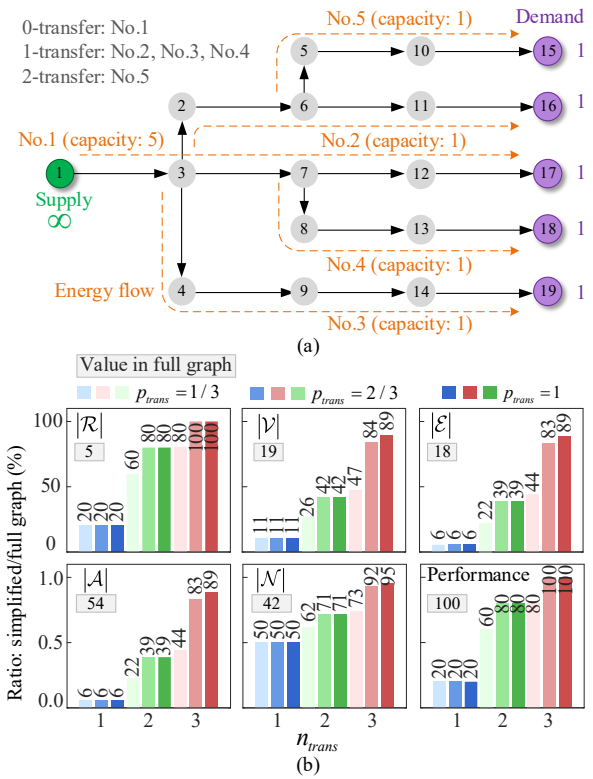


Fig. 3. Example of the directed acyclic graph and performance-size trade-off. (a) Graph structure with supply, demand, and routes. (b) impact of parameters on performance (supply-demand balance) and the model size.

The complexity of the baseline model is decided by $|\mathcal{R}|$ and $|\mathcal{V}|$. In time-invariant VEN, the graph contains $|\mathcal{V}|+|\mathcal{R}|(\bar{m}+1)$ nodes, and $3|\mathcal{R}|\bar{m}+Q$ arcs, where $\bar{m}$ denotes the average length of the route, and $Q$ is the total number of supply junctions. For the time-varying model, $(|\mathcal{V}|+|\mathcal{R}|(\bar{m}+1))T_0$ and $2|\mathcal{R}|\bar{m}T_0+|\mathcal{R}|\bar{m}(T_0-\bar{L})+K$ are the numbers of nodes and arcs, where $K$ is the number of instances that a junction has positive supply at a time slot during all duration over all junctions. In the generalized flow model, the number of arcs and nodes corresponds to the number of variables and constraints. The theoretical complexity of solving such LPs depends on the algorithm used, e.g., simplex or interior-point methods, typically polynomial but with worst-case scenarios reaching exponential time [37, 38]. Large-scale VENs with large $|\mathcal{R}|$ and $|\mathcal{V}|$ can lead to excessive memory usage and computational burden, so graph reduction is crucial for ensuring scalability and practical applicability.

In practice, not all vehicular routes and junctions are equally relevant to energy transfer. Routes that do not connect supply and demand nodes, or low-flow junctions, contribute little to optimal routing but add significant computational overhead. To mitigate this, we propose a flow-guided graph reduction method based on progressive energy propagation from supply to demand nodes via high-impact junctions. The process begins with identifying selected node set $\mathcal{V}_{sel}$ and route set $\mathcal{R}_{sel}$.

$$\mathcal{V}_{\text{sup}}^{(0,0)}=\{i\in\mathcal{V}\mid B_i(0)>0\} \tag{4}$$

$$\mathcal{V}_{\text{dem}}=\{i\in\mathcal{V}\mid \exists t, B_i(t)<0\} \tag{5}$$

$$\mathcal{V}_{\text{sel}}=\mathcal{V}_{\text{dem}}\cup\mathcal{V}_{\text{sup}}^{(0,0)},\quad \mathcal{R}_{\text{sel}}=\varnothing \tag{6}$$

Specially, the demand node set $\mathcal{V}_{\text{dem}}$ is fixed as persistent energy sinks, while the supply set $\mathcal{V}_{\text{sup}}^{(t,e)}$ updates iteratively through relay selection as dynamic energy sources. In each iteration, candidate routes connecting current supply nodes to downstream demand nodes are selected.

$$\mathcal{R}_{\text{cand}}^{(t,e)}=\left\{r_j\in\mathcal{R}\mid \exists v_s\in\mathcal{V}_{\text{sup}}^{(t,e)},\exists v_d\in\mathcal{V}_{\text{dem}}:\text{pos}_{r_j}(v_s)<\text{pos}_{r_j}(v_d)\right\} \tag{7}$$

A set of candidate nodes is defined, and their weights are calculated to reflect the cumulative flow of associated routes:

$$\mathcal{V}_{\text{cand}}^{(t,e)}=\bigcup_{r_j\in\mathcal{R}_{\text{cand}}^{(t,e)}} r_j \tag{8}$$

$$w_v^{(t,e)}=\sum_{r_j\in\mathcal{R}_{\text{cand}}^{(t,e)}}\sum_{v\in r_j} f_j \tag{9}$$

Next, top-ranked junctions are selected as intermediate relays.

$$k^{(t,e)}=\left\lceil p_{trans}\left|\mathcal{V}_{\text{cand}}^{(t,e)}\right|\right\rceil \tag{10}$$

$$\mathcal{V}_{\text{relay}}^{(t,e)}=\arg\max_{\mathcal{S}\subset\mathcal{V}_{\text{cand}}^{(t,e)},|\mathcal{S}|=k^{(t,e)}}\sum_{v\in\mathcal{S}} w_v^{(t,e)} \tag{11}$$

The selected sets are updated:

$$\mathcal{V}_{\text{sup}}^{(t,e+1)}=\mathcal{V}_{\text{relay}}^{(t,e)} \tag{12}$$

$$\mathcal{V}_{\text{sel}}\leftarrow\mathcal{V}_{\text{sel}}\cup\mathcal{V}_{\text{sup}}^{(t,e+1)},\ \mathcal{R}_{\text{sel}}\leftarrow\mathcal{R}_{\text{sel}}\cup\mathcal{R}_{\text{cand}}^{(t,e)} \tag{13}$$

where $\leftarrow$ indicates the update process. Final selected sets are determined after the iterations. Subsequently, $\mathcal{R}_{\text{sel}}$ is trimmed as:

$$r_j^{\text{trim}}=\left\{v\in r_j\mid v\in\mathcal{V}_{\text{sel}}\right\} \tag{14}$$

$$\mathcal{R}_{\text{sel}}^{\text{trim}}=\left\{r_j^{\text{trim}}\mid r_j\in\mathcal{R}_{\text{sel}}\right\},\ r_j^{\text{trim}}=\left\{v\in r_j\mid v\in\mathcal{V}_{\text{sel}}\right\} \tag{15}$$

The final simplified graph is built based on $\mathcal{V}_{\text{sel}}$ and $\mathcal{R}_{\text{sel}}^{\text{trim}}$. The procedure is outlined in Algorithm 3, which is applicable to both time-invariant and time-varying scenarios.

The proposed graph reduction method exhibits a polynomial-time complexity of $O(T_0 n_{trans}(|\mathcal{R}|\bar{m}+|\mathcal{V}|\log|\mathcal{V}|))$. Fig. 1(d) and 1(e) illustrate the reduction outcome under $n_{trans}$=1, and only supply and demand junctions are included in $\mathcal{V}_{sel}$. To evaluate the computational impact, we consider interior-point methods as a representative LP solver with a complexity of $O((|\mathcal{A}|+|\mathcal{N}|)^{3.5})$. The time-invariant cases show a complexity reduction from $O((16+11)^{3.5})$ to $O((4+4)^{3.5})$, achieving an approximate 98.6% decrease. In the time-varying case over 3 periods, the complexity decreases from $O((42+33)^{3.5})$ to $O((9+12)^{3.5})$, corresponding to a 96.5% reduction. These results theoretically validate the substantial efficiency gains enabled by our graph reduction approach, particularly under conditions with limited demand nodes, intensive flow routes, and sufficient energy supply.

The proposed method is designed to align closely with the optimization objective, i.e., minimizing energy loss. Energy loss is primarily caused by charging and discharging operations. Hence, fewer route transfers correspond to lower loss. Our proposed method follows this principle by first identifying directly connected supply-demand pairs, then gradually expanding to include routes with one transfer, and so on. However, when the parameters are set low, critical routes or junctions may be excluded. Therefore, rather than optimality of the loss, the proposed method has a more direct impact on feasibility, which relates to the demand satisfaction.

Therefore, the sufficient condition of preserving feasibility and optimality is: The routing topology induced by $\mathcal{R}$ is a

directed acyclic graph whose leaf nodes are demand nodes. *Proof.* For a feasible generalized flow problem, $\forall\ i \in \mathcal{V}_{dem}$, must $\exists$ a set of $x_{ji}$ meet (2) and (3). That is, for every demand node, there exists at least one feasible route, combination of concatenated routes, that delivers required energy from supply nodes. For a directed acyclic graph, it can be formed based on a recursive forward-expansion process, proceeding from backbone chains, to branches, and subsequently to sub-branches. Accordingly, our method first identifies supply-demand route as backbone path. Then, we iteratively expand the supply pool, which consists of subset of the nodes previously visited, and then search for the sub-branches. When $n_{trans}$ and $p_{trans}$ are set sufficiently large, the algorithm is capable of discovering all feasible branches. Since all leaf nodes are demand nodes, all routes leading to each demand node can be effectively found. Therefore, the feasibility and optimality are preserved. A typical example is illustrated in Fig. 3(a). Based on the proposed method, in the first-round searching, the direct route No.1 can be included, followed by one-transfer routes, i.e., No. 2, 3, and 4, discovered in the second iteration. Route No. 5 is identified in the third iteration. This demonstrates that, with sufficiently large parameter settings, the complete graph can be recovered with ensured performances. However, in networks that rely on short relay routes or fragmented flows, the method may remove critical paths, potentially leading to infeasibility.

Next, the connectivity preservation bounds for demand nodes are discussed. The reachable demand nodes are classified into levels based on their proximity to the supply nodes. Let $p^1_{dem}$ denote the proportion of demand nodes directly reachable (one expansion step); $p^2_{dem}$ for those reachable via two expansion steps, and so on. When $p_{trans}$=1, since the minimum value of $n_{trans}$=1, the method guarantees a minimum connectivity of $p^1_{dem}$. As $n_{trans}$ increases to 2, the connectivity also increases to $p^1_{dem}+ p^2_{dem}$, this pattern continues recursively. When $p_{trans}$ decreases, the algorithm selects only a subset of the candidate junctions, which may lead to partial connectivity loss. Therefore, in the worst case, up to $(1-p^1_{dem})$ of demand nodes may be disconnected. Then, for the approximation error of supply-demand balance, in the optimal-case scenario, all demand nodes receive their required energy $B_{dem}$. Based on above analysis, in the worst-case scenario, only $p^1_{dem}$ demand nodes can be reached. Given the maximum total energy flow capacity to these nodes is $U^1_{dem}$, the maximum deliverable energy is $Z_d U^1_{dem}$. Therefore, the worst-case approximation error in terms of supply-demand balance can be upper-bounded by $|B_{dem}- Z_d U^1_{dem}|$.

The trade-off between model size and performance is primarily controlled by the parameters $n_{trans}$ and $p_{trans}$. Performance is measured based on demand satisfaction, while size is characterized by several indicators, including $|\mathcal{V}|$, $|\mathcal{E}|$, and $|\mathcal{R}|$ from the road network, as well as $|\mathcal{A}|$ and $|\mathcal{N}|$ from the VEN model. Take Fig. 3(a) an example, we evaluate several discrete parameter settings. Efficiency loss is not considered in this analysis. It can be noticed that $n_{trans}$ is the main factor influencing the results, as it decides the depth of exploration. However, $p_{trans}$ also affects the outcome as it controls the selection of starting points for exploration. In the case, settings of $n_{trans}$ = 2 and $p_{trans}$ = 2/3 provide a relatively balanced result. It retains 80% original performance, while reducing $|\mathcal{A}|$ to 39% and $|\mathcal{N}|$ to 71%. These two metrics are directly associated with the computational efficiency of LP. In more complex cases, where multiple routing options are available for each demand node, there is even greater potential to reduce graph size without compromising performance.

**Algorithm 4:** Route-Guided Time-Expanded Graph Construction

**Input:** Supply matrix $B$, vehicular route set $\mathcal{R}$, number of periods $T_0$
**Output:** $\mathcal{G}^{T_0} = (\mathcal{N}^{T_0}, \mathcal{A}^{T_0})$

**for** $n = 1$ **to** $|\mathcal{R}|$ **do**
  **for** $t = 0$ **to** $T_0 - 1$ **do**
    **for** $m = 1$ **to** $|R_n| - 1$ **do**
      Let $L$ be the travel time between them;
      **if** $t + L < T_0$ **then**
        **if** $J^n_{mt}$ *is not yet created* **then**
          Create junction node $J^n_{mt}$ and assign supply $B^t$;
        **if** $J^n_{m+1t+L}$ *is not yet created* **then**
          Create junction node $J^n_{m+1t+L}$ and assign supply $B^{t+L}$;
        Create nodes $J^n_m.n.m_t$ and $J^n_{m+1}.n.(m+1)_{t+L}$;
        **if** $m = 1$ **then**
          Add arc $(J^n_{mt},\ J^n_m.n.m_t)$ with cost $1 - Z_c$, multiplier $Z_c$, capacity $\infty$;
        **else if** $m = |R_n| - 1$ **then**
          Add arc $(J^n_{m+1}.n.(m+1)_{t+L},\ J^n_{m+1t+L})$ with cost $1 - Z_d$, multiplier $Z_d$, capacity $\infty$;
        **else**
          Add bidirectional arcs between $J^n_m.n.m_t$ and $J^n_{mt}$ with cost $1 - Z_{c/d}$, multiplier $Z_{c/d}$, capacity $\infty$;
        Add arc $(J^n_m.n.m_t,\ J^n_{m+1}.n.(m+1)_{t+L})$ with cost 0, multiplier 1, capacity $WF^t_n T$;
**foreach** *node* $J^n_{mt}$ *with* $B^t > 0$ **do**
  Add self-loop arc $(J^n_{mt}, J^n_{mt})$ with cost 0, multiplier 0.5, capacity $\infty$;

## B. Route-Guided Time-Expanded Graph Construction

Time-expanded graph is a classical modeling approach used to represent multi-period problems by unrolling a static network across discrete time steps. Specifically, the original time-invariant graph is replicated at each time point, creating a layer for every timestamp within the planning horizon. Inter-layer arcs are then added between successive time layers to represent the passage of time. While effective, this approach leads to substantial structural redundancy and computational overhead, especially in long-horizon or high-resolution scenarios. For example, in Fig. 1(c), nodes $4_1$, $4.1.4_1$, and $4.2.3_1$, are only connected locally without engaging in inter-period or inter-node energy transfer. These nodes are redundant from an optimization perspective and can be safely pruned.

To address the structural inefficiency introduced by exhaustive replication, the modeling strategy is redefined in Algorithm 4. The inputs include the supply matrix $B$, a set of vehicular routes $\mathcal{R}$, and the planning horizon length $T_0$. For each route $r_j$, the algorithm first computes the travel time $L$ between origin and destination nodes. It then checks whether the corresponding temporal junction nodes have already been created; if not, they are instantiated and supplied accordingly. This event-driven node creation avoids unnecessary replication and reduces the graph size. For each valid transition, the corresponding temporal arcs are constructed to represent energy-carrying vehicle movement. These arcs are annotated with model-consistent attributes: traversal cost, energy conversion multiplier, and transmission capacity. Additionally, bidirectional arcs and self-loop arcs are selectively introduced to preserve energy flow feasibility and continuity over time. By aligning temporal graph construction

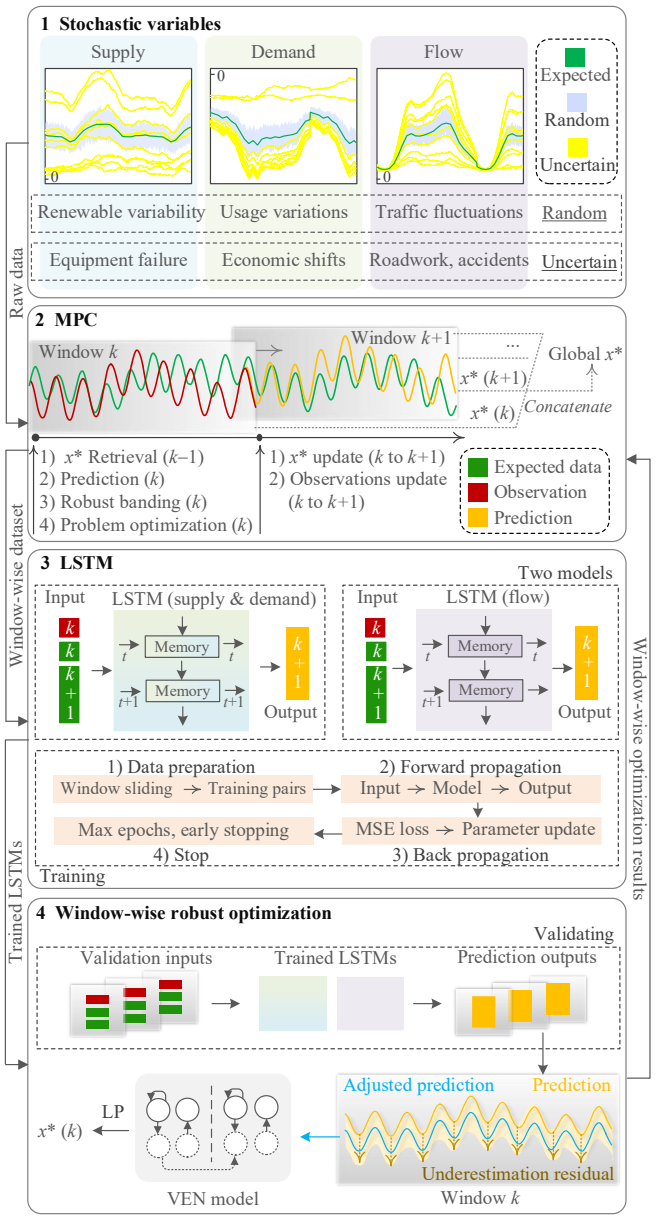


Fig. 4. Overview of the robust optimization framework of the LSTM-MPC.

with actual vehicle routes and energy propagation constraints, it enables a sparser graph representation, which improves the tractability and scalability of downstream flow routing optimization. Furthermore, the proposed construction procedure is inherently compatible with the flow-guided graph reduction method introduced in Part A. The simplified static graph serves as the structural foundation at each time step during the time-expanded construction. Travel times along vehicular routes should be re-evaluated.

*Lemma 1*. Let $\mathcal{G}_{\text{full}}^{T_0}$ be full time-expanded graph. Let $\mathcal{G}_{\text{route}}^{T_0}$ be the route-guided time-expanded graph. Under the same $\mathcal{R}$ and $T_0$, $\mathcal{G}_{\text{route}}^{T_0}$ yields the same feasible set as $\mathcal{G}_{\text{full}}^{T_0}$. *Proof*. In $\mathcal{G}_{\text{full}}^{T_0}$, the time-invariant graph is replicated for all discrete time layers $t$=1, …, $T_0$, revising with temporal connections according to the route with the corresponding travel delays, ensuring preserving all feasible time-feasible set. In $\mathcal{G}_{\text{route}}^{T_0}$, construction begins by traversing $R_n \in \mathcal{R}$. For each route, relevant nodes and arcs are formed as long as the cumulative travel time does not exceed the time horizon $T_0$, otherwise, generation is truncated. Since all routes are exhaustively processed and all time-feasible transitions are preserved, the method captures the complete valid set relevant to the energy routing task. Therefore, $\mathcal{G}_{\text{route}}^{T_0}$ and $\mathcal{G}_{\text{full}}^{T_0}$ represent identical feasible path sets.

### C. LSTM-MPC Robust Optimization

In our paper, a hybrid stochastic graph is modelled, where each node or arc may exhibit randomness, uncertainty, or both [39]. Randomness refers to inherent fluctuations that follow known probability distributions, such as Gaussian variations. In contrast, uncertainty reflects distribution-free deviations, caused by unpredictable factors. For example, renewable energy supply often exhibits Gaussian fluctuations driven by wind speed or solar irradiance, while it may shift due to equipment aging or unexpected failures. User demand may follow a probabilistic daily profile but is changed by factors such as economic events or seasonal patterns [40]. Similarly, traffic flow may vary randomly over time yet be constrained by uncertain capacity limits caused by road maintenance, accidents, or construction activities [41]. To address long-term stochasticity in dynamic environments, a LSTM-MPC robust optimization method is newly proposed. The overall framework is shown in Fig. 4. It combines data-driven prediction via LSTM with robust optimization under a rolling MPC scheme, enabling resilient and anticipatory routing decisions.

In the baseline, the overall graph size scales linearly with $T_0$, rendering long-horizon optimization computationally demanding. To address this, MPC is employed to decompose the global problem into a sequence of rolling subproblems [42]. The full phase $T_0$ can be divided into overlapping windows $[t_s^k, t_e^k]$.

$$t_s^k = (k-1)t_{step} + 1, \quad t_e^k = t_s^k + H - 1 \tag{16}$$

where $H$ is the window length and $t_{step}$ is the step size. To ensure temporal consistency in the time-expanded network, the time window size $H$ is constrained by the minimum travel time across all route segments. Specifically, $H$ must not exceed the shortest segment duration, such that each segment occupies at least one discrete time step after mapping. This prevents zero-step assignments that would result in route loss during graph construction. After partitioning, a subgraph $\mathcal{G}^k$ is formed for each window, incorporating vehicular routes and flows within the current horizon. To maintain temporal consistency across rolling windows, decisions in overlapping regions are fixed based on the previous window's outputs. The subsequent window only optimizes new variables, effectively ensuring stability in rolling decision-making. Also, Algorithms 3 and 4 can be applied within each subgraph. To ensure temporal consistency across windows, a cross-window state propagation mechanism is introduced. For each junction node $i$, the updated supply/demand in window $k$ is:

$$B_i^k \leftarrow B_i^k + \sum_{j:(j,i)\in\mathcal{A}} M_{ji} x_{ji}^{k-1} - \sum_{j:(j,i)\in\mathcal{A}} x_{ij}^{k-1} \tag{17}$$

This adjustment reflects inter-window energy exchanges, maintaining supply-demand coherence over time. Under dynamic and stochastic conditions, strict feasibility may be difficult to guarantee. Infeasibility in any window could

disrupt the entire planning process. Therefore, slack variables $s_i^k$ are introduced for each demand node. These variables allow temporary constraint violation while penalizing it in the objective:

$$\min \sum_{(i,j)\in\mathcal{A}^k} C_{ij}x_{ij}^k + \gamma \sum_{i\in\mathcal{N}_d^k} s_i^k \tag{18}$$

subject to

$$\sum_{j:(i,j)\in\mathcal{A}^k} x_{ij}^k - \sum_{j:(j,i)\in\mathcal{A}^k} M_{ji}x_{ji}^k = B_i^k + s_i^k \ , \quad \forall i \in \mathcal{N}_d^k \tag{19}$$

$$\sum_{j:(i,j)\in\mathcal{A}^k} x_{ij}^k - \sum_{j:(j,i)\in\mathcal{A}^k} M_{ji}x_{ji}^k = B_i^k \ , \quad \forall i \in \mathcal{N}^k \setminus \mathcal{N}_d^k \tag{20}$$

$$0 \le x_{ij}^k \le U_{ij}, \qquad \forall (i,j) \in \mathcal{A}^k \tag{21}$$

$$s_{ij}^k \ge 0, \qquad \forall i \in \mathcal{N}_d^k \tag{22}$$

The rolling procedure continues over successive windows, with results stitched together to form a complete solution trajectory. Compared to full-horizon optimization, the proposed method effectively reduces the optimization scale at each step of inter-window coordination.

In the presence of historical data, predictive modeling allows us to transform part of the uncertainty into estimable quantities. Compared to conventional robust optimization methods, such as distributionally robust optimization (DRO) or worst-case optimization, learning-based forecasts offer a data-driven way to proactively adapt to anticipated fluctuations rather than react passively, thus improving robustness performances. LSTM is utilized as it enhances sequence modeling by integrating gating mechanisms and memory cells, offering significant improvements over conventional deep neural networks [33]. It is able to learn the mapping between the historical input and the predicted output sequence. In our design, two separate models are trained to forecast the key stochastic components: supply/demand $B$ and flow capacity $U$. For the $k$-th window, the inputs of two models are organized as a matrix formed by concatenating node-wise ($B$) or edge-wise ($U$) vectors:

$$u^{k,B} = \left[ u_1^{k,B}, u_2^{k,B}, ..., u_{|\mathcal{N}|}^{k,B} \right]^T \in \mathbb{R}^{|\mathcal{N}|\times 3H} \tag{23}$$

$$u_i^{k,B} = [B_i^k, B_i^{k,\exp}, B_i^{(k+1),\exp}] \tag{24}$$

$$u^{k,U} = \left[ u_1^{k,U}, u_2^{k,U}, ..., u_{|\mathcal{A}|}^{k,U} \right]^T \in \mathbb{R}^{|\mathcal{A}|\times 3H} \tag{25}$$

$$u_{i,j}^{k,U} = [U_{i,j}^k, U_{i,j}^{k,\exp}, U_{i,j}^{(k+1),\exp}] \tag{26}$$

where the superscript $k$ denotes the index of the $k$-th window, and the subscript *exp* refers to the expected value derived from historical data. Therefore, for the $i$-th node or edge ($i$,$j$), the input vector is constructed by concatenating three components: the real-time observed value in window $k$, the historical expected value in window $k$, and the historical expected value in window $k$+1. Next, the cell state $s^k$ regulates temporal information across time steps. It is updated at each time step by integrating previous memory and new candidate information as follows:

$$s^k = f^k \odot s^{k-1} + g^{k,(in)} \odot \hat{s}^k \tag{27}$$

The forget gate $f^k$ determines the extent to which the previous memory $s^{k-1}$ is preserved, while the input gate regulates the incorporation of new information.

$$f^k = \sigma\left( W^{(forg)} \left[ h^{k-1}, u^k \right] + d^{(forg)} \right) \tag{28}$$

$$g^{k,(in)} = \sigma\left( W^{(in)} \left[ h^{k-1}, u^k \right] + d^{(in)} \right) \tag{29}$$

$$\tilde{s}^k = \tanh\left( W^{(cand)} \left[ h^{k-1}, u^k \right] + d^{(card)} \right) \tag{30}$$

The output is computed as:

$$o^k = \sigma\left( W^{(out)} \left[ h^{k-1}, u^k \right] + d^{(out)} \right) \tag{31}$$

$$h^k = o^k \odot \tanh(s^k) \tag{32}$$

$$\tilde{y}^k = \phi\left( W^{(y)} h^k + d^{(y)} \right) \tag{33}$$

The transformation is governed by the weight $W^{(forg)}$, $W^{(in)}$, $W^{(cand)}$, $W^{(out)}$, and $W^{(y)}$, associated bias terms including $b^{(forg)}$, $b^{(in)}$, $b^{(cand)}$, $b^{(out)}$, and $b^{(y)}$. The activation functions include $\sigma(\cdot)$, $\phi(\cdot)$ and $\tanh(\cdot)$. The target output for training is the real observed curve in the next window $U_{i,j}^{k+1}$ and $B_i^{k+1}$. The loss function employed is the mean squared error (MSE) between the predicted and observed values.

LSTM models are utilized to predict supply, demand, and flow. However, the predictions cannot be perfect due to the data-driven nature. To enhance the robustness under randomness and uncertainty, a residual-based correction mechanism is developed to refine LSTM predictions conservatively before optimization, which does not rely on any predefined statistical distribution. It is a data-driven approach that extracts the worst-case overestimation error within each time window to construct a downward correction band. The degree of conservativeness is controlled by a tunable robustness parameter, which prevents overly optimistic forecasts from leading to infeasible scheduling decisions. Implementation details are as follows.

Step 1: Model training and test data preparation. Two independent LSTMs are trained for supply/demand and flow prediction, respectively. A sufficient and independent testing dataset is reserved for evaluation without data leakage.

Step 2: Sliding-window setup and residual calculation. Based on the testing set, predictions are formed based on trained LSTMs. The residual is calculated as the difference between actual observed value and the LSTM-predicted value of each time interval. A negative residual indicates that the predicted value exceeds the actual value, meaning an overestimated prediction. For the $k$-th window, the worst-case estimations are extracted by selecting the most negative prediction errors:

$$\varepsilon_i^{k,B} = \min\left\{ B_i^k - \tilde{B}_i^k \middle| \left( B_i^k, \tilde{B}_i^k \right) \in \mathcal{D}_i^k \right\} \tag{34}$$

$$\varepsilon_{i,j}^{k,U} = \min\left\{ U_{i,j}^k - \tilde{U}_{i,j}^k \middle| \left( U_{i,j}^k, \tilde{U}_{i,j}^k \right) \in \mathcal{D}_i^k \right\} \tag{35}$$

where $\mathcal{D}_i^k$ denotes the testing set with the $k$-th window, and the symbol ~ represents predictions.

Step 3: robust inputs construction. Within each sliding window, a linear robustness correction is applied to the predicted result as:

$$\tilde{B}_i^k = \tilde{B}_i^k + \lambda \varepsilon_i^{k,B} \ , \ \tilde{U}_{i,j}^k = \tilde{U}_{i,j}^k + \lambda \varepsilon_{i,j}^{k,U} \tag{36}$$

where $\lambda$ is a parameter that controls the level of conservativeness, ranging from 0 to 1. A larger value yields more conservative inputs. For supply, downward deviations indicate generation shortfalls or surging external demand, while reductions in flow capacity imply diminished vehicle mobility. This formulation aligns with robust optimization practices for bounded uncertainty [43]. Finally, substituting

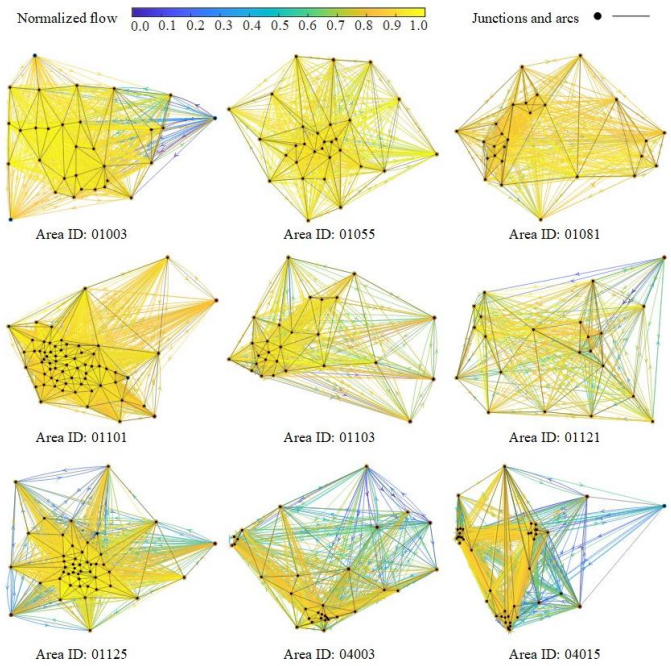


Fig. 5. Normalized OD flow and underlying road networks for 9 representative areas from 100 cases.

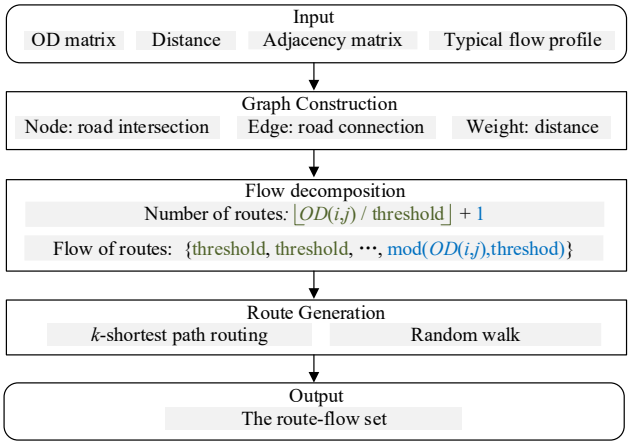


Fig. 6. Flowchart of route generation.

the adjusted variables into (18) to (22) enables reformulation of the stochastic problem as a deterministic LP.

## IV. Experiment Performance on Real-World Data

To evaluate the proposed method under diverse and realistic spatial-temporal scenarios, we conduct experiments on a large-scale dataset covering 3,233 heterogeneous areas across the U.S. [44]. In the dataset, the commuting origin-destination (OD) matrix, combined with regional attributes including demographics and point-of-interests of each region in that area are provided. A representative subset of 100 areas is randomly selected for evaluation. Each area's road network is defined by its adjacency matrix and node positions. Fig. 5 visualizes 9 areas from the 100 test cases to illustrate the spatial distribution of commuting demand. In these graphs, black nodes and arcs represent junctions and roads, and directed edges denote OD pairs. The color of each edge encodes the normalized flow magnitude. The area IDs are consistent with the naming convention used in the public dataset. As observed,

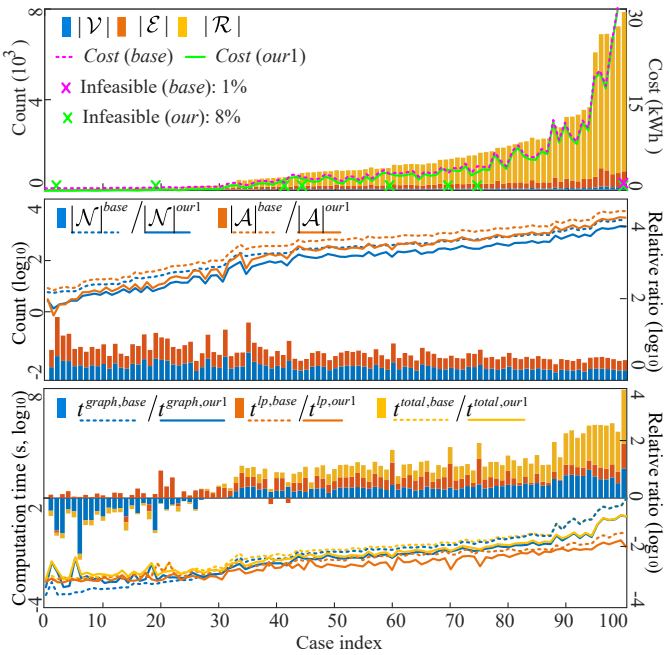


Fig. 7. Performance comparison for time-invariant VENs. Top: junctions, edges, routes, costs, and feasibility. Middle: nodes and arcs related results. Bottom: computation time for graph construction, LP solving, and total runtime.

regions with more centralized structures tend to exhibit denser OD connections in the core areas with higher flow intensities, while peripheral zones appear more sparsely connected, reflecting fragmented commuting patterns. For time-varying VENs, flow data is obtained from [45], and averaged values are used in time-invariant cases. The demand at each region is modeled from [46], with random fluctuations within ±10% to simulate typical variability. Renewable energy supplies are generated using wind power data from [47]. Nodes with predominantly outgoing flows are selected as supply nodes, while those with incoming flows are designated as demand nodes. In each VEN instance, all nodes and flows exhibit random properties, while 30% are further assigned as uncertain, reflecting distribution-free deviations. The total energy supply is set higher than the total demand. Each EV carries 1 kWh of energy, with charging/discharging efficiencies of 0.95 and storage efficiency of 0.97.

The flowchart of route generation is shown in Fig. 6. There are four inputs, including OD, distance, adjacency matrix, and typical flow profile. A digraph is then constructed. For the OD of junction $i$ and $j$, the value is decomposed into multiple sub-demands by dividing it into as many full units of the threshold, plus a residual. The threshold is calculated as the mean value, augmented by 0.5 times the standard deviation for ensuring route diversity. Routes are formed based on $k$-shortest path method, with a random walk used with no feasible path exists. Finally, the output is the route-flow set.

To compare the effectiveness of the proposed components, we define the following naming conventions. There are three main methods proposed in our framework: (1) Graph reduction/simplification method, (2) Time-expanded graph construction method, (3) LSTM-MPC method. In the subsequent comparisons, *our1* refers to the version that only applies method 1; *our12* includes both methods 1 and 2; and

TABLE II
SUMMARY OF NETWORKS FOR TIME-INVARIANT VENS

| Case | 1~33 | 34~66 | 67~100 |
|---|---|---|---|
| $\lvert\mathcal{V}\rvert$ | 8±3 | 27±5 | 61±28 |
| $\lvert\mathcal{E}\rvert$ | 29±18 | 135±33 | 330±165 |
| $\lvert\mathcal{R}\rvert$ | 63±62 | 629±196 | 2536±1780 |

TABLE III
COMPARISON PERFORMANCE FOR TIME-INVARIANT VENS

| Case | | 1~33 | 34~66 | 67~100 |
|---|---|---|---|---|
| $\lvert\mathcal{N}\rvert$ | *base* | 185±203 | 2402±848 | 12770±10311 |
| | *our1* | 71±89 | 1029±429 | 6411±5343 |
| $\lvert\mathcal{A}\rvert$ | *base* | 344±415 | 5238±1954 | 30521±25547 |
| | *our1* | 122±176 | 2226±991 | 15357±13250 |
| $t^{graph}$ (s) | *base* | 0.0±0.0 | 0.1±0.0 | 7.5±17.6 |
| | *our1* | 0.0±0.0 | 0.0±0.0 | 1.2±2.3 |
| $t^{lp}$ (s) | *base* | 0.0±0.0 | 0.0±0.0 | 0.2±0.2 |
| | *our1* | 0.0±0.0 | 0.0±0.0 | 0.0±0.0 |
| $t^{total}$ (s) | *base* | 0.0±0.0 | 0.1±0.1 | 7.7±17.8 |
| | *our1* | 0.0±0.0 | 0.0±0.0 | 1.3±2.4 |
| *error* (%) | | 0.0±0.0 | 0.0±0.0 | 0.0±0.0 |

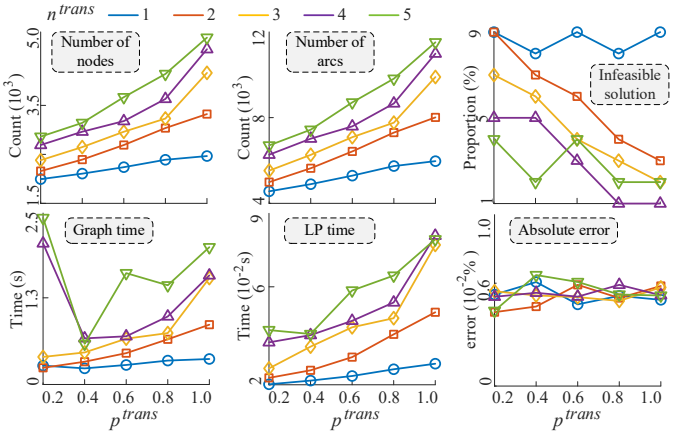


Fig. 8. Sensitivity analysis of key metrices under different $p^{trans}$ and $n^{trans}$.

*our123* represents the full version incorporating all three methods. This naming scheme is used consistently throughout the experimental analysis. The baseline approach presented in [28] is denoted as *base*. All non-integer results are rounded to one decimal place in the presentation. All experiments were conducted on a desktop machine equipped with an Intel® Core™ i5-10505 CPU (12 logical cores, 2.30GHz base frequency, up to ~3.2GHz) and 16 GB of RAM. The operating system is Windows 11 Enterprise 64-bit. The proposed methods were solved using MATLAB's built-in solver *linprog*. The convergence tolerances are set to the default values.

### *A. Performances Under Deterministic Time-Invariant VENs*

This section evaluates the effectiveness of the proposed graph reduction method. Runtime performance is assessed by graph construction time $t^{graph}$ and LP solving time $t^{lp}$. The model size is reflected by the number of nodes $|\mathcal{N}|$ and arcs $|\mathcal{A}|$, which correspond to the number of constraints and decision variables in optimization. Cases are characterized by the number of junctions $|\mathcal{V}|$, edges $|\mathcal{E}|$, and routes $|\mathcal{R}|$, capturing the scale and complexity of real-world road networks. This allows us to examine their impact on optimization results.

As shown in Fig. 7, 100 cases are sorted in descending order of route count. The parameter settings of *our1* are: $p^{trans}$=0.6 and $n^{trans}$=1. The top subplot shows $|\mathcal{V}|$, $|\mathcal{E}|$, and $|\mathcal{R}|$, and cost. The three metrics grow proportionally, reflecting significant variation in network complexity and confirming test case diversity. For feasible instances, both methods yield consistent costs, indicating preserved solution quality. Infeasible cases caused by supply-demand imbalance are marked with crosses. For the *base*, which retains the full graph, results in 1% infeasibility, indicating that the instance is inherently unsolvable. Our proposed method introduces an extra 7% infeasibility due to overly aggressive simplification under the given parameter settings, which leads to supply-demand mismatch and thus solution failure. Both middle and bottom subplots use a $\log_{10}$ scale to capture value variation across cases. Line plots (left y-axis) show absolute metrics, while bar plots (right y-axis) indicate performance ratios between the baseline and our method. Specifically, the middle subplot shows consistent reductions in nodes and arcs with our method, with larger gains under complex networks, demonstrating strong scalability. The bottom subplot reveals similar improvements in computation time, notably in graph construction and LP solving for large-scale cases. These results confirm that our method effectively reduces model size and runtime, making it well-suited for real-world vehicular energy routing.

Table II provides a numerical summary with both the mean and standard deviation of $|\mathcal{V}|$, $|\mathcal{E}|$, and $|\mathcal{R}|$. To illustrate performance variation, test cases are grouped into three segments by network scale. Metric values increase markedly across segments, and the large standard deviation of $|\mathcal{R}|$, reaching 2,536±1,780 in cases 67 to 100, reflects the structural heterogeneity of real-world networks. Table III presents the comparative results on $|\mathcal{N}|$, $|\mathcal{A}|$, computational time, and error. In all segments, the proposed method achieves significant reductions in model size and runtime. For cases 67 to 100, the number of nodes decreases from 12,770±10,311 to 6,411±5,343, while the number of arcs decreases from 30,521±25,547 to 15,357±13,250, representing reductions of 49.8% and 49.7% respectively. Total computation time is reduced by 83.1%, from 7.7±17.8 s to 1.3±2.4 s. Similar trends are observed in smaller segments. In all cases, the optimization error remains zero, confirming that efficiency is improved without compromising solution quality.

Sensitivity analysis evaluates how $p^{trans}$ and $n^{trans}$ affect six key metrics, as shown in Fig. 8. Increasing $n^{trans}$ adds more routing layers, while higher $p^{trans}$ includes more high-flow nodes per step. This leads to steady growth in model size and LP complexity. LP solving time scales accordingly. For graph construction time $t^{graph}$, a non-monotonic trend is observed when $n^{trans}\geq 4$: the runtime first decreases due to reduced redundancy in node additions, then increases sharply as more new nodes are introduced. This indicates the existence of a performance-optimal region where sufficient connectivity is achieved with minimal overhead. Total runtime is mainly driven by $t^{graph}$. In terms of feasibility, higher $n^{trans}$ generally improves solution viability by enabling more demand-supply paths, but $p^{trans}$ may still lead to infeasible cases due to insufficient node inclusion. Solution error remains consistently below one percent, confirming the method's reliability. Also, from the above discussion in Fig. 7 and Table III, our method consistently outperforms the baseline except infeasibility rate.

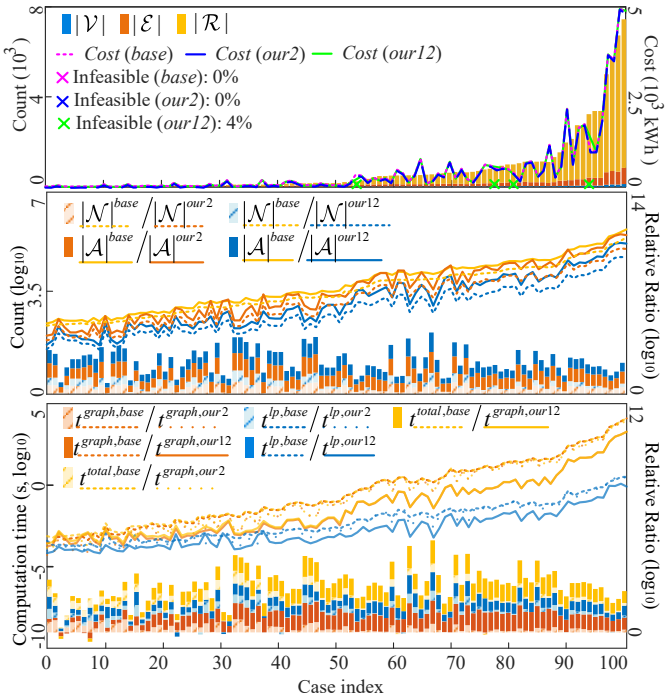


Fig. 9. Performance comparison for time-varying VENs. Top: junctions, edges, routes, costs, and feasibility. Middle: nodes and arcs related results. Bottom: computation time for graph construction, LP solving, and total runtime.

TABLE IV
SUMMARY OF NETWORK FOR TIME-VARYING VENS

| Case | 1~33 | 34~66 | 67~100 |
|---|---|---|---|
| $\|\mathcal{V}\|$ | 5±1 | 14±5 | 43±27 |
| $\|\mathcal{E}\|$ | 15±7 | 62±28 | 228±157 |
| $\|\mathcal{R}\|$ | 26±15 | 191±132 | 1619±1631 |

TABLE V
COMPARISON PERFORMANCE FOR TIME-VARYING VENS

| Case | | 1~33 | 34~66 | 67~100 |
|---|---|---|---|---|
| $\|\mathcal{N}\|$ | *base* | 1495±860 | 12780±9539 | 158013±192349 |
| | *our2* | 747±614 | 6838±6360 | 75350±92855 |
| | *our12* | 341±215 | 2516±2305 | 34766±47350 |
| $\|\mathcal{A}\|$ | *base* | 2134±1327 | 21548±17302 | 301069±383610 |
| | *our2* | 1369±1317 | 16538±17365 | 223262±295318 |
| | *our12* | 596±453 | 5856±6018 | 103633±150150 |
| $t^{graph}$ (s) | *base* | 0.1±0.1 | 3.9±4.7 | 1310.1±3066.0 |
| | *our2* | 0.0±0.0 | 3.1±4.4 | 1013.0±2355.5 |
| | *our12* | 0.0±0.0 | 0.6±0.8 | 263.5±665.1 |
| $t^{lp}$ (s) | *base* | 0.0±0.0 | 0.1±0.1 | 2.4±3.9 |
| | *our2* | 0.0±0.0 | 0.1±0.1 | 2.1±3.6 |
| | *our12* | 0.0±0.0 | 0.0±0.0 | 0.8±1.5 |
| $t^{total}$ (s) | *base* | 0.1±0.1 | 4.0±4.8 | 1312.5±3069.7 |
| | *our2* | 0.0±0.1 | 3.2±4.5 | 1015.1±2359.0 |
| | *our12* | 0.0±0.0 | 0.6±0.8 | 264.3±666.5 |
| *error* (%) | | 0.0±0.0 | 0.0±0.0 | 0.0±0.0 |

Notably, when $p^{trans}$=0.8 and $n^{trans}$=4, the infeasibility rate matches the baseline, while model complexity and runtime remain significantly lower, further demonstrating the effectiveness of the proposed approach. These results highlight the critical role of parameter tuning in balancing graph complexity and computational cost. Moderate expansion achieves efficient computation without sacrificing solution quality, underscoring the value of graph reduction for scalable energy routing.

## B. Performances Under Deterministic Time-Varying VENs

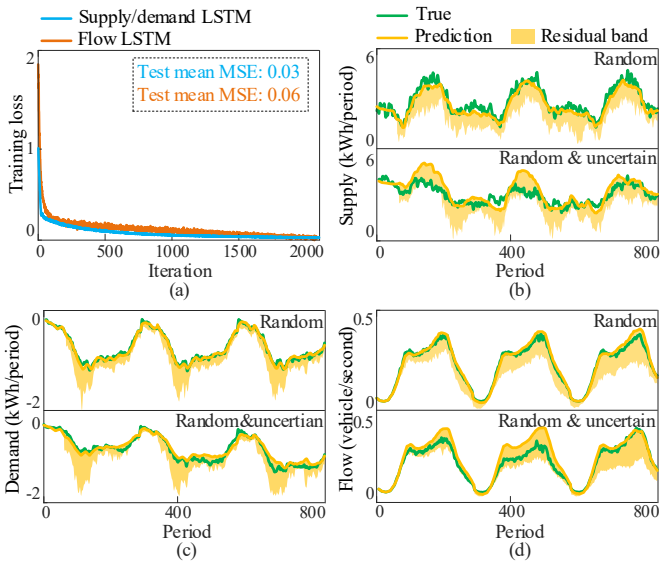


Fig. 10. LSTM training and robust forecast performance. (a) Training loss curve. (b) Supply comparison. (c) Demand comparison. (d) Flow comparison.

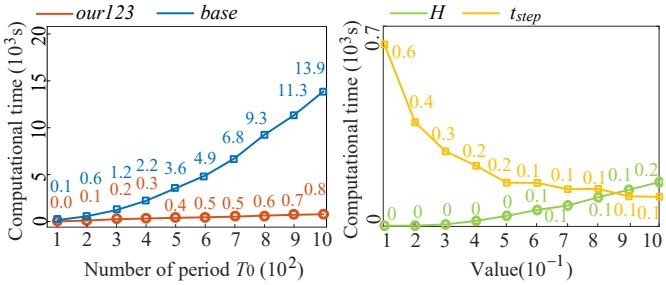


Fig. 11. Computational time. (a) Comparison between with MPC and without MPC. (b) Sensitivity analysis of $H$ and $t_{step}$.

Performance under time-varying VENs is evaluated with a period length of 100 s and a total of 800 periods $T_0$. $p^{trans}$=0.6 and $n^{trans}$=1. The proposed methods include *our*2 alone and a combined strategy *our*12, which integrates graph reduction and redundancy elimination. These variants are compared to assess their individual and joint effects. Evaluation metrics are consistent with those in Part A.

Fig. 9 presents the results. The upper subplot shows $|\mathcal{V}|$, $|\mathcal{E}|$, and $|\mathcal{R}|$, and optimization cost for both the baseline and proposed methods. Network metrics increase markedly, indicating substantial structural variability across instances. The strategy *our2* matches the baseline in both cost and feasibility, confirming that redundancy elimination preserves solution validity. The strategy *our12* leads to 4% infeasibility due to the aggressive pruning, reflecting a trade-off between complexity reduction and feasibility. The middle subplot shows consistent reductions in node and arc counts under *our2* and *our12*, with greater improvements in large-scale cases. The bottom subplot reveals significant runtime savings in both graph construction and LP solving. Overall, *our2* enhances scalability without loss of accuracy, while *our12* offers further gains at a modest feasibility cost.

Table IV summarizes the statistics of time-varying VENs. Network metrics grow substantially across segments. For example, $|\mathcal{R}|$ increases from 26±15 in cases 1~33 to 1619±1631 in cases 67~100, reflecting pronounced variation

TABLE VI
AVERAGE PERFORMANCE FOR STOCHASTIC CONDITIONS

| Indicator | *base* | *our3* | *DRO* | *Column decomposition* | *Heuristic* | *ARIMA* | *Prophet* | *Boost* | *our123* |
|---|---|---|---|---|---|---|---|---|---|
| Transmission loss ($\times 10^4$) | 2.2±0.0 | 1.5±0.2 | 3.0±0.0 | 2.3±0.0 | 2.3±0.0 | 1.7±0.2 | 1.6±0.1 | 1.5±0.2 | 1.5±0.2 |
| Over-supply ratio (%) | 4.4±4.6 | 9.3±3.7 | 38.5±9.2 | 4.4±4.6 | 51.1±5.3 | 25.7±12.1 | 31.6±2.3 | 12.5±1.9 | 9.3±3.7 |
| Unmet demand ratio (%) | 6.2±8.1 | 0.9±1.0 | 0.3±0.4 | 6.4±8.1 | 79.0±0.9 | 0.3±0.4 | 0.0±0.0 | 1.0±1.3 | 0.9±1.0 |
| Flow violation ratio (%) | 5.5±2.1 | 0.0±0.0 | 5.4±3.6 | 5.8±3.4 | 0.3±0.1 | 0.0±0.0 | 0.0±0.0 | 0.0±0.0 | 0.0±0.0 |
| Total delivered energy ($\times 10^5$) | 2.1±0.0 | 2.4±0.3 | 2.7±0.0 | 2.1±0.0 | 1.5±0.0 | 2.7±0.4 | 2.7±0.1 | 2.4±0.2 | 2.4±0.3 |
| Fairness index | 98.0±3.4 | 99.9±0.1 | 99.8±0.1 | 98.0±3.4 | 84.7±0.5 | 99.9±0.1 | 99.9±0.0 | 99.9±0.0 | 99.9±0.1 |
| Time (s) | 3014.3±23.5 | 280.9±7.4 | 3027±53.7 | 2994.1±23.5 | 73.9±0.6 | 144.9±2.1 | 250.7±11.0 | 183.2±5.1 | 125.2±1.2 |

TABLE VII
95-TH PERCENTILE PERFORMANCE FOR STOCHASTIC CONDITIONS

| Indicator | *base* | *our3* | *DRO* | *Column decomposition* | *Heuristic* | *ARIMA* | *Prophet* | *Boost* | *our123* |
|---|---|---|---|---|---|---|---|---|---|
| Transmission loss ($\times 10^4$) | 2.2 | 1.7 | 3.0 | 2.2 | 2.3 | 2.1 | 1.7 | 1.9 | 1.7 |
| Over-supply ratio (%) | 16.0 | 14.2 | 49.4 | 15.8 | 61.1 | 38.6 | 35.9 | 17.8 | 14.2 |
| Unmet demand ratio (%) | 27.7 | 3.0 | 0.9 | 27.8 | 81.8 | 0.9 | 0.0 | 3.1 | 3.0 |
| Flow violation ratio (%) | 9.5 | 0.0 | 7.8 | 14.2 | 0.6 | 0.0 | 0.0 | 0.0 | 0.0 |
| Total delivered energy ($\times 10^5$) | 2.1 | 2.9 | 2.7 | 2.1 | 1.5 | 3.3 | 2.7 | 3.0 | 2.9 |
| Fairness index | 99.9 | 99.9 | 99.9 | 99.9 | 86.1 | 99.9 | 99.9 | 99.9 | 99.9 |

in network complexity. Table V compares the optimization results. In case 67~100, the baseline method yields an average of 158,013 nodes, which is reduced to 75,350 by *our2* and further to 34,766 by *our12*, corresponding to reductions of 52.3% and 78.0%. A similar trend is observed in $|\mathcal{A}|$, decreasing from 301,069 in the baseline to 223,262 with *our2* and 103,633 with *our12*, representing reductions of 25.9% and 65.6%. For computation time, the baseline requires 1,312.5 s, with 1,310.1 s for graph construction and 2.4 s for LP solving. Using *our2* alone reduces the total time to 1,015.1 s, representing a 22.6% reduction. For *our12*, the total computation time drops further to 264.3 s, achieving an overall reduction of 79.9%. These results confirm the effectiveness of our approach in significantly accelerating computation under large-scale, time-varying VENs.

### C. Performance of Long-Time-Horizon Energy Routing

This section evaluates the proposed rolling-horizon MPC method using the case of area ID 01001, with $|\mathcal{V}|$=12, $|\mathcal{E}|$=52, and $|\mathcal{R}|$=134. The settings are: $H$=57, $t_{step}$=47, $p^{trans}$=0.5, $n^{trans}$=3, and $\lambda$=0.2. The LSTM models share the same structure, consisting of a sequence input layer, one LSTM layer with 256 hidden units, a fully connected layer, and a regression output layer, and Adam optimizer is applied. Fig. 10 illustrates the LSTM model performance. Fig. 10(a) shows training loss convergence with test MSE below 0.06, confirming model accuracy. Fig. 10 (b) to (d) compare predictions and true values under random, and random plus uncertain settings. The latter introduces irregularities due to distribution-free noise. While predictions are accurate to some extent, directly using the predicted values may lead to overestimation, because it may exceed the real values. By introducing residual bands that covers the true values, the predictions can be shifted downward, enabling more conservative computation and hence enhancing the robustness.

Firstly, computational time of different parameters are presented in Fig. 11. The computation time grows rapidly of *base*, exceeding 13,000 s at 1000 periods. In contrast, *our123* maintains runtime under 1000 s across all cases, achieving over 90% reduction. This indicates that our method is well-suited for long-horizon energy routing. Fig. 11(b) illustrates the impact of $H$ and $t_{step}$ under $T_0$=200. Both $H$ and $t_{step}$ are represented as proportions: $H$ is set as a fraction of the maximum allowable window, and $t_{step}$ is defined relative to $H$. $H$ cannot be arbitrarily set as large as the entire planning horizon; instead, it is constrained by the travel time differences across various paths to preserve the temporal dynamics and sequencing of vehicle movements. Increasing $H$ from 0.1 to 1.0, computational time increases from 600 s to 1200 s, increasing $t_{step}$ reduces computation time sharply from 1,200 s to 100 s. Therefore, from the perspective of runtime alone, a larger $H$ and $t_{step}$ are preferable. This confirms that fine-grained rolling windows improve efficiency.

Next, for a more comprehensive evaluation, several metrics are selected, including: (1) Transmission loss, the primary minimization objective; (2) Over-supply ratio: the proportion of excess energy delivered beyond demand; (3) Unmet-demand ratio: the fraction of unmet energy demand; (4) Flow violation ratio: the extent to which flow exceeds arc capacity constraints. (5) Total deliver energy: the total amount of energy successfully delivered to demand nodes; (6) Fairness index: the normalized Jain's fairness index, which quantifies the equity of energy distribution among demand nodes. (7) Testing time.

$$fairness = \left(\sum_{i\in\mathcal{V}_{\text{dem}}} e_i^{deliver}/B_i\right)^2 \Big/ |\mathcal{V}_{\text{dem}}| \sum_{i\in\mathcal{V}_{\text{dem}}} \left(e_i^{deliver}/B_i\right)^2 \tag{37}$$

where $e_i^{deliver}$ denotes the total deliver energy at demand node $i$.

The following methods are selected for comparison. For methods that cannot handle uncertainty and randomness, the statistical expectations under stochastic scenarios are used as deterministic inputs. (1) *base* [39]: full time-expanded min-cost flow method, using the same system parameters as the proposed approach. (2) *our3*: only LSTP-MPC is applied, without the proposed graph reduction and construction methods. (3) *DRO* [48]: A robust method using means and standard deviations of historical flow and supply samples. A 90% confidence interval is applied for conservative adjustment. (4) *Column decomposition* [49]: An LP acceleration method that starts with 25% of variables and iteratively adds up to 50% more with negative reduced cost

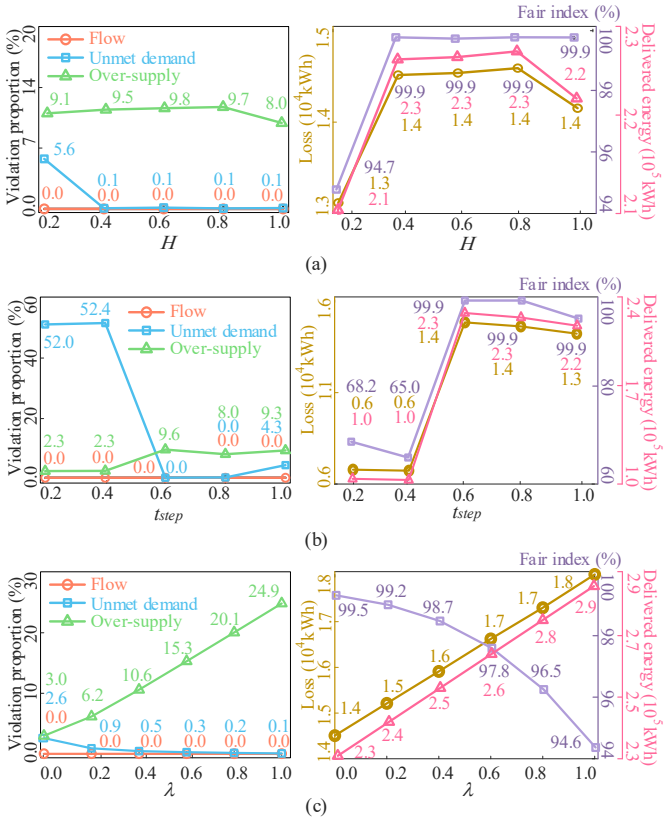


Fig. 12. Sensitivity analysis of violations, loss, fair index, and delivered energy of different parameters. (a) $H$. (b) $t_{step}$. (c) $\lambda$.

over a maximum of 10 iterations or until convergence. The reduced cost threshold is −0.001. (5) *Heuristic* [50]: A demand-driven model free method. The direct feasible paths from supply to demand are iteratively searched with the constraints of available energy and capacities. The algorithm stops until demand satisfaction or infeasibility. Also, three forecasting variants are employed as alternative models, with other components unchanged compared with *our123* (full method). Similar to LSTM, each model is trained using data from the current window to predict the next, and separate models are built for supply/demand and flow, respectively. (6) *ARIMA* [51]: Uses first-order differencing ($d$=1), no autoregressive ($p$=0) or moving average ($q$=0) terms. (7) *Prophet* [52]: Applied with default settings in the python environment, which include piecewise linear trend modeling and additive seasonality, without manual hyperparameter tuning. (8) *Boost* [53]: For each supply/demand nodes and routes, a set of gradient boosting regression models are trained with the LSBoost method to predict the average value over the next window, used with 100 learning cycles, a maximum of 10 splits per tree, and a learning rate of 0.1.

Both average performance (mean ± 95% confidence interval) and 95-th percentile performance are adopted to present numerical results, thus providing a comprehensive assessment of solution quality and worst-case performance. Results under ten trials are shown in Tables VI and VII. For average performance, forecasting-based methods, including *ARIMA*, *Prophet*, *Boost*, *our3*, and *our123*, show clear advantages in reducing transmission loss. Without forecasting, the loss remains above 2.0. Among all methods, *our123* and *Boost* achieve the best performance, with loss around 1.5. Also, forecasting methods also perform well in minimizing flow violations and maintaining a fairness index close to 99.9%, because most demands are satisfied. In terms of computation time, methods involving MPC significantly reduce runtime compared to full-scale optimization. For instance, our method requires only 125.2 s on average, while full computation takes around 3,000 s. The fastest method is the heuristic approach, as it is model-free and avoids constructing the VEN model. However, due to its greedy design, it suffers the worst performance in constraint violations. Therefore, in terms of average performance, our method achieves the best and most balanced results in both efficiency and effectiveness. As for the 95-th percentile performance, *our123* maintains low values in transmission loss, over-supply, and unmet demand, demonstrating strong reliability even in worst-case scenarios, with consistently high fairness and zero violations. Therefore, the effectiveness and robustness of our method are validated.

Fig. 12 presents a sensitivity analysis of violations, loss, fair index, and delivered energy of different parameters. The range of $H$ and $t_{step}$ have been previously introduced, where $H$ denotes a percentage of the maximum allowable window length, and $t_{step}$ is a fraction of $H$. As shown in Fig. 12(a), when $H$ is set small, e.g., 0.2, the system performance is unsatisfied, with unmet demand reaching 5.6%. Although energy loss appears minimal in this case, it results from insufficient energy delivery. The fairness index is also low, indicating imbalanced allocation. Thus, a larger $H$ improves performance by enabling better planning. In our case, $H$=1.0 achieves the most balanced performance. The setting of $H$ also needs to consider computational cost and the time granularity imposed by vehicle travel durations. Regarding $t_{step}$ in Fig. 12(b), performance degrades when $t_{step}$ is smaller than 0.4. Limited look-ahead restricts the decision-making horizon, reducing solution quality. A step size between 0.8 and 1.0 yields the most balanced results, with low unmet demand and reasonable energy loss. In Fig. 12(c), our method achieves strong performance in flow constraints with wide range of $\lambda$. Also, the unmet demand is kept below 3%, indicating effective demand satisfaction. As $\lambda$ increases, the decision becomes more conservative in supply to avoid shortages, leading to notable rise in over-supply. The model tends to deliver more energy, resulting in higher transmission loss with increased $\lambda$. Meanwhile, fairness declines slightly, indicating a trade-off between robustness and equity in energy distribution. In our case, a value between 0.0~0.2 offers a good balance.

## V. Discussion

To handle extreme events such as sudden surges or drops in supply, demand, and flow, a dual mechanism combining prediction and robust optimization is proposed. Firstly, LSTM models are used to forecast future trends of key variables, allowing the model to detect risk early and plan energy routing in advance. Secondly, a residual correction band is constructed based on historical prediction errors, with a tunable parameter $\lambda$ controlling the level of conservativeness. The proposed strategy combines data-driven learning with robustness consideration, providing both anticipatory and resilient decision-making.

The effectiveness of the proposed method relies heavily on the predictive capability of the LSTM model. In the ideal case, if predictions are perfect, the residual correction band becomes unnecessary, and the model can achieve optimal performance. However, in practical applications, prediction errors are inevitable. When forecasted values consistently overestimate the actual values, a larger residual correction band must be introduced. Conversely, if predictions frequently underestimate actual values, the system inherently adopts a conservative stance. Notably, when both overestimation and underestimation errors are large, the correction band may become excessively wide. If the prediction is already lower than the actual value, applying this downward adjustment makes it even more conservative, leading to unnecessary caution and reduced performance. In general, larger prediction errors result in greater performance loss. Therefore, to address this limitation, a potential improvement is to dynamically adjust $\lambda$ based on real-time system feedback and the probability of extreme events, ensuring system robustness while avoiding performance degradation caused by excessive conservativeness under normal conditions.

Also, there are several limitations that warrant further study. Firstly, the proposed graph reduction method is less effective with short and sparse routes. Enhancing its adaptability across diverse conditions is an important direction. Secondly, practical deployment of VEN relies on effective infrastructure design, such as the placement and storage configuration, which serve as the physical foundation. Thirdly, user participation and operator profitability are critical to real-world adoption. Frequent charging and discharging may raise concerns about battery degradation. The system must balance user compensation with the sustainable operation. Future work should focus on enhancing the generalizability and scalability of the algorithm, optimizing infrastructure layout, and designing incentive-compatible business models to support the practical deployment of VEN.

## VI. Conclusion

This study presents a robust and scalable energy routing optimization framework for stochastic VENs aimed at minimizing transmission loss, under both time-invariant and time-varying conditions. This is achieved through an integrated methodology that combines structural simplification, stochastic-aware chance-robust optimization, and model predictive coordination, ensuring computational efficiency and routing reliability across diverse operational VEN scenarios. The code is available at https://github.com/tytzh/VEN.

Extensive evaluations on 100 real-world U.S. cases confirm the model's effectiveness in reducing computational complexity, managing randomness and uncertainty, and enabling long-horizon planning. Notably, in large-scale instances, the proposed methods deliver significant improvements. In time-invariant scenarios, node and arc counts are reduced by approximately 49.8% and 49.7%, respectively, alongside an 83.1% reduction in total computation time. In time-varying settings, the benefits are further amplified, with reductions of up to 78.0% in nodes, 65.6% in arcs, and 79.9% in runtime. Under stochastic conditions, the proposed robust optimization framework lowers average and worst-case constraint violations, while incurring only a modest cost increase. For long-horizon optimization, the predictive control scheme reduces runtime by over 90%, and the LSTM-MPC achieves balanced performances in the aspect of transmission loss, fairness index, and violation-related indicators. These results underscore the practical value of the proposed framework to support the integration of VENs into next-generation IoT-enabled smart grid systems.

## References


[1] F. Ahsan *et al.*, "Data-driven next-generation smart grid towards sustainable energy evolution: techniques and technology review," *Protection and Control of Modern Power Systems,* vol. 8, no. 3, pp. 1-42, Jul. 2023.

[2] S. Alsubai, A. Alqahtani, A. Alanazi, and M. Bhatia, "Digital-twin-inspired IoT-assisted intelligent performance analysis framework for electric vehicles," *IEEE Internet of Things Journal,* vol. 11, no. 10, pp. 18880-18887, Feb. 2024.

[3] M. S. Alam, F. S. Al-Ismail, A. Salem, and M. A. Abido, "High-level penetration of renewable energy sources into grid utility: challenges and solutions," *IEEE Access,* vol. 8, pp. 190277-190299, Oct. 2020.

[4] S. M. A. A. Abir, A. Anwar, J. Choi, and A. S. M. Kayes, "IoT-enabled smart energy grid: applications and challenges," *IEEE Access,* vol. 9, pp. 50961-50981, Mar. 2021.

[5] A. Y. S. Lam, K. C. Leung, and V. O. K. Li, "Vehicular energy network," *IEEE Transactions on Transportation Electrification,* vol. 3, no. 2, pp. 392-404, Jan. 2017.

[6] P. Yi, T. Zhu, B. Jiang, B. Wang, and D. Towsley, "An energy transmission and distribution network using electric vehicles," in *2012 IEEE International Conference on Communications (ICC)*, 2012, pp. 3335-3339.

[7] K. T. Chau, *Pure Electric Vehicles. In Alternative Fuels and Advanced Vehicle Technologies for Improved Environmental Performance – Towards Zero Carbon Transportation*. (Ed.) R. Folkson. Woodhead Publishing, 2014, pp. 655-684.

[8] W. Liu, K. T. Chau, C. C. T. Chow, and C. H. T. Lee, "Wireless energy trading in traffic internet," *IEEE Transactions on Power Electronics,* vol. 37, no. 4, pp. 4831-4841, Oct. 2022.

[9] Y. Cao, C. Zhao, Y. Zhang, and Y. Jin, "Optimizing resource allocation and energy efficiency in vehicle mobile-edge computing with blockchain integration," *IEEE Internet of Things Journal,* vol. 12, no. 18, pp. 36807-36818, May. 2025.

[10] B. Bibak and H. Tekiner-Moğulkoç, "A comprehensive analysis of vehicle to grid (V2G) systems and scholarly literature on the application of such systems," *Renewable Energy Focus,* vol. 36, pp. 1-20, Mar. 2021.

[11] C. Liu, K. T. Chau, D. Wu, and S. Gao, "Opportunities and challenges of vehicle-to-home, vehicle-to-vehicle, and vehicle-to-grid technologies," *Proceedings of the IEEE,* vol. 101, no. 11, pp. 2409-2427, Jul. 2013.

[12] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, "Internet of things: a survey on enabling technologies, protocols, and applications," *IEEE Communications Surveys & Tutorials,* vol. 17, no. 4, pp. 2347-2376, Jun. 2015.

[13] D. Emad, O. Abdel-Rahim, W. Rohouma, and S. M. Abdelkader, "Energy internet: state of the art and challenges," *IEEE Access,* vol. 12, pp. 143131-143148, Aug. 2024.

[14] A. A. K. Majhi and S. Mohanty, "A comprehensive review on internet of things applications in power systems," *IEEE Internet of Things Journal,* vol. 11, no. 21, pp. 34896-34923, Aug. 2024.

[15] Y. Cao, Y. Wang, D. Li, and X. Chen, "Joint routing and wireless charging scheduling for electric vehicles with shuttle services," *IEEE Internet of Things Journal,* vol. 10, no. 16, pp. 14810-14819, Oct. 2023.

[16] *Global EV Outlook 2024* [Online]. Accessed: Apr. 2025. Available: https://www.iea.org/reports/global-ev-outlook-2024

[17] W. Liu, K. T. Chau, X. Tian, H. Wang, and Z. Hua, "Smart wireless power transfer — opportunities and challenges," *Renewable and Sustainable Energy Reviews,* vol. 180, p. 113298, Jul. 2023.

[18] Z. Xue, W. Liu, C. Liu, and K. T. Chau, "Critical review of wireless charging technologies for electric vehicles," *World Electric Vehicle Journal,* vol. 16, no. 2, Jan. 2025.

[19] S. Thangavel, D. Mohanraj, T. Girijaprasanna, S. Raju, C. Dhanamjayulu, and S. M. Muyeen, "A comprehensive review on electric vehicle: battery management system, charging station, traction motors," *IEEE Access,* vol. 11, pp. 20994-21019, Feb. 2023.

[20] Y. Tang, W. Liu, K. T. Chau, Y. Hou, and J. Guo, "Stochastic behavior modeling and optimal bidirectional charging station deployment in EV energy network," *IEEE Transactions on Intelligent Transportation Systems,* pp. 1-17, Mar. 2025.

[21] W. Liu, T. Placke, and K. T. Chau, "Overview of batteries and battery management for electric vehicles," *Energy Reports,* vol. 8, pp. 4058-4084, Nov. 2022.

[22] R. Wang, J. Wu, Z. Qian, Z. Lin, and X. He, "A graph theory based energy routing algorithm in energy local area network," *IEEE Transactions on Industrial Informatics,* vol. 13, no. 6, pp. 3275-3285, Jun. 2017.

[23] R. Razi, M. C. Pham, A. Hably, S. Bacha, Q. T. Tran, and H. Iman-Eini, "A novel graph-based routing algorithm in residential multimicrogrid systems," *IEEE Transactions on Industrial Informatics,* vol. 17, no. 3, pp. 1774-1784, May. 2021.

[24] H. Hua *et al.*, "Carbon emission flow based energy routing strategy in energy internet," *IEEE Transactions on Industrial Informatics,* vol. 20, no. 3, pp. 3974-3985, Oct. 2024.

[25] P. Yi, T. Zhu, G. Lin, and Q. Zhang, "Routing renewable energy using electric vehicles in mobile electrical grid," in *2013 IEEE 10th International Conference on Mobile Ad-Hoc and Sensor Systems*, 2013, pp. 19-27.

[26] A. Y. S. Lam and V. O. K. Li, "Opportunistic routing for vehicular energy network," *IEEE Internet of Things Journal,* vol. 5, no. 2, pp. 533-545, Sept. 2018.

[27] T. Fu, C. Wang, and N. Cheng, "Deep-learning-based joint optimization of renewable energy storage and routing in vehicular energy network," *IEEE Internet of Things Journal,* vol. 7, no. 7, pp. 6229-6241, Jan. 2020.

[28] C. C. T. Chow, A. Y. S. Lam, W. Liu, and K. T. Chau, "Multisource–multidestination optimal energy routing in static and time-varying vehicular energy network," *IEEE Internet of Things Journal,* vol. 9, no. 24, pp. 25487-25505, Aug. 2022.

[29] S. Arnborg, B. Courcelle, A. Proskurowski, and D. Seese, "An algebraic theory of graph reduction," *Journal of the ACM,* vol. 40, no. 5, pp. 1134–1164, 1993.

[30] M. a. G. Hashemi, S. and Ni, J. and Fan, W. and Prakash, B. A. and Jin, W., "A comprehensive survey on graph reduction: sparsification, coarsening, and condensation," presented at the Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence (IJCAI), Jeju, Korea, 2024, 2024.

[31] G. a. D. Desaulniers, J. and Solomon, M. M., *Column Generation*. Springer Science & Business Media, 2006.

[32] S. Sethi and G. Sorger, "A theory of rolling horizon decision making," *Annals of Operations Research,* vol. 29, no. 1, pp. 387-415, Dec. 1991.

[33] K. Huang, K. Wei, F. Li, C. Yang, and W. Gui, "LSTM-MPC: a deep learning based predictive control method for multimode process control," *IEEE Transactions on Industrial Electronics,* vol. 70, no. 11, pp. 11544-11554, Dec. 2023.

[34] E. Gaetan, L. Giarré, and P. Falcone, "Scenario-tree model predictive control for vehicle interactions in highway setting," *IEEE Control Systems Letters,* vol. 8, pp. 1162-1167, May. 2024.

[35] J. Li, H. Zeng, and Y. Xie, "Unit commitment with joint chance constraints in multi-area power systems with wind power based on partial sample average approximation," *Journal of Modern Power Systems and Clean Energy,* vol. 13, no. 1, pp. 241-252, Jul. 2025.

[36] R. K. Ahuja, T. L. Magnanti, and J. B. Orlin, *Network Flows: Theory, Algorithms and Applications*. Upper Saddle River, NJ: Prentice-Hall, 1993.

[37] L. G. Khachiyan, "A polynomial algorithm in linear programming," *Soviet Mathematics Doklady,* vol. 20, no. 1, pp. 191–194, 1979.

[38] N. Karmarkar, "A new polynomial-time algorithm for linear programming," *Combinatorica,* vol. 4, no. 4, pp. 373–395, 1984.

[39] S. Abdi, F. Baroughi, and B. Alizadeh, "The minimum cost flow problem of uncertain random network," *Asia-Pacific Journal of Operational Research,* vol. 35, no. 03, p. 1850016, Apr. 2018.

[40] Y. Jiang, T. Ortmeyer, and M. Fan, "Data-driven fast uncertainty assessment of distribution systems with correlated EV charging demand and renewable generation," *IEEE Transactions on Sustainable Energy,* vol. 14, no. 3, pp. 1446-1456, Jan. 2023.

[41] Z. Liu, C. Lyu, Z. Wang, S. Wang, P. Liu, and Q. Meng, "A Gaussian-process-based data-driven traffic flow model and its application in road capacity analysis," *IEEE Transactions on Intelligent Transportation Systems,* vol. 24, no. 2, pp. 1544-1563, Jan. 2023.

[42] J. B. Rawlings and D. Q. Mayne, *Model Predictive Control: Theory and Design*. Madison, WI: Nob Hill Publishing, 2009.

[43] B. Liu, *Uncertainty Theory: A Branch of Mathematics for Modeling Human Uncertainty*, 3rd ed. (Studies in Fuzziness and Soft Computing). Springer, 2010.

[44] C. R. J. D. Y. L. Y. Li, "A large-scale benchmark dataset for commuting origin-destination matrix generation," *arXiv preprint arXiv:2407.15823,* Jul. 2024.

[45] *Performance Measurement System* [Online]. Accessed: Apr. 2025. Available: https://pems.dot.ca.gov/

[46] K. Zhou, D. Hu, R. Hu, and J. Zhou, "High-resolution electric power load data of an industrial park with multiple types of buildings in China," *Scientific Data,* vol. 10, no. 1, p. 870, Dec. 2023.

[47] *Wind Power Generation Data* [Online]. Accessed: Apr. 2025. Available: https://www.kaggle.com/datasets/jorgesandoval/wind-power-generation?resource=download&select=TransnetBW.csv

[48] R. Hamed and M. Sanjay, "Frameworks and results in distributionally robust optimization,", *Open Journal of Mathematical Optimization,* vol. 3, pp. 1-85, Jul. 2022.

[49] G. Desaulniers, J. Desrosiers, and M. M. Solomon, *Column generation*. Springer Science & Business Media, 2006.

[50] V. Sharma and A. K. Tripathi, "A systematic review of meta-heuristic algorithms in IoT based application," *Array,* vol. 14, p. 100164, Jul. 2022.

[51] S. L. Ho and M. Xie, "The use of ARIMA models for reliability forecasting and analysis," *Computers & Industrial Engineering,* vol. 35, no. 1, pp. 213-216, Oct. 1998.

[52] S. J. Taylor and B. Letham, "Forecasting at scale," *The American Statistician,* vol. 72, no. 1, pp. 37-45, Jul. 2018.

[53] R. E. Schapire, "Theoretical views of boosting and applications," presented at the Algorithmic Learning Theory, Berlin, Heidelberg, May. 1999.

**Yao Tang** (Student Member, IEEE) received the B.Eng. degree from Beijing Forestry University, China in 2019, and received the M.Eng. degree from Hunan University, China in 2021. She is currently working toward the Ph.D. degree in electrical and electronic engineering at the Department of Electrical and Electronic Engineering, the University of Hong Kong. Her current research interests include electric vehicle, optimization theory and artificial intelligence for industrial applications.

**Wei Liu** (Senior Member, IEEE) received the B.Eng. and M.Eng. degrees in electrical engineering from China University of Petroleum, Qingdao, China, and a Ph.D. degree in electrical and electronic engineering from The University of Hong Kong (HKU), Hong Kong, China, in 2014, 2017, and 2021, respectively.

He is currently an Assistant Professor at the Research Centre for Electric Vehicles and the Department of Electrical and Electronic Engineering at the Hong Kong Polytechnic University (PolyU). Dr. Liu served as a Postdoctoral Fellow and a Research Assistant Professor from 2021 to 2023, and he is now an Honorary Assistant Professor at the Department of Electrical and Electronic Engineering, HKU. He also worked as a Visiting Researcher with Nanyang Technological University, Singapore (NTU), in 2019. His research interests include wireless power transfer, power electronics, biomedical power electronics, and electric vehicle technologies.

Dr. Liu was the recipient of the Power Engineering Prize from HKU, the Excellent Paper Award, and the Best Presentation Award from international conferences in the area of Electric Vehicles and Transportation Electrification. He is also a Guest Associate Editor of *IEEE Journal of Emerging and Selected Topics in Power Electronics (JESTPE)*, Guest Editor of international journals, and Session Chair of international conferences.

**Yunhe Hou** (Fellow, IEEE) received the B.E. and Ph.D. degrees in electrical engineering from the Huazhong University of Science and Technology, Wuhan, China, in 1999 and 2005, respectively. He was a Postdoctoral Research Fellow with Tsinghua University, Beijing, China, from 2005 to 2007, and a Postdoctoral Researcher with Iowa State University, Ames, IA, USA, and the University College Dublin, Dublin, Ireland, from 2008 to 2009. He was also a Visiting Scientist with the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA, USA, in 2010. He has been a Guest Professor with the Huazhong University of Science and Technology, China, since 2017, and an Academic Adviser of China Electric Power Research Institute since 2019. He joined as a Faculty with the University of Hong Kong, Hong Kong, in 2009, where he is an Associate Professor in the Department of Electrical and Electronic Engineering. He was an Associate Editor of IEEE TRANSACTIONS ON SMART GRID from 2016 to 2021. He is currently an Associate Editor of IEEE TRANSACTIONS POWER SYSTEMS and the Journal of Modern Power Systems and Clean Energy.

**K. T. Chau** (Fellow, IEEE) received the B.Sc. (Eng.), M.Phil., and Ph.D. degrees in electrical and electronic engineering from The University of Hong Kong, Hong Kong, in 1988, 1991, and 1993, respectively. Currently, he serves as Chair Professor of Electrical Energy Engineering at the Research Centre for Electric Vehicles and Department of Electrical and Electronic Engineering, The Hong Kong Polytechnic University. His research interests include electric and hybrid vehicles, power electronics and drives, and renewable energies. He is the author of nine books and more than 350 journal papers.

Prof. Chau is a Fellow of the Institution of Engineering and Technology (IET), U.K., and of the Hong Kong Institution of Engineers. He is a Chartered Engineer. He was the recipient of the Changjiang Chair Professorship from the Ministry of Education, China, and the Environmental Excellence in Transportation Award for Education, Training, and Public Awareness from the Society of Automotive Engineers International.