

Discovering EV Charging Site Archetypes Through Few Shot Forecasting

The First U.S.-Wide Study

Kshitij Nikhal
technikhal@alphagrid.ai

Luke Ackerknecht
luke@alphagrid.ai

Benjamin S. Riggan
briggan2@unl.edu

Phil Stahlfeld
phil@alphagrid.ai

MOTIVATION AND CONTRIBUTIONS

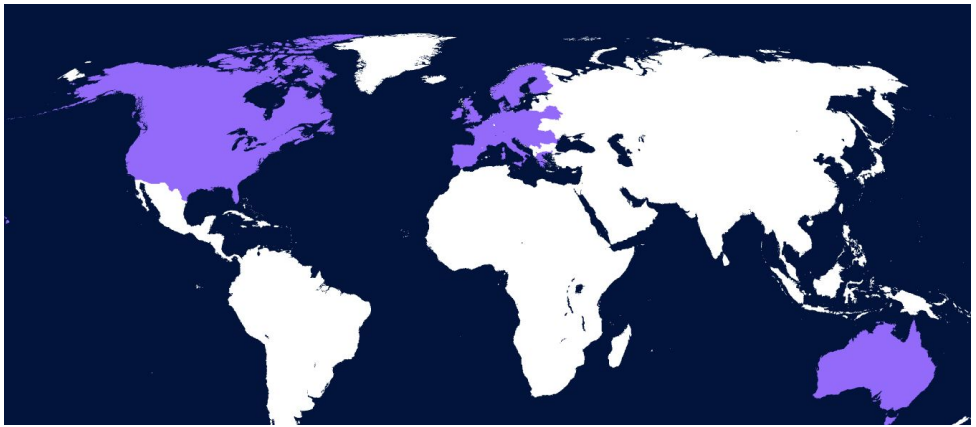
ELECTRIC VEHICLES (EVS) ARE CENTRAL TO ACHIEVING NET ZERO EMISSIONS

	Existing Research		*Contributions*
DATASET	Limited, small datasets	→	High resolution US Wide L3 dataset
INSIGHTS	Regional insights	→	Large-scale, high resolution analysis
GENERALIZATION	Poor generalization to new sites	→	Knowledge transfer through expert
ACCURACY	High Forecast Error Rates	→	Low Forecast Error Rates
INTERPRETATION	No semantic interpretation	→	Clustering to uncover archetypes

Accurate forecasting of EV charging demand informs infrastructure **planning**, grid **stability**, dynamic **pricing**, and climate **policy**.

DATASET

- The AGCharging dataset includes session data filtered to the top five US charging networks.
- Daily site energy consumption is derived by aggregating session energy (kWh).
- Test set is restricted to 28 days of operational history.
- A slice of this dataset will be released to promote research.



>90%
coverage across Americas, Europe, and Australia

400k+
Connectors in our system

100+
networks

50+
Countries

METHODOLOGY

1. Input time series dataset aggregated at a day level.

$$X = \{x_1, x_2, \ldots, x_n\}$$
$$x_i = \{(t_{ij}, y_{ij})\}_{j=1}^{T_i}$$

2. Characterize site-level demand into weekly utilization profile and catch22 features.

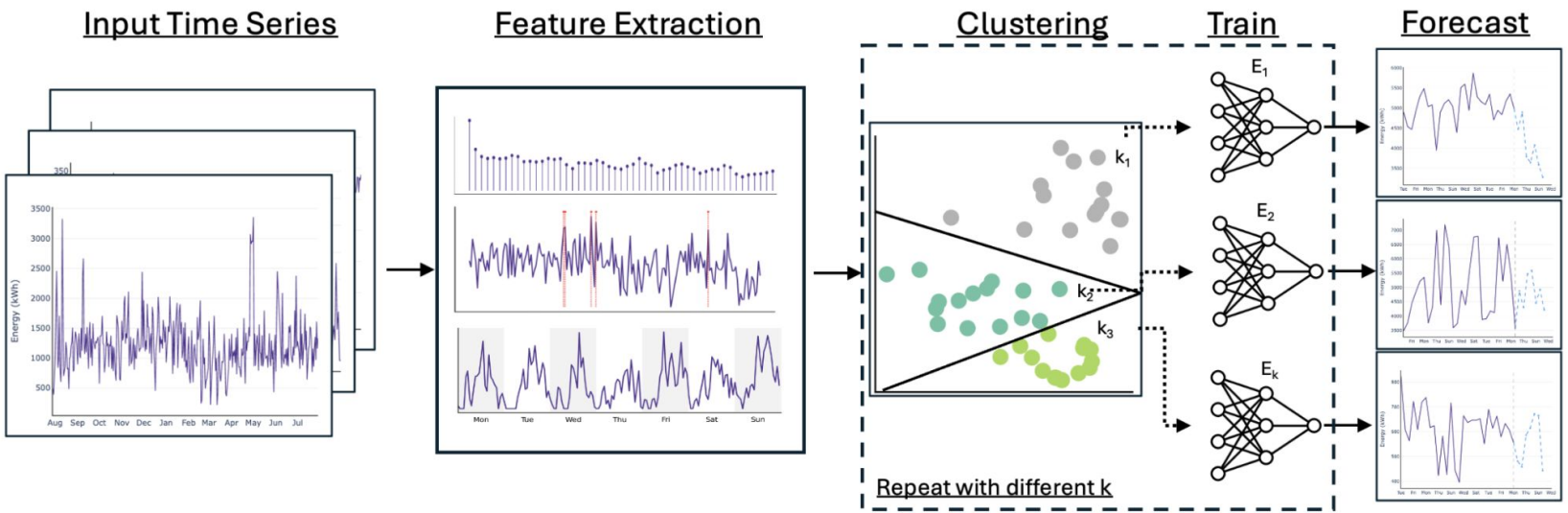
$$\Phi_i = \left[\, u_i,\ \phi(u_i) \,\right]$$

3. Cluster and train k-models for each cluster.

$$\arg\min_k \|\Phi_i - \mu_k\|_2^2$$
$$\Theta = \{\theta_1, \ldots, \theta_k\}$$

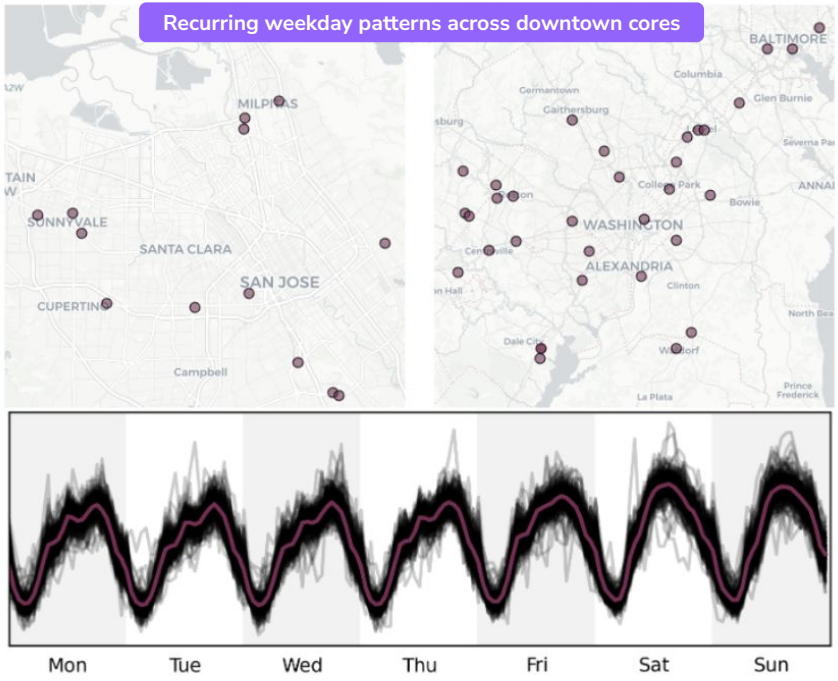
4. Forecast the next H days using past L days.

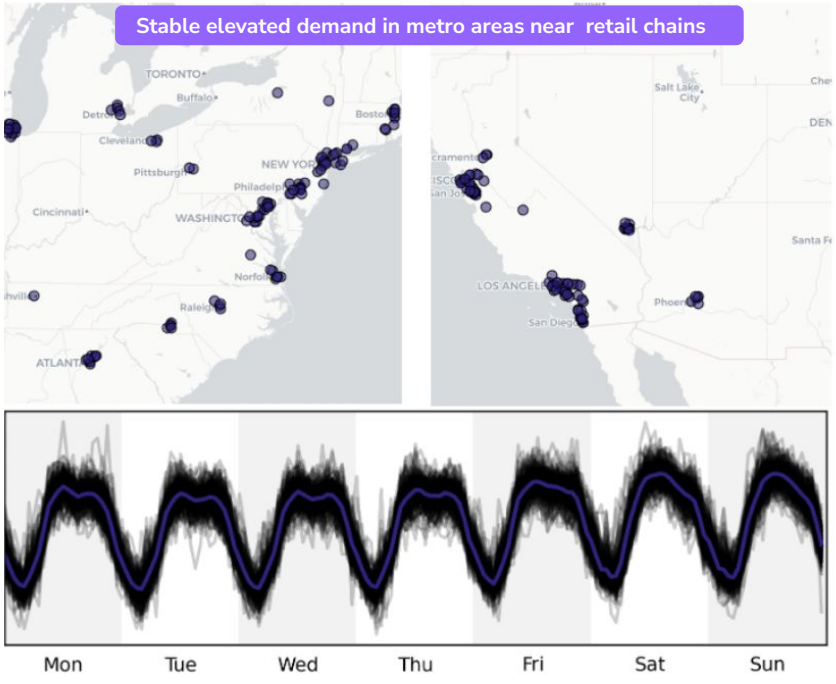
$$\hat{\mathbf{y}}_{t+1:t+H|t} = f_\theta(\mathbf{y}_{t-L+1:t})$$



DISCOVERED ARCHETYPES

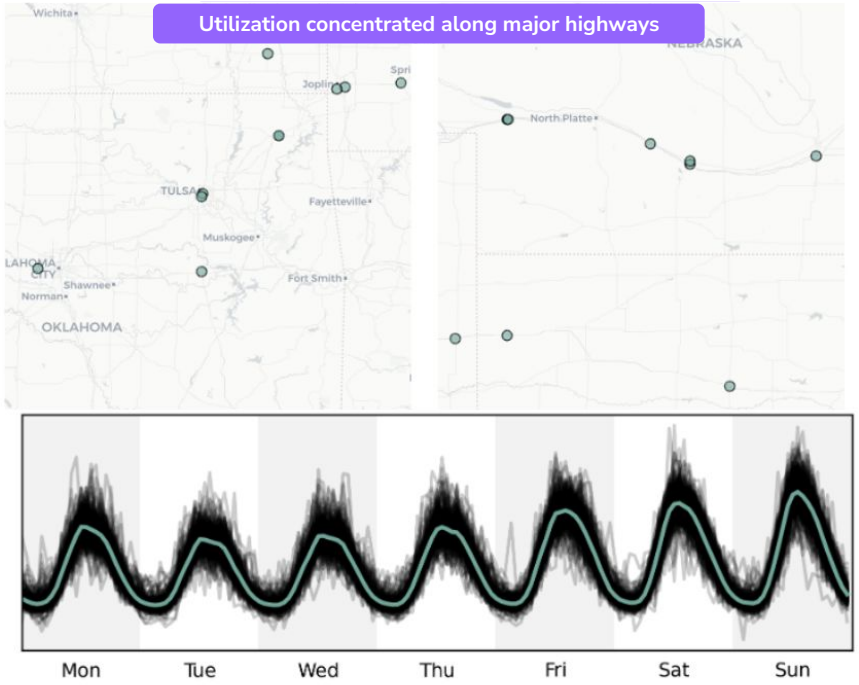
Archetypes with semantic labels mapped geographically

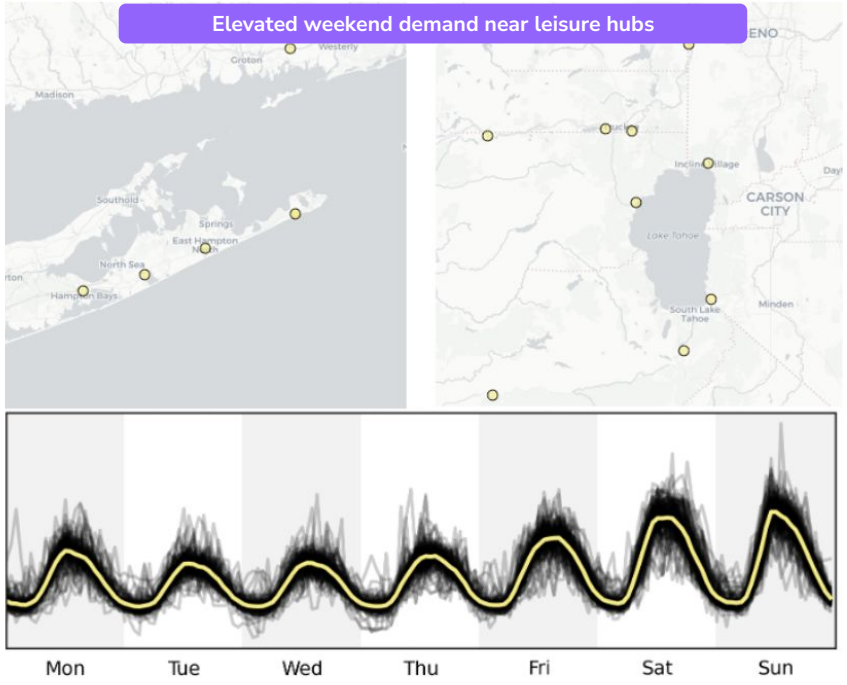




DISCOVERED ARCHETYPES

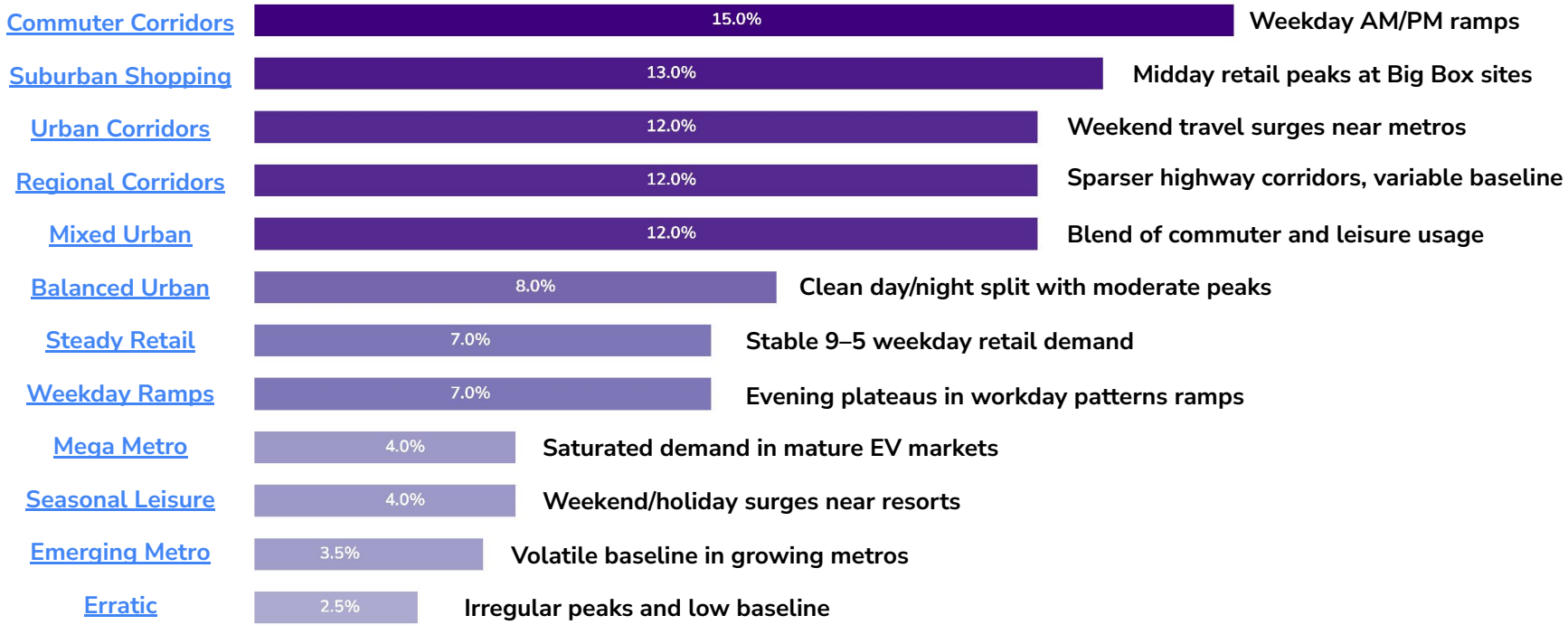
Archetypes with semantic labels mapped geographically





DISCOVERED ARCHETYPES

Archetypes and their share of sites in the US



See Appendix A.1 for complete geospatial distribution and demand profiles.

QUANTITATIVE RESULTS

Performance of global (G) vs. expert (E) models.

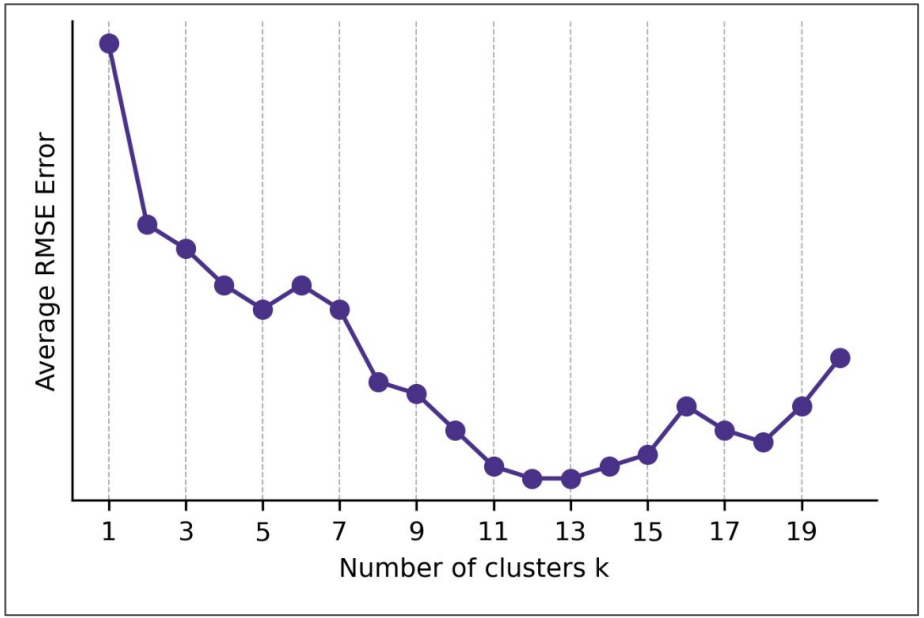
- Across all test sites, archetype-specific experts outperform the single global model in both sMAPE and RMSE.
- The largest gains appear in volatile clusters—corridors and vacation destinations—where specialized temporal structure really matters.
- This shows that specialization improves generalization: smaller, context-aware models beat a single global one.

Archetype (% of sites) and Description	sMAPE		RMSE	
	G	E	G	E
A1. Steady Retail (7%): Stable all-week 9AM-5PM demand (↑ 1.3σ), predominantly located near retail chains in cities such as Los Angeles, Chicago, Denver, and Washington DC.	16.15	**15.82**	**614**	616
A2. Urban Corridors (12%): Large weekend peaks (↑ 2.6σ) co-located with travel plazas near popular destinations and metros.	24.17	**20.76**	728	**654**
A3. Regional Corridors (12%): Similar peaks (↑ 2.1σ) to A2 but lower baseline and higher variance, reflecting sparser corridors.	30.23	**28.60**	638	**608**
A4. Mixed Urban (12%): Blend of commuter and leisure usage (↑ 1.5σ), observed in dense downtowns with diverse land use.	24.77	**22.87**	799	**735**
A5. Balanced Urban (8%): Similar to A1, but cleaner day/night split, moderate peaks (↑ 1.21σ), and deep troughs (↓ 1.75σ).	23.46	**21.80**	650	**608**
A6. Commuter Corridors (15%): Weekday AM/PM ramps exhibit smoother peaks (↑ 1.40σ) compared to sharper weekend peaks (↑ 1.9σ), reflecting work–home travel rhythms.	24.50	**23.23**	614	**575**
A7. Mega Metro (4%): Consistent, saturated (↑ 1.0σ) profile characteristic of large EV markets such as Southern California, reflecting large-scale adoption.	**12.30**	12.91	1126	**1005**
A8. Weekday Ramps (7%): Pronounced weekday ramps with evening plateaus (↑ 1.5σ), seen in San Francisco, Dallas and Seattle.	12.72	**8.69**	1232	**1131**
A9. Suburban Shopping (13%): Spread across most states with midday peaking (↑ 1.55σ), tied to grocery/Big Box stores.	17.86	**17.19**	829	**792**
A10. Emerging Metro (3.5%): Found in Houston, Atlanta, Orlando, and Dallas, where workplace, retail, and leisure usage combine but baseline demand remains volatile.(↓ 1.76σ and ↑ 1.21σ).	33.67	**32.89**	682	**664**
A11. Seasonal Leisure (4%): Stronger weekend/holiday surges (↑ 2.3σ) than weekdays (↑ 1.3σ), near resorts and vacation homes.	34.48	**31.73**	762	**705**
A12. Erratic (2.5%): Low baseline demand with irregular peaks (↑ 1.24σ) and troughs (↓ 0.94σ).	65.46	**61.16**	457	**377**

ABLATION STUDY

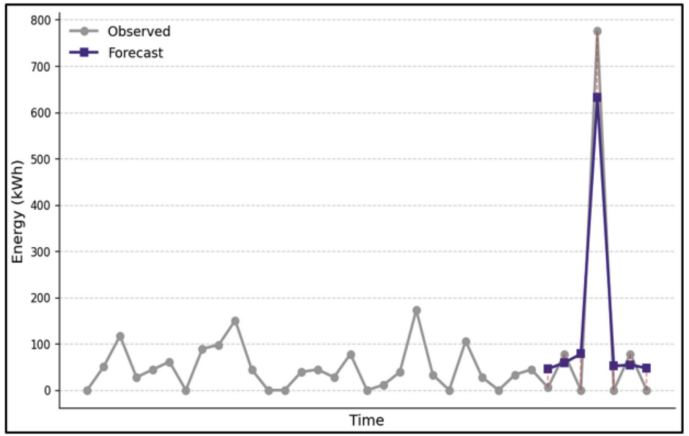
Trade-off between global and expert models.

- The global model (k=1) provides a strong baseline, while specialized models yield substantial gains up to an optimal k=12.
- Balancing cluster size is important with k > 12 reducing predictive power.
- Twelve distinct site archetypes capture dominant demand patterns and enable robust few-shot inference.

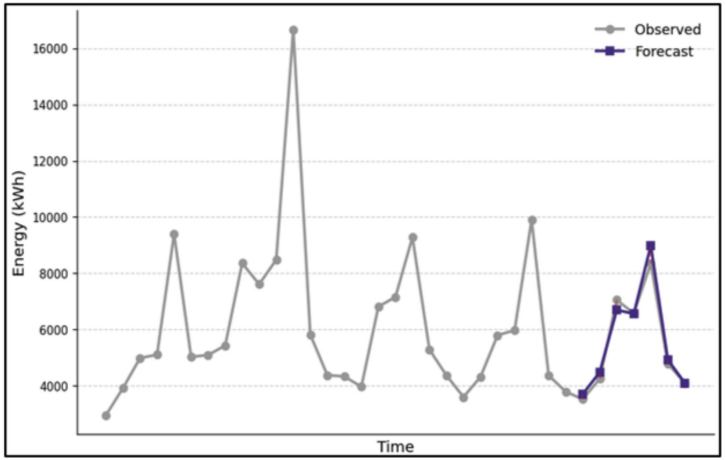


FORECASTING

Effectively transfers archetype-specific knowledge to forecast both peak surges and recurring demand trends.



Predicting surges with no history of surge data using knowledge transfer.



Predicting weekly routine trends with accurate energy forecast.

CONCLUSION AND FUTURE WORK

- We present the first nationwide assessment of the US fast-charging market, highlighting utilization patterns and site-level heterogeneity.
- Using forecast-guided evaluation, we identify the optimal number of archetypes and show that cluster-aware expert models outperform global models in few-shot inference.
- Beyond accuracy gains, these insights strengthen interconnection studies, guide incentives toward underserved areas, and inform underwriting through archetype membership.
- Future work will extend to longer-term horizons, soft clustering, and external signals (e.g., events).

THANK YOU!

technikhal@alphagrid.ai