Auditable Surge Planning: Network-Aware Causal Inference Meets Prescriptive Optimization

Sanjay Kumar Patnala & Vedika Sanjeev Lakhanpal

CML KDD 2025

Georgia Tech

Motivation - why tactical surges matter

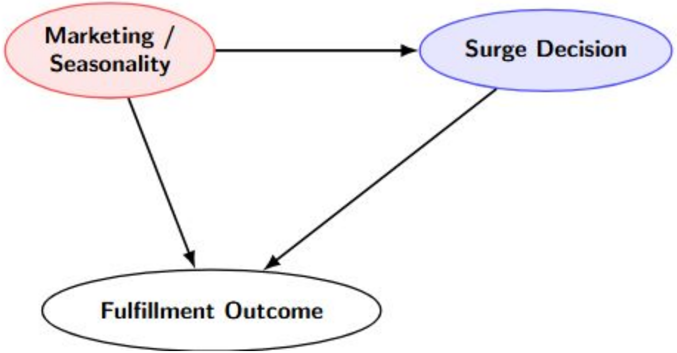
- Consumer-goods plants increasingly run short, 30-40 % “surge” runs to rescue service levels when demand spikes or inventory dips.
- Today these surges are triggered by **heuristics** – spreadsheet rules like “boost the SKU with the biggest forecast gap.”
- Result: ad-hoc firefighting → stockouts, premium freight, lost sales.
- We ask: **Can we learn the true impact of a surge and prescribe the optimal set of SKUs to surge next, under real plant constraints?**

Georgia Tech

Key Technical Challenges

Resource coupling (network spillover)

- A filler line shared by Sports Drink (A) and Iced Tea (B) means surging A can starve B.



Confounding (causal attribution)

- Surges coincide with promotions, seasonality, unlogged fixes.
- Naïve before/after comparison confuses surge effect with these shocks.

Need a graph-aware causal framework that separates true surge uplift from confounders and quantifies spillovers on neighbours.

Georgia Tech

Related Work

Theme	Key Insights
Production interventions as exogenous shocks	Early works like *Gupta et al. (2019)* used DiD to estimate surge effects. However, DiD assumes surge timing is unrelated to demand shifts a weak assumption in real plants.
Improved causal methods at coarse granularity	*Ahmed et al. (2024)* used doubly robust models but at the plant-month level, masking SKU-level variation. *Lin et al. (2023)* studied overtime spillovers but did not model SKU heterogeneity.
Network interference underexplored in industry	*Kim & Hollingsworth (2022)* and *Vasiliev & Weng (2023)* analyzed COVID-19 and retail shocks in networks but lacked fine-grained SKU-level or surge-relevant modeling.
Prescriptive optimization over causal effects	*Bertsimas et al. (2020)* and *Poupart et al. (2022)* optimize over treatment effects but ignore estimation uncertainty. Our work incorporates bootstrap quantiles in a chance-constrained knapsack.

Georgia Tech

Detecting Surge Episodes

Rule/Heuristic :

- Compute 3-day moving average output: $\mu_i(3)(t)$
- Compute 14-day baseline: $\mu_i(14)(t)$
- Surge starts at day **t if: $\mu_i(3)(t) \geq 1.3 \times \mu_i(14)(t)$** and condition holds for t, t+1, t+2.
- Enforce 10-day cool-down ⇒ 17 632 non-overlapping episodes across 2 104 SKUs, spanning Jan–Aug 2023.

Why this matters:

✓ Provides an objective treatment label for causal analysis.
✓ Matches plant KPI for “tactical surge,” boosting practitioner trust.

Georgia Tech

Key contributions

Contribution	Description
Graph-fused propensity model	Smooths treatment probabilities over the SKU resource graph, boosting overlap and keeping coefficients interpretable.
Dynamic marginal structural models	Recovers day-by-day direct surge effects and 7-day cumulative uplift.
Spillover & heterogeneous uplift estimation	Honest uplift forests quantify how much each neighbour benefits or suffers, with per-SKU confidence intervals.
Chance-constrained knapsack optimizer	Selects SKU surge set that maximizes 5th-percentile fulfilment gain under labour / line-hour budgets; solves in under 90 ms.

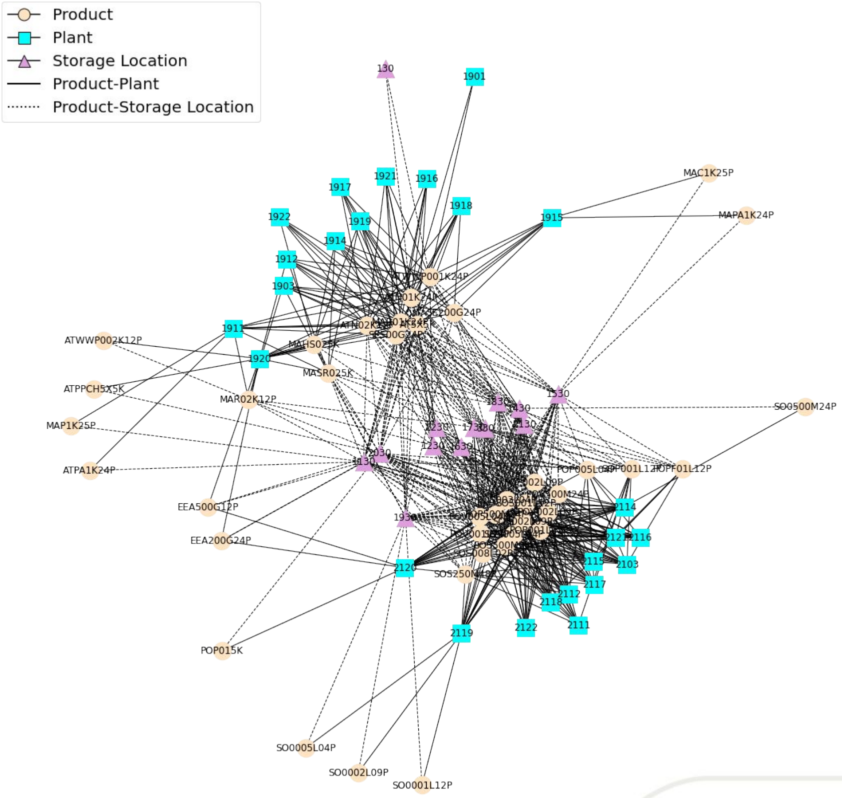
Georgia Tech

Dataset

- Data – SupplyGraph: Open dataset of 2,104 SKUs over 243 days, yielding 17,632 detected surge episodes.

- Each episode has a start day and spans one week of outcome observation. Resource-sharing network: SKUs are nodes connected if they share a production line, primary storage site, or product subgroup.

- The resulting undirected graph has 115,372 edges (average degree ~55), reflecting how capacity and materials link the SKUs. This network encodes potential interference between the SKUs' production.



Causal Assumptions behind the Pipeline

- **Sequential ignorability**

  After conditioning on covariates X_{it}, surge assignment A_{it} and neighbour exposure E_{it} are independent of future outcomes.

- **Partial interference**

  Treatment effects propagate **only** along edges in G; no long-range spillover beyond the immediate resource-sharing graph.

- **Positivity (overlap)**

  Every SKU appears in both surged and non-surged states across 243 days.

- **Control checks**

  ✓ Post-weight absolute SMD < 0.05 for every feature group.

  ✓ Only 0.8 % rows have weight > 10 → no extreme extrapolation.

Graph Based Propensity Estimation

Model and Graph penalty

$$\Pr(A_{it} = 1 \mid X_{it}) = \sigma(\alpha + \gamma_i + X_{it}^\top \beta)$$

$$\min_{\alpha,\beta,\gamma} -\sum_{i,t} \Big[A_{it} \log p_{it} + (1 - A_{it}) \log(1 - p_{it}) \Big] + \lambda \sum_{(i,j)\in E} W_{ij} (\gamma_i - \gamma_j)^2.$$

Why fuse?

- Nearby SKUs share crews & maintenance schedules → similar propensities.
- Fusion shrinks noisy intercepts, enlarges overlap.

Metric	Plain Logit	+Graph Fusion
Avg |SMD|	0.06	**0.04**
ESS (Effective Sample Size)	910	**1,470** (+62%)

Georgia Tech

Dynamic Marginal Structural Model (MSM)

Weighted regression (h = 0…7)

$$Y_{i,t+h} = \alpha_h + \psi_h A_{it} + \eta_h E_{it} + X_{it}^{\top}\beta + \varepsilon_{i,h}, \qquad h = 0, \ldots, 7.$$

Key findings

- Day 0: ψ_0 = −2.6 pp (congestion)
- Day 3: ψ_3 = +1.9 pp
- Day 7: ψ_7 = +5.8 pp (95 % CI +4.6,+6.9)

Interpretation → backlog clears within 5 days; net uplift after 1 week.

Heterogeneous Uplift with Honest Forests

Why heterogeneous?

- 30 % of SKUs show strong overlap → tailor surges where payoff is largest.

Two learners evaluated

Model	RMSE ↓	R^2 ↑	PICP
DR-Learner (GBM + RF)	1.31	0.28	92%
Honest Uplift Forest	**1.18**	**0.34**	**94%**

Forest captures non-linear interactions → better targeting.

Quantifying Network Spillovers

Neighbour model

$$Y_{j,t_k+7} = \alpha + \beta\, A_{i,t_k} + \eta\, E_{i,t_k} + X_{j,t_k}^{\top}\theta + \varepsilon_{jk}$$

Result

• Average spillover β̂ = +0.89 pp (95 % CI +0.48,+1.31)

→ Coordinated surges can lift fulfilment across shared lines.

Operational insight

✓ Capacity synergy outweighs cannibalisation in this plant network.

Georgia Tech

Chance-Constrained Knapsack Optimizer

Objective

$$\max_{x\in\{0,1\}^{|P|}} \sum_{i\in P} \tau_i\, x_i \quad \text{s.t.} \quad \sum_{i\in P} c_i\, x_i \;\le\; K_p$$

- τ_i = 5th-percentile bootstrap CATE   c_i ≈ 0.84 h / SKU
- Why 5th-percentile? → Guarantees uplift in worst 5 % scenarios.

Runtime & optimality

- • Greedy = optimal for K ≤ 5; MILP (CPLEX) < 90 ms for 372 SKUs.
- • Daily end-to-end pipeline = 7.4 s on commodity CPU.

Metric	Heuristic	Ours
Mean uplift	+0.8 pp	**+5.4 pp**
5th-pct uplift	+0.2 pp	**+4.7 pp**
Worst-case risk	—	**−88%**

Georgia Tech

Prescriptive Results vs. Incumbent Heuristic

Offline replay

Policy	Mean ↑	5th-pct ↑	Std-dev ↓
Volume-rank heuristic	+0.8 pp	+0.2 pp	0.6 pp
Point-estimate knapsack	+6.1 pp	+3.3 pp	1.9 pp
Chance-constrained (ours)	**+5.4 pp**	**+4.7 pp**	**1.1 pp**

Key takeaways

- **4.9×** higher worst-case uplift than heuristic.
- Trades 0.7 pp mean to slash downside risk by **88 %**.
- Prevents ≈ 1.9 M late cases over 8 months.

Georgia Tech

Auditability, Robustness, Deployment

Transparency

✓ Propensity weights, MSM coefficients, CATE intervals, solver logs all serialised for domain review.

Robustness checks

- Placebo surge dates → ATT collapses to 0 (no spurious effect).
- Weight-clip 1 ↔ 99 pct → $<$ 0.4 pp drift in ATT.
- Rosenbaum bound: uplift $>$ 0 for $\Gamma \leq 1.35$ (unmeasured bias).

Deployment footprint

- Python + scikit-learn + CPLEX; one cron job per plant.
- CPU-only: 7.4 s total per day, fits inside MES scheduler SLA.

Conclusion

Graph-aware causal ML turns noisy surge logs into reliable, SKU-level uplift and spillover estimates (ESS +62 %, SMD < 0.05).

Risk-aware optimisation lifts forward-week fulfilment by **+5.4 pp** on average and protects **≥ 4.7 pp** in the worst 5 %.
→ **5×** improvement over today's spreadsheet heuristic.

Glass-box & fast: full audit trail, sub-second solve times, deployable with existing MES.

THANK YOU!

Georgia Tech