A Controlled Replication of Stochastic N-k Interdiction on RTS-GMLC: Unconditional Optimality at $k \leq 2$, and Why Ranking Beats Search

Ricardo Marquez, Ph.D. rmarquez123@gmail.com

***Abstract*—We reproduce the stochastic N-k interdiction benchmark of Sundar *et al.* on RTS-GMLC—identical 73-bus network, identical 200 background-outage scenarios, identical DC minimum-load-shed inner problem—with an independent Python/HiGHS implementation, and use the exact replication to answer a question the original could not: are the published optima *globally* optimal? All ten published objectives reproduce to $3.5e-05$ MW (the rounding of the reference files). Enumerating the complete symmetry-quotient candidate space, we prove *unconditionally* that the published $k = 1$ (216 candidates) and $k = 2$ (15,602 candidates) interdictions are global optima—a guarantee neither published method provides, the exact one resting on assumed dual bounds and the heuristic on cuts that are supervalid, not valid. We then test whether a bound-guided best-first search over the interdiction lattice recovers the optima with fewer evaluations than a trivial comparator. It does not: top-k single-component ranking recovers the published objective at *every* k for a cost independent of the budget, while the search exceeds a pre-registered evaluation budget by $k = 3$. The mechanism is measured, not asserted: pairwise interaction among the large units is strongly super-additive (mean +197.0 MW over 78 pairs) but near-uniform, because generator-interdiction value on this transmission-rich, capacity-adequate system is set by the aggregate reserve margin—a scalar—so the problem is, on this instance, a knapsack. We report this negative result against a pre-registered comparator, explain when it would not hold, and delimit what a genuine demonstration of inference-based (rather than search-based) methods would require.**

***Index Terms*—power-grid interdiction, stochastic programming, DC optimal power flow, global optimality, RTS-GMLC, benchmark replication**

I. Introduction

Sundar *et al.* [1] introduce the stochastic N-k interdiction problem for power grids: choose up to k components whose removal maximises the expected minimum load shed, where the expectation is a sample average over a fixed set of background outage scenarios and the operator re-dispatches optimally in response. They give an exact extensive-form MILP (dualising the inner LP, McCormick-linearising the bilinear terms) and a Wood-type cutting-plane heuristic [2], [3], and show the two agree on all k for the RTS-GMLC test case.

We ask three questions the original leaves open. **(RQ1)** Can the exact problem be reproduced by an independent toolchain? **(RQ2)** Are the published optima globally optimal—a claim neither published method establishes, since the exact method assumes dual bounds it does not compute and the heuristic uses cuts it notes are not valid in general (Braess' paradox) [4]? **(RQ3)** Does a bound-guided search over the interdiction lattice recover the optima with materially fewer inner solves than a trivial ranking comparator, and if not, why?

Contributions. (i) An exact, open-stack (scipy/HiGHS, no commercial solver) replication of all ten published objectives to reference precision; (ii) the first *unconditional* proof of global optimality at $k \leq 2$ by complete enumeration of the symmetry-quotient space; (iii) a pre-registered, negative evaluation-reduction result—the search does not beat single-component ranking—with a measured mechanism (aggregate-reserve-set super-additivity); and (iv) a delimitation of when the negative result would not hold and what a non-search (inference-based) method would need to demonstrate. We do not claim a search advantage; on this instance none exists, and we say so with the metric and comparator attached.

II. Exact replication of the inner problem (RQ1)

The inner problem is the DC minimum-load-shed LP: minimise $\sum_i p_i^d(1 - x_i^d)$ over per-bus curtailment $x^d \in [0, 1]$, DC power balance, the flow–angle coupling $x_\ell p_\ell = \theta_f - \theta_t$, thermal limits, and generator limits with the outaged components' bounds forced to zero. We replicate the published implementation's conventions exactly, including three non-standard choices documented in the study's `conventions.md`: the DC susceptance is $1/x$ (ignoring resistance and tap ratio), minimum generation is enforced, and islands are handled natively by the LP (a free coupling row per out-of-service branch) rather than by bus deactivation—the last a trap that otherwise returns

Companion to study directory `stochastic_nk_interdiction`. Every number is read from committed, test-pinned artifacts. Upstream data: `github.com/kaarthiksundar/stochastic_nk` @ `56dc3ec` (MIT), 24 files verified by SHA-256. Wall-clock is one workstation's; evaluation counts are the portable quantity.

TABLE I
LEVEL-1 REPLICATION: PUBLISHED OBJECTIVE VS. THIS STUDY, EVALUATED AT THE PUBLISHED INTERDICTION SETS.

k	published z (MW)	this study (MW)	Δz (MW)
1	151.8247	151.8247	$-3.5e-05$
2	481.3227	481.3227	$-1.5e-05$
3	835.3350	835.3350	$+5.7e-13$
4	1190.3350	1190.3350	$+0.0e+00$
5	1545.3350	1545.3350	$+9.1e-13$
6	1900.3350	1900.3350	$+2.0e-12$
7	2255.3350	2255.3350	$+5.5e-12$
8	2605.3350	2605.3350	$+9.1e-13$
9	2942.5850	2942.5850	$+1.4e-12$
10	3274.5100	3274.5100	$+1.8e-12$

TABLE II
THE $k = 2$ PLATEAU: EVERY CANDIDATE WITHIN 1 MW OF THE GLOBAL OPTIMUM, FROM COMPLETE QUOTIENT ENUMERATION.

rank	candidate	J (MW)	below opt. (MW)
1	{g18@118, g74@121}	481.3227	0.0000
2	{g71@323, g74@121}	480.8473	0.4754
3	{g57@313, g74@121}	480.8392	0.4834
4	{g68@321, g74@121}	480.8140	0.5087
5	{g67@318, g74@121}	480.8104	0.5123

125 MW where the reference is 15 MW on the deterministic $k = 1$ case. All ten *deterministic* reference points reproduce to $1.8e-12$ MW and all ten *stochastic* points to $3.5e-05$ MW—the granularity of the published files, not a modelling difference (Table I). The disagreement, localised across the five candidate sources (input, formulation, scenario, solver, aggregation), is solver rounding alone.

III. UNCONDITIONAL OPTIMALITY AT $k \leq 2$ (RQ2)

Because generators at the same bus with identical limits and identical scenario membership are interchangeable, the 216 interdictable components collapse to 177 symmetry classes; candidates are *multisets* of classes with multiplicity capped by class size (a within-class pair such as the twin units at bus 323 is a legitimate candidate, and the published $k = 5$ optimum contains exactly such a pair). Enumerating this quotient space completely—216 candidates at $k = 1$, 15,602 at $k = 2$—establishes that the published $k = 1$ interdiction $\{g74\}$ and $k = 2$ interdiction $\{g18, g74\}$ are the *global* optima, the latter winning its plateau by 0.475 MW (Table II). This is the paper's central verification result: the extensive-form MILP's optimality certificate is conditional on assumed dual bounds and the heuristic's on supervalid cuts, so neither establishes global optimality on its own; complete enumeration does, and independently confirms the published answers. At $k = 3$ the same machinery finds *eleven* distinct optima at 10^{-4} MW, so exact set recovery is ill-posed beyond $k = 2$ and we report solution equivalence as membership of the reference class in the complete tie set.

IV. SEARCH DOES NOT BEAT RANKING (RQ3)

We test a bound-guided best-first search over the interdiction lattice: internal nodes scored by an LP-free root Wood coefficient, only surviving leaves solved exactly, pruned against a greedy incumbent. The pre-registered success bar is recovery of the published optimum within $B_k = 500{,}000$ unique inner states, and the binding comparator is the cheapest arm that also finds the optimum—top-k single-component ranking. The result is negative (Table III): ranking recovers the published objective at every k for a budget-independent ∼42,000 states, while the search beats it only at $k = 1$ and exceeds B_k by $k = 3$. Budgets $k =$ were not run, recorded as a scope decision rather than a silent cap: the comparator settles the verdict.

TABLE III
MODE B SEARCH VS. ITS BINDING COMPARATOR (SINGLE-COMPONENT RANKING). UNIQUE INNER STATES; RATIO BELOW 1 MEANS THE SEARCH WAS CHEAPER.

k	J (MW)	#opt	Mode B	ranking	ratio	B_k
1	151.8247	1	12,062	42,426	0.28	pass
2	481.3227	1	165,841	42,626	3.89	pass
3	835.3350	11	500,173	42,626	11.73	
textbffail						
4	1190.3350	14	500,118	42,626	11.73	
textbffail						
5	1545.3350	11	500,072	42,626	11.73	
textbffail						
6	1900.3350	5	500,032	42,626	11.73	
textbffail						
7	2255.3350	1	500,164	42,626	11.73	
textbffail						
8	2605.3350	1	500,025	42,626	11.73	
textbffail						
9	2942.5850	1	500,044	42,626	11.73	
textbffail						
10	3274.5100	1	500,023	42,626	11.73	
textbffail						

Why (mechanism). Over the 78 pairs of units ≥ 350 MW, the interaction $I(a, b) = J(\{a, b\}) - J(\{a\}) - J(\{b\}) + J(\emptyset)$ runs +183.2 to +219.6 MW, mean +197.0 MW: strongly super-additive (removing a second large unit costs far more than its solo value, the first removal having spent the 526 MW reserve margin) but *near-uniform*. A near-uniform interaction translates every candidate by roughly the same amount and does not reorder the ranking, so a sorted list of single-component values is already optimal. Generator-interdiction value on this transmission-rich, capacity-adequate system is governed by the aggregate reserve—a scalar—so the problem is a knapsack; only 15–49 of 200 scenarios ever bind the network. A supporting island-capacity lower bound tracks the exact objective to 0.13–4.6% at 7.3× the speed of the warm-started LP (1.78 ms).

V. FAILURE-LATTICE STRUCTURE AS A SEARCH DRIVER

The directive asks whether TERRA's failure-lattice geometry—the structure that organises *contingency* states by probability and severity—can drive the interdiction search. We built the search it specifies and measured it: rank the candidates by a structural lattice score and screen or order the search by it. The answer is negative, and Table IV shows why. A deterministic N-1 severity score needs the top 80 of 216 components, and a top-5%-severity-state

TABLE IV
SCREEN QUALITY: SMALLEST TOP-m OF 216 COMPONENTS WHOSE PREFIX CONTAINS THE PUBLISHED OPTIMUM, BY THREE SCORES. LARGER IS WORSE.

score	top-m to cover A_{10}
N-1 severity (deterministic, low-order lattice)	80
top-5%-severity-state frequency (lattice)	51
stochastic objective $J(\{c\})$ (not structural)	11

frequency score the top 51, to merely *contain* the ten optimal units—against 11 for the stochastic objective. Ranking the interdiction by N-1 severity misses the published objective by up to −2893 MW. The mechanism: failure-lattice severity measures *deterministic single-outage* damage, dominated by line-islanding, whereas interdiction value is a *stochastic capacity* quantity—a large generator removed from a healthy grid merely redispatches (zero N-1 severity) yet is the optimal interdiction under background stress. The two are orthogonal. A lattice screen to the top 51 does retain the optima, so a greedy search within it recovers them (9k–86k states) — a weak candidate-space reduction, but dominated by plain stochastic ranking, which needs no lattice signal at all.

The other lattice signals confirm the same partial picture: lattice-frequency predicts A_{10} membership only weakly (AUC 0.829, below a 0.90 bar, dominated by background-scenario membership); low-order screening recovers the optimum's class in the top 5%; and a locality probe predicts marginal interdiction value from island/corridor structure to a median error ratio of 0.0000. A load-margin "regime" test intended to show the mechanism's collapse as margin widens is confounded by construction—scaling load down manufactures per-island infeasibility—and is reported with that caveat rather than repaired.

Three limits bound the claims. **One instance:** every result is RTS-GMLC at peak load with one scenario set. **One consequence model:** the negative search result follows from generation being fungible on a transmission-adequate system; a model where value is topological (cascades) would not be a knapsack. **Search, not capability:** the tested search is a generic branch-and-bound whose pruning information is the reference method's own Wood cut; it reuses a research engine's architecture but exercises none of its distinctive inference machinery, because the scenario set is a fixed sample by construction. Establishing an inference-based advantage—a certified expectation over an *inferred* distribution, and evidence-conditioned re-optimisation—is a distinct problem (a certified rather than sampled inner expectation, in the sense of stochastic programming's value of the stochastic solution [5]) and out of scope here.

VI. Conclusion

On identical data and inner physics, an independent open-stack implementation reproduces the Sundar *et al.* benchmark exactly and adds what the published methods cannot: an unconditional proof that the $k \leq 2$ optima are global. On the search question the honest answer is negative—no lattice search beats single-component ranking, because the instance is capacity-dominated and its interdiction objective is effectively a knapsack—and we give the measured mechanism that predicts where that ceases to hold. The benchmark is a sound demonstration of a stochastic formulation; it is, for the same reason, unable to discriminate between interdiction *search* methods.

References

[1] K. Sundar, A. Mastin, M. Garcia, R. Bent, and J.-P. Watson, "Exact and heuristic approaches for the stochastic N-k interdiction in power grids," *arXiv:2402.00217*, 2024.

[2] R. K. Wood, "Deterministic network interdiction," *Mathematical and Computer Modelling*, vol. 17, no. 2, pp. 1–18, 1993.

[3] J. Salmeron, K. Wood, and R. Baldick, "Worst-case interdiction analysis of large-scale electric power grids," *IEEE Transactions on Power Systems*, vol. 24, no. 1, pp. 96–104, 2009.

[4] B. Schäfer, D. Witthaut, M. Timme, and V. Latora, "Understanding braess' paradox in power grids," *Nature Communications*, vol. 13, p. 5396, 2022.

[5] J. R. Birge and F. Louveaux, *Introduction to Stochastic Programming*. Springer, 2011.