What Survives

Nine planning questions from one census of a non-convex power-flow model, with gradients from the solve itself

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

September 2026

Abstract

A companion study answered nine planning questions — shadow prices, variance attribution, screening, re-planning under a new belief, tail and worst-case floors — from one census of a load-shedding model at zero or near-zero further solves, with one-sided guarantees on half of the answers. That model was a linear program: convex, with exact gradients from its duals. This study asks what survives when neither property is given. The model is the IEEE 14-bus AC power flow, a non-convex nonlinear system solved by Newton's method; the uncertain inputs are its 8 largest loads; the consequence is the total line overload above a set of ratings, in a congested regime where every line is loaded and an intermittent one where the overload is often zero. Gradients come from the solver itself: at a converged solution one factorization of the Jacobian Newton already formed gives the derivative of every unknown with respect to every input, so a census point costs one solve and one factorization instead of the 17 solves that finite differences of the consequence needed before. Each answer is labeled by what it rests on: a theorem under a declared and tested hypothesis (certified), a one-sided claim that rests on $L \leq G$ and held on every one of 12,000 fresh draws (empirical), or a number with a standard error (estimate). The result: exact gradients survive unconditionally and carry the screening guarantee into the non-convex setting; the floors on the mean, the tail and the worst case survive as observations in the congested regime and fail, by at most 7.2×10^{-4} of $\mathrm{sd}[G]$, in the intermittent one; and every estimate is at or above the accuracy of the sampling alternative at a fraction of its solves. A duct-heater model from a different physical domain, run through the same calls, shows that nothing in the construction is specific to power flow.

1 Introduction

When a model's inputs are uncertain, the planning questions asked of it are functionals of the distribution of a consequence G: its expectation, the sensitivity of that expectation to each input, the share of its variance each input causes, its tail, its worst case, and how all of these change when the belief about the inputs changes. Monte Carlo sampling answers each of these correctly and answers each with a sample designed for it. A companion study[1] showed that when every solve of the model returns not only G but its gradient ∇G, and when G is convex in the inputs, a single census of (input, value, gradient) triples answers all nine at zero or near-zero further solves, and that half of the answers come with a direction: a floor or a ceiling that holds by theorem.

Both conditions were gifts of that model. A linear program's dual variables are exact derivatives of its value, and its value is convex in its right-hand sides. Most physical models are neither. A power flow is a system of nonlinear equations solved by Newton's method; a thermal or hydraulic network is the same; their consequences are as often concave as convex in the inputs, and no dual variable is on offer.

This study takes those conditions away one at a time and asks what survives. The gradient condition is replaced by a general mechanism: at a converged solution of any smooth equation

[1] *One Census, Nine Questions*, the same nine questions on the RTS-GMLC minimum-load-shed linear program.

system, the implicit-function theorem gives the derivative of every unknown with respect to every input from one factorization of the Jacobian the solver already formed. The convexity condition is not replaced; it is dropped, and every one-sided statement is instead checked on fresh draws and labeled by what was observed. The nine questions are then asked of the IEEE 14-bus AC power flow with its eight largest loads uncertain, in two operating regimes, and of a duct heater from the thermal domain.

The software. The model, the sensitivity mechanism and the surrogate machinery are components of TERRA, a modeling engine for networked physical systems; Section 3.1 describes it from the beginning and states what the method requires of any solver.

Outline. Section 2 describes the network, the model, the uncertainty, the two regimes and the questions. Section 3 describes the engine, derives the sensitivity accessor, states the surrogate and what its certificate means on a non-convex model, and specifies both routes for every question. Section 4 reports the answers with their labels, and the duct-heater section. Section 5 states what the comparison establishes. Appendices give the power-flow equations, the accessor's derivation, the estimators, and the software calls.

2 Problem statement

2.1 The test system

The IEEE 14-bus system is a small transmission network in long use as a reference case for power-flow methods (a portion of the American Electric Power system as of 1962, distributed through the University of Washington test-case archive). It has 14 buses, 20 branches (lines and transformers), 5 generator buses and 11 loads totaling 259 MW (Figure 1). The 8 largest loads are the uncertain inputs; the three remaining loads are fixed at their nominal values.

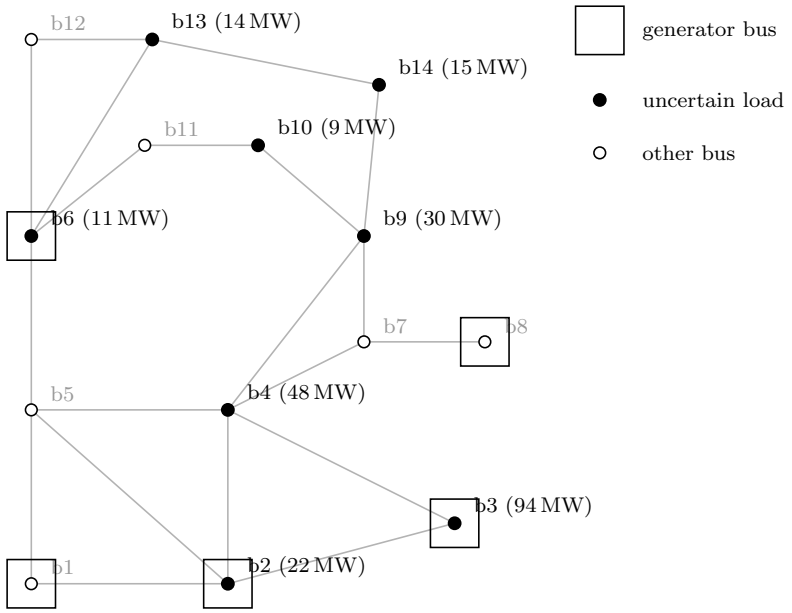


Figure 1: The test system: the IEEE 14-bus network, 14 buses and 20 branches (schematic layout). Squares mark the 5 generator buses (the slack bus b1 and four voltage-controlled buses). The 8 largest loads (filled, with their nominal active power) are the uncertain inputs; the three remaining loads are held at their nominal values. Total load 259 MW.

2.2 The model

The AC power flow (Appendix A) determines the voltage magnitude and angle at every bus, and the active and reactive power on every branch end, from the generation and load at each bus and the branch impedances. It is a system of 114 nonlinear algebraic equations in as many unknowns, solved by Newton's method from a flat start or a warm start; it has no optimization inside it, no dual variables, and its outputs are not convex functions of its inputs. Every solve in this study is one such Newton solution for one vector of loads.

2.3 The uncertainty

Let $d \in \mathbb{R}^8$ be a vector of multipliers, one per uncertain load, scaling both the active and reactive power of that load (constant power factor). Under *belief 1* each d_i is uniform on $[0.85, 1.15]$ independently of the others. Two further beliefs are used in Section 4: *belief 2*, every interval shifted to $[0.8, 1.1]$ (a milder season), and a *shared-weather* belief with belief 1's marginals coupled by a one-factor Gaussian copula of pairwise latent correlation $\rho = 0.6$.

2.4 The consequence and the two regimes

The consequence is the total apparent-power overload above a set of line ratings, in MW on the system base:

$$G(d) \;=\; 100 \sum_{\ell=1}^{20} \max\bigl(0,\; |S_\ell(d)| - r_\ell\bigr),$$

where $|S_\ell|$ is the larger of the two end apparent powers of branch ℓ at the solution for loads d, and r_ℓ its rating. Two rating sets define two regimes. In the *congested* regime every rating is 90% of

the branch's loading at the nominal loads, so every line is overloaded throughout the box and G is the sum of 20 smooth loadings. In the *intermittent* regime every rating is 105% of the nominal loading, so a line is overloaded only when the loads run high; G is zero on 10% of the draws, and its active set changes across the box. The relu makes G non-smooth at the boundary of each line's active set; the loadings themselves are smooth but not convex.

2.5 The questions, and what an answer may rest on

Table 1 lists the nine questions. Each answer in this study carries one of three labels. **Certified**: a theorem under a hypothesis the evaluator declares and the study tests. *Empirical*: a one-sided claim (a floor or a ceiling) that rests on the surrogate inequality $L \le G$, which is a theorem only for a convex G; here it was checked on every holdout and reference draw and observed to hold — an observation, not a guarantee. *Estimate*: a number with a standard error and no direction. The labels are assigned by rule, not by judgment, and Table 8 is the study's central result.

Table 1: The nine questions, the quantity each asks for, and the decision it informs.

	quantity	decision
Q1	$\partial\,\mathbb{E}[G]/\partial d_i$	which load's growth costs the most overload
Q2	$S_{T,i} = \mathbb{E}[\mathrm{Var}(G \mid d_{-i})]/\mathrm{Var}[G]$	which loads drive the uncertainty
Q3	upper bound on $S_{T,i}$	which loads can be ignored, with a guarantee
Q4	$S_{1,i}\,\mathrm{Var}[G] = \mathrm{Var}(\mathbb{E}[G \mid d_i])$	the variance removed by metering load i
Q5	$\mathbb{E}[G]$ under a shifted belief	re-planning when the load forecast changes
Q6	$\mathbb{E}[G]$ under correlated loads	loads that share weather
Q7	$\mathbb{E}[G \mid d_i = d]$ over d	what growth at one bus does
Q8	$P(G > g)$, VaR_α, CVaR_α	compliance: the tail
Q9	$\max_{\mathrm{box}} G$	the worst outcome the loads admit

Every method is charged in *engine solves*: one Newton solution of the power flow for one vector of loads. The sensitivity accessor adds a Jacobian factorization per census point, which is counted separately and costs about a tenth of a solve.

3 Method

3.1 TERRA, from the beginning

TERRA is software for computing the state of a networked physical system from a description of its parts. The description is a graph. Each *node* is a place where a conserved quantity balances: at a bus the active power entering equals the active power leaving, and the same for reactive power; at a duct the enthalpy carried in by the air plus the heat added equals the enthalpy carried out. Each *edge* is a flow between two nodes. Each edge and node carries *closures*, the constitutive relations that determine a flow from the state: for a transmission line the AC equations that give its power flow from the voltages at its two ends; for a duct the product of mass flow and enthalpy. Some quantities are *inputs*: numbers the modeler pins rather than solves for — a load, a generator's voltage set point, an ambient temperature. Given the description, the engine assembles one system of equations, $F(z, u) = 0$: a balance equation per node and domain, a closure equation per relation, with z the vector of unknowns and u the vector of inputs. A *solve* finds $z^*(u)$ by Newton's method, which at each iteration assembles the residual F and its Jacobian $J_z = \partial F/\partial z$ and solves one sparse linear system.

Two things follow from the equations being declared rather than coded. First, the engine knows how F depends on u as well as on z: the input Jacobian $J_u = \partial F/\partial u$ is available for the same price as J_z. Second, a model is one object, so the same solve serves every question asked of it and the engine's plugins — packaged models and analysis methods — combine freely. This study uses three plugins: `power_flow` builds the AC model with its loads declared as inputs; `thermo` builds the duct heater; `uncertainty` provides the census, surrogate, certificate and queries as generic operations on (input, value, gradient) triples and never sees the physics. Table 10 lists the calls.

3.2 The sensitivity accessor

At a converged solution $F(z^*, u) = 0$. If J_z is non-singular there, the implicit-function theorem says that z^* is a smooth function of u near this point and that its derivative satisfies

$$J_z\, S = -J_u, \qquad S = \frac{\partial z^*}{\partial u} \in \mathbb{R}^{n \times K_u}.$$

Newton has already assembled J_z and factorized it; J_u is the block the forward solve did not need. One assembly of J_u and K_u triangular solves with the existing factorization give every derivative of every unknown with respect to every input. Figure 2 shows the chain: the loads d map affinely onto the model's 19 load inputs (constant power factor), S gives the flows' derivatives, and the chain rule through $|S_\ell|$ and the relu, with the active set taken at the solution, gives $\nabla_d G$. The accessor is exact where the relu is differentiable; on the measure-zero boundary of an active set it returns the one-sided derivative of the centre point. Its output was checked against central differences of full re-solves at 4.0×10^{-10} relative on this system, and against closed forms at machine precision on the duct heater.

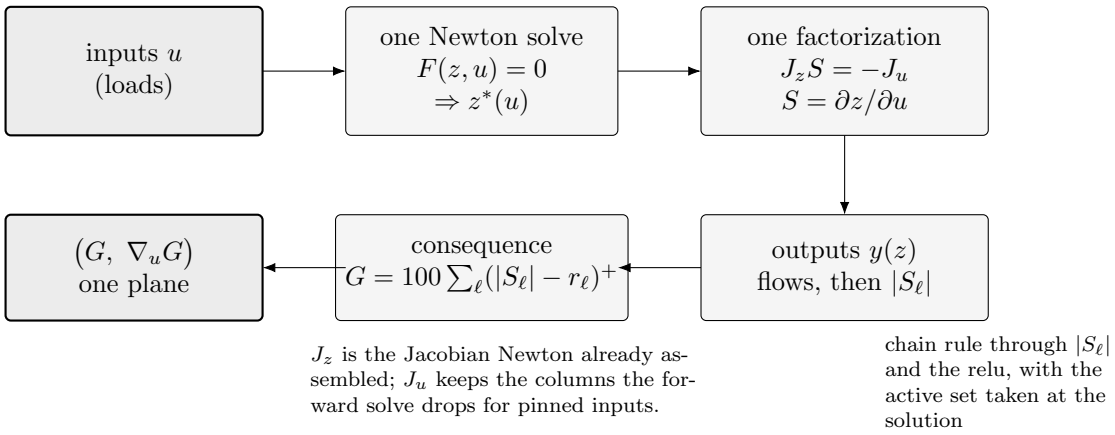


Figure 2: The sensitivity accessor. At a converged solution the implicit-function theorem gives the derivative of every unknown with respect to every input from one factorization of the Jacobian Newton already formed; the consequence's gradient follows by the chain rule. One solve and one factorization per census point, against $2K + 1 = 17$ solves for central differences of the solve.

For the census this means one solve and one factorization per point. The superseded route — central differences of the consequence itself — cost $2K + 1 = 17$ solves per point and, across a kink of the relu, returned an average of two slopes that is a gradient of neither side.

3.3 The surrogate, and what its certificate means here

Let $\{(d_j, G_j, c_j)\}_{j=1}^{C}$ be the census. The surrogate is

$$L(d) \;=\; \max\Big(0,\; \max_{j}\big[G_j + c_j \cdot (d - d_j)\big]\Big),$$

the maximum of the tangent planes and of the floor 0. For a convex G every plane lies below G and $L \le G$ everywhere; that inequality is what turned the companion study's floors into theorems. For the model here G is not convex, a plane can lie above G (Figure 3), and $L \le G$ is a hypothesis to be checked, not a fact. The check is the *holdout*: G is solved on 2,000 fresh draws and the excess $L - G$ recorded; the reference sample of 10,000 draws repeats the check at larger scale. The evaluator declares `convex = False`, so the software never labels a floor certified on this model whatever the check finds; where the check finds nothing, the label is *empirical*.

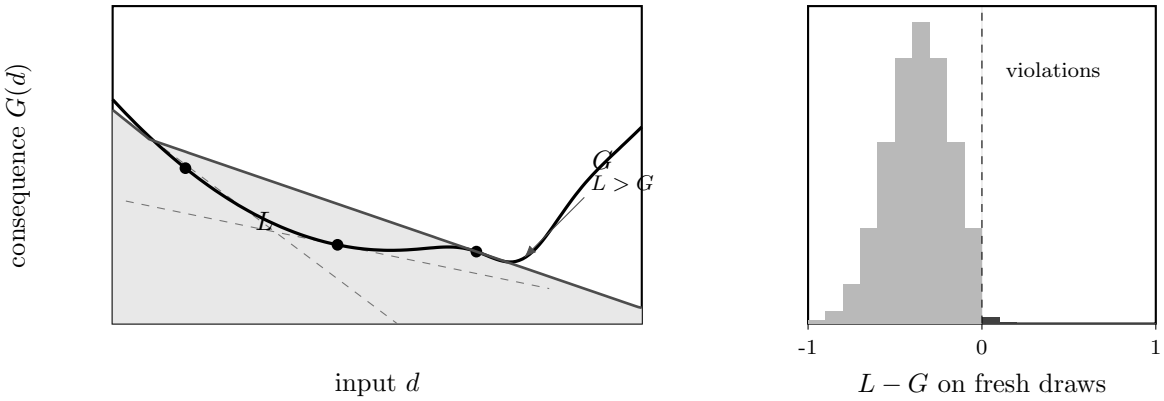


Figure 3: Left: the cutting-plane surrogate on a non-convex consequence (one input shown). Each census point contributes the tangent plane $G_j + c_j \cdot (d - d_j)$; L (grey) is the maximum of the planes and the floor 0. Where G is convex the planes lie below it; where it bends the other way a plane can rise above G (arrow), and $L \le G$ fails there. Right: the certificate on a non-convex model is therefore an observation — the distribution of $L - G$ on fresh draws, with its mass at or below zero and any violations counted — not a theorem.

One hypothesis does hold, and it is the one the screening bound needs. The derivative-based bound of Sobol' and Kucherenko, $S_{T,i} \le w_i^2\, \mathbb{E}[(\partial G/\partial d_i)^2]/(\pi^2 \operatorname{Var}[G])$ for d_i uniform on an interval of width w_i, requires that G be absolutely continuous along each input axis with the harvested slopes its almost-everywhere gradients; it does not require convexity. The solution map is smooth on the box and the relu is Lipschitz, so the accessor's gradients satisfy it; the evaluator declares `exact_gradients = True`, and the bound is a theorem in both regimes. Its two ingredients $\mathbb{E}[(\partial G/\partial d_i)^2]$ and $\operatorname{Var}[G]$ remain Monte Carlo estimates.

3.4 Monte Carlo and its designs

The sampling alternatives are the standard ones, used as black boxes that return G only: plain and quasi-Monte Carlo for expectations (scrambled Sobol' sequences with independent scramblings as the interval); central differences of the mean with common random numbers for Q1 ($2KN$ solves); Saltelli's pick-freeze design (Figure 4) with Jansen's estimators for Q2 and Q4 ($(K+2)N$ solves); Morris's elementary-effect screening for Q3 ($R(K+1)$ solves, a ranking with no bound); conditional

Monte Carlo per grid point for Q7; the empirical tail for Q8; and, for Q9, the value of G at every vertex of the box, which for a convex G is the exact maximum and for this G is an estimate like the surrogate's. The estimators are in Appendix B.

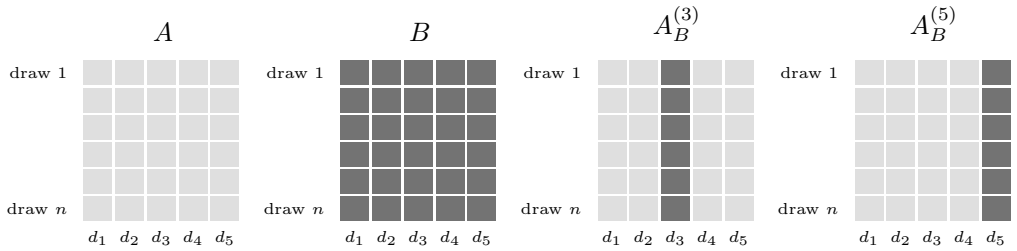


Figure 4: The pick-freeze design behind Q2 and Q4 (five inputs shown). Two independent sample matrices A and B are drawn; the hybrid $A_B^{(i)}$ takes column i from B and every other column from A. Rows k of A and of $A_B^{(i)}$ agree on all inputs except d_i, and the average squared difference of f across such pairs measures how much d_i alone moves the output — the total index. Every matrix is a set of solves on the model, or of cheap evaluations on the surrogate.

3.5 The nine queries, both routes

Figure 5 shows the TERRA route. For each question the TERRA route is a function of the census matrix and the surrogate; the sampling route is the design above at the budget stated, per regime.

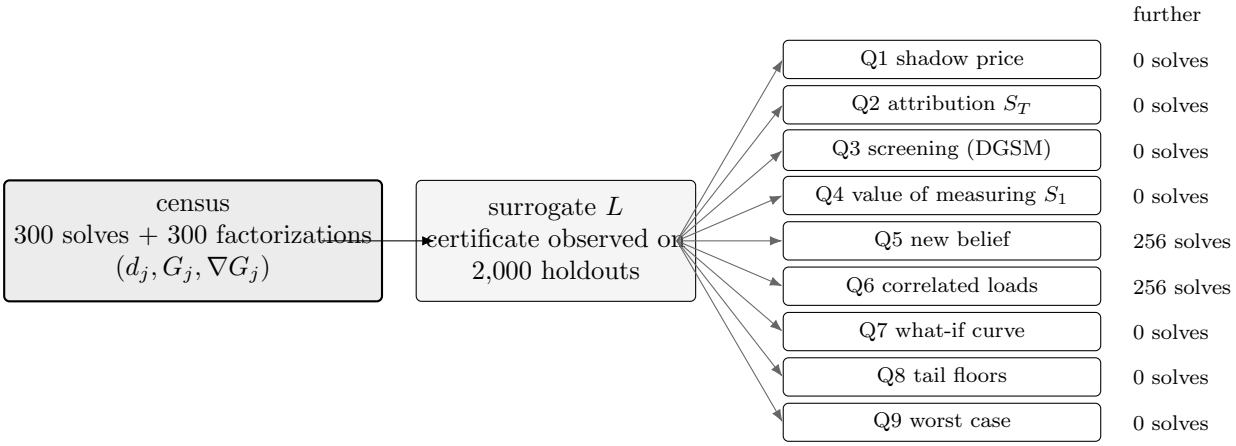


Figure 5: One census, nine questions, per regime. The census and its certificate are paid once; each question is then a function of the surrogate and of cheap draws from the input distribution. Only the two questions that change the input distribution (Q5, Q6) solve again, for a small residual sample.

Q1, shadow price of load. $\partial\,\mathbb{E}[G]/\partial d_i = \mathbb{E}[\partial G/\partial d_i]$: the column mean of the census gradients with its standard error, zero solves; against central differences of the mean, $N = 500$ common

draws per shift, 8,000 solves. The reference is the mean gradient over the 10,000 reference draws, which return gradients too.

Q2, Q4, attribution and value of metering. Jansen's total and first-order estimators on the surrogate with $n = 20,000$ rows per matrix (zero solves, bootstrap standard errors); against the same design on the model at $N = 1,000$, 10,000 solves; reference $N = 3,000$.

Q3, screening. The DGSM bound from the census gradients, zero solves, a load excluded below 1%; against Morris with $R = 20$ trajectories, 180 solves.

Q5, Q6, a new belief. $\mathbb{E}[L]$ under the new belief over 2^{21} scrambled-Sobol' draws (zero solves) plus a residual $\mathbb{E}[G - L]$ from 256 solves at fresh draws in 8 scrambles; against a fresh quasi-Monte Carlo study of 2,048 solves in four scrambles; reference 8,192. $\mathbb{E}[L]$ alone is a floor on $\mathbb{E}[G]$ only where $L \leq G$.

Q7, what-if curve. $\mathbb{E}[L \mid d_i = d]$ on a grid of 7 values for the most influential load, the others drawn once ($n = 20,000$), zero solves; against conditional Monte Carlo at $N = 300$ per grid point; reference $N = 1,500$.

Q8, tail floors. $P(L > g)$ at the reference quantiles, the α-quantile of L and its conditional tail mean on the reference draws, zero solves; against the empirical tail of 5,000 solves.

Q9, worst case. $\max_{\text{box}} L$ in closed form — each plane is separable over the box, so its maximum is at the vertex that takes each coordinate to the end its slope favors, and $\max L$ is the largest over the planes, at $O(CK)$ arithmetic — followed by one solve of G at that vertex; against G at all $2^8 = 256$ vertices.

4 Results

4.1 Cost of the census, and the certificate

In both regimes the census of 300 points cost exactly 300 solves and 300 factorizations, one of each per point; the superseded route cost 17 solves per point. No solve needed a fallback attempt. Figure 6 shows the residual $G - L$ and the violations of $L \leq G$ as the census grows.

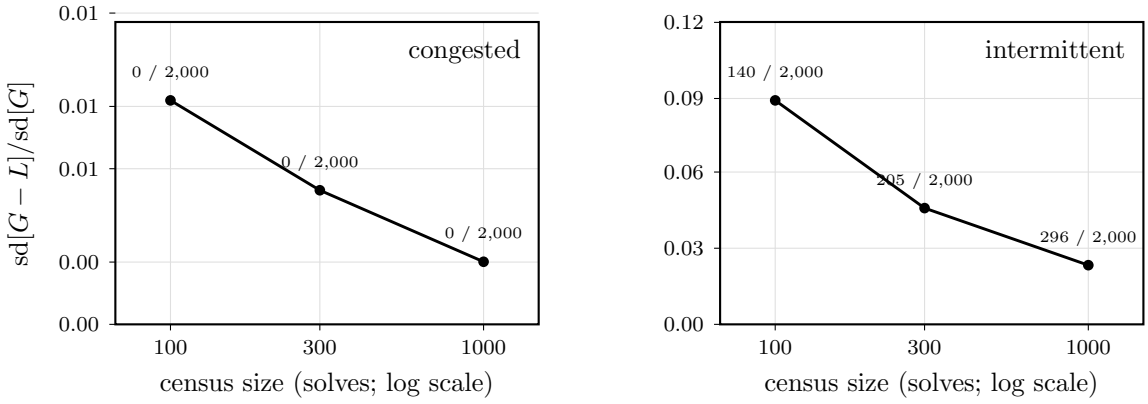


Figure 6: The residual $G - L$ on the 2,000-draw holdout as the census grows, in both regimes; the number above each point is the count of violations of $L \leq G$ among the holdout draws. Congested: the surrogate captures all but 0.5% of the spread of G at census 300 with no violation; intermittent: 4.3% with violations on 10% of the draws, each below 7.2×10^{-4} of sd$[G]$.

Congested. On the 2,000 holdout draws there were 0 violations beyond the solver tolerance of 1.0×10^{-4} MW, and 0 on the 10,000 reference draws. The residual $G - L$ has standard deviation 0.128 MW against 27.68 MW for G (ratio 0.0046). The surrogate is below G everywhere it was looked at, and close.

Intermittent. There were 205 violations on the holdout (10% of the draws) and 1149 on the reference, the largest excess 4.6×10^{-3} MW, which is 7.2×10^{-4} of sd$[G] = 6.45$ MW. The residual ratio is 0.0434. The surrogate is an accurate interpolant of G but not a bound on it: the one-sided claims of Q5–Q9 are estimates in this regime, and Table 8 says so.

4.2 Cost of the questions

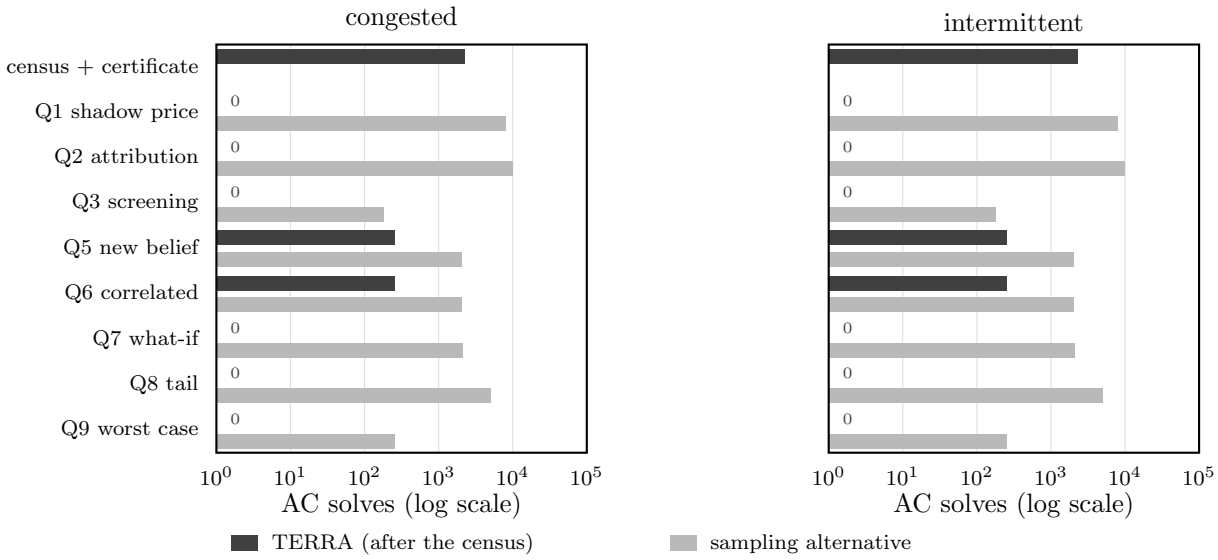


Figure 7: AC solves per question and regime: the surrogate route (after the census) against the standard sampling method at the budgets of Section 3.5. Q4 is answered from the same run as Q2 on both sides. Totals over both regimes: 9,738 (TERRA, census included) and 71,264 (alternatives).

Table 2: Solve ledger per regime. References are excluded. The census also spends one Jacobian factorization per point (not a solve).

	TERRA solves			alternative solves	
question	cong.	interm.	alternative	cong.	interm.
census (values and gradients)	300	300	—	—	—
holdout certificate	2,000	2,000	—	—	—
Q1 shadow price of load	0	0	finite differences of the mean	8,000	8,000
Q2 variance attribution	0	0	Saltelli on the AC solve	10,000	10,000
Q3 screening	0	0	Morris elementary effects	180	180
Q4 value of metering a load	0	0	the same Saltelli run	0	0
Q5 a new belief	256	256	fresh QMC study	2,048	2,048
Q6 correlated loads	256	256	fresh QMC study	2,048	2,048
Q7 what-if curve	0	0	conditional Monte Carlo	2,100	2,100
Q8 tail floors	0	0	empirical tail	5,000	5,000
Q9 worst case over the box	0	0	G at every vertex	256	256
total	3,068	3,068		29,632	29,632

4.3 Expected overload under belief 1

The control variate $\mathbb{E}[G] = \mathbb{E}[L] + \mathbb{E}[G - L]$ with a 256-solve residual gives 72.40 ± 0.01 MW (congested) and 4.45 ± 0.04 MW (intermittent) against the references 72.41 ± 0.28 and 4.52 ± 0.06; the residual variance ratios are 58697 and 613. Everything below is answered from the same census.

4.4 Q1: shadow price of load

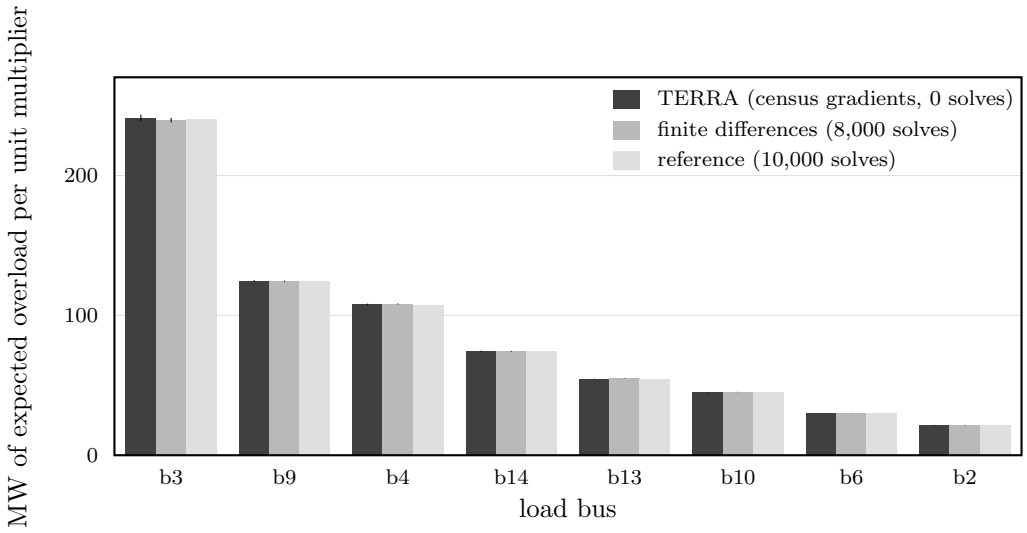


Figure 8: Q1 (congested regime): shadow price of load, $\partial\,\mathbb{E}[G]/\partial d_i$, for the 8 uncertain loads. Error bars are one standard error.

In the congested regime the largest error of the zero-solve answer over the 8 loads is 0.94 MW per unit multiplier on a scale of 240; finite differences at 8,000 solves reach 0.71. In the intermittent regime the figures are 1.64 against 1.78 on a scale of 62. Both routes and the reference rank the same three buses first in the congested regime (b3, b9, b4); the intermittent reference's top three are b3, b9, b4 against the census's b3, b9, b4. A shadow price of s means that growing that load by one per cent adds $s/100$ MW of expected overload; it is the number that ranks demand-side programs by bus. Both answers are estimates: the census mean is unbiased for the expectation but carries the standard error of 300 draws.

4.5 Q2–Q4: attribution, screening, value of metering

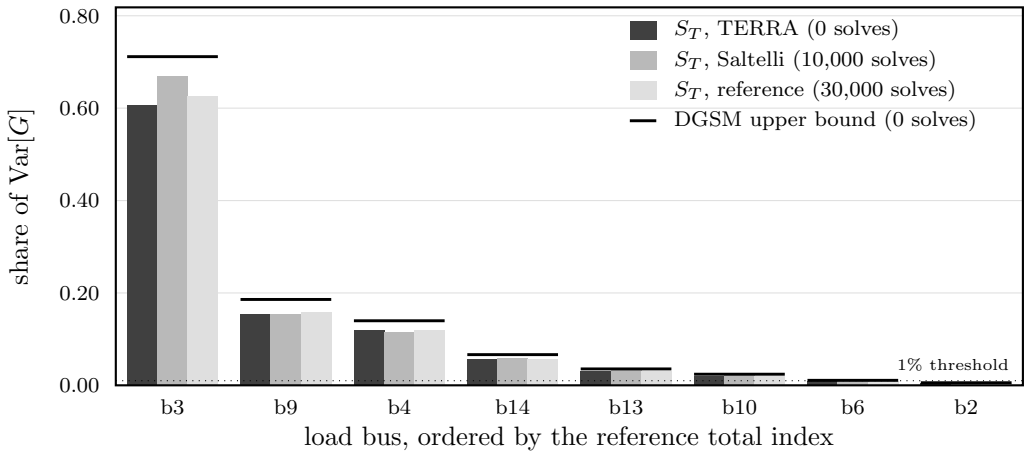


Figure 9: Q2 and Q3 (congested regime): total Sobol' index of every uncertain load by the three routes, with the DGSM upper bound from the census gradients as a cap on each group. The cap is a theorem here (the gradients are exact); a load whose cap lies below the 1% line is excluded with a guarantee.

Table 3: Q2–Q4, congested regime, ordered by the reference total index. S_T: TERRA from the surrogate (zero solves), Saltelli on the AC solve (10,000 solves), the reference at 30,000 solves. S_1: TERRA and reference. DGSM: the certified upper bound on S_T from the census gradients; μ^*: the Morris mean elementary effect (180 solves).

	S_T			S_1		DGSM	Morris
load bus	TERRA	Saltelli	ref.	TERRA	ref.	bound	μ^*
b3	0.606	0.670	0.627	0.616	0.603	0.711	240.5
b9	0.154	0.153	0.158	0.162	0.149	0.186	121.0
b4	0.118	0.115	0.119	0.129	0.110	0.140	102.3
b14	0.056	0.059	0.057	0.072	0.049	0.066	72.8
b13	0.030	0.033	0.030	0.041	0.030	0.035	54.2
b10	0.021	0.022	0.020	0.034	0.022	0.024	42.4
b6	0.009	0.009	0.009	0.020	0.014	0.011	29.2
b2	0.005	0.005	0.005	0.019	0.003	0.006	20.9

Table 4: Q2–Q4, intermittent regime, ordered by the reference total index. S_T: TERRA from the surrogate (zero solves), Saltelli on the AC solve (10,000 solves), the reference at 30,000 solves. S_1: TERRA and reference. DGSM: the certified upper bound on S_T from the census gradients; μ^*: the Morris mean elementary effect (180 solves).

	S_T			S_1		DGSM	Morris
load bus	TERRA	Saltelli	ref.	TERRA	ref.	bound	μ^*
b3	0.805	0.893	0.832	0.638	0.628	2.656	76.9
b4	0.180	0.189	0.183	0.028	-0.001	0.322	16.2
b9	0.119	0.094	0.134	0.029	0.014	0.286	14.5
b14	0.064	0.066	0.067	0.037	0.018	0.105	18.3
b10	0.026	0.030	0.028	0.009	0.004	0.035	19.1
b13	0.026	0.027	0.027	0.006	-0.009	0.047	12.0
b2	0.017	0.016	0.015	0.000	-0.004	0.022	4.7
b6	0.009	0.008	0.009	-0.003	-0.018	0.012	6.3

Total indices from the surrogate are within 0.021 (congested) and 0.027 (intermittent) of the reference on every load; Saltelli on the model at 10,000 solves is within 0.043 and 0.060. The three most influential loads are b3, b9, b4 (congested) and b3, b4, b9 (intermittent) by the reference; the surrogate gives b3, b9, b4 and b3, b4, b9. The DGSM bound lies above the reference total index on 8/8 loads in the congested regime and 8/8 in the intermittent one, as the theorem requires; it excludes 1 loads below 1% in the congested regime (b2) and 0 in the intermittent one. This is the one guarantee of the companion study that survives intact: it never depended on convexity, only on the gradients being exact, and the accessor's are. Morris, at 180 solves, ranks the bottom three as b2, b6, b10 against the reference's b2, b6, b10 (congested), with no statement of how much any of them could matter.

4.6 Q5–Q6: a new belief, correlated loads

Table 5: Q5–Q6: the census re-marginalized under a different input distribution, both regimes. TERRA: $\mathbb{E}[L]$ under the new distribution plus a 256-solve residual sample, with the 95% half-width, and $\mathbb{E}[L]$ alone (a floor only where $L \le G$); alternative: a fresh scrambled-Sobol' study of 2,048 solves (one standard error over four scrambles); reference: 8,192 solves. All in MW.

	congested				intermittent			
belief	TERRA	$\mathbb{E}[L]$	QMC	ref.	TERRA	$\mathbb{E}[L]$	QMC	ref.
belief 2, $U(0.80, 1.10)$	39.90 ± 0.04	39.45	39.91 ± 0.00	39.90	1.10 ± 0.01	1.09	1.11 ± 0.00	1.11
shared weather, $\rho = 0.6$	73.59 ± 0.03	73.40	73.57 ± 0.02	73.60	9.34 ± 0.04	9.21	9.39 ± 0.02	9.36

Under belief 2 the expected overload falls to 39.90 MW (congested) and 1.11 MW (intermittent); the surrogate harvested under belief 1 recovers these within 0.00 and 0.01 MW from 256 residual solves, against 0.01 and 0.00 MW for a fresh quasi-Monte Carlo study of 2,048 solves. Under the copula the figures are 0.01 / 0.02 MW (TERRA) and 0.03 / 0.03 MW (fresh study). The floor $\mathbb{E}[L] \le \mathbb{E}[G]$ was observed to hold under both beliefs in the congested regime (yes, yes) and also in the intermittent one; in neither is it a theorem.

4.7 Q7: what-if curve

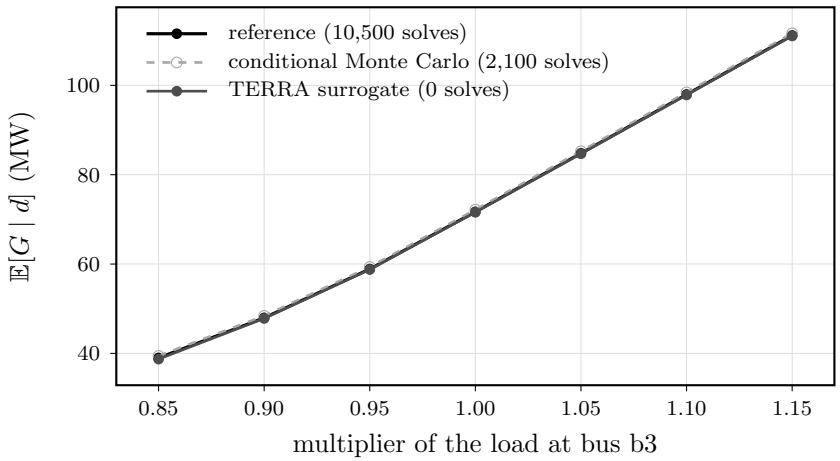


Figure 10: Q7 (congested regime): expected overload as a function of the most influential load's multiplier, the other 7 averaged over belief 1.

The surrogate curve is within 0.28 MW (congested) and 0.38 MW (intermittent) of the reference across the range of the load at bus b3; conditional Monte Carlo at 2,100 solves deviates by up to 0.48 and 0.14 MW. The curve lay below the reference at every grid point in the congested regime (yes) and also in the intermittent one.

4.8 Q8–Q9: tail and worst-case floors

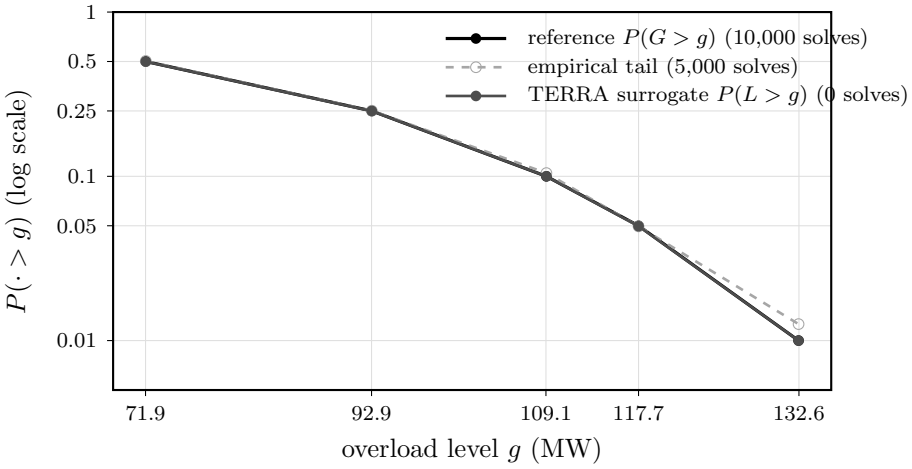


Figure 11: Q8 (congested regime): exceedance probability of the overload at the reference quantiles.

Table 6: Q8, congested regime: exceedance probabilities at the reference quantiles. TERRA: $P(L > g)$ at zero solves; alternative: the empirical tail of 5,000 solves with its binomial standard error; reference: 10,000 solves. VaR_{95} / CVaR_{95}: TERRA 117.67 / 127.19, empirical 117.65 / 127.54, reference 117.70 / 127.22 MW.

quantile	g (MW)	TERRA	empirical	reference
0.50	71.92	0.4996	0.5036 ± 0.0071	0.5000
0.75	92.92	0.2498	0.2508 ± 0.0061	0.2500
0.90	109.14	0.1000	0.1046 ± 0.0043	0.1000
0.95	117.70	0.0497	0.0496 ± 0.0031	0.0500
0.99	132.57	0.0100	0.0126 ± 0.0016	0.0100

Table 7: Q8, intermittent regime: exceedance probabilities at the reference quantiles. TERRA: $P(L > g)$ at zero solves; alternative: the empirical tail of 5,000 solves with its binomial standard error; reference: 10,000 solves. VaR_{95} / CVaR_{95}: TERRA 18.41 / 25.06, empirical 18.67 / 25.76, reference 18.67 / 25.34 MW.

quantile	g (MW)	TERRA	empirical	reference
0.50	1.78	0.4852	0.5044 ± 0.0071	0.5000
0.75	5.77	0.2414	0.2552 ± 0.0062	0.2500
0.90	13.09	0.0956	0.1010 ± 0.0043	0.1000
0.95	18.67	0.0486	0.0500 ± 0.0031	0.0500
0.99	29.19	0.0100	0.0120 ± 0.0015	0.0100

In the congested regime the surrogate's tail statements held at every level and for both risk measures (yes); in the intermittent regime they held as well. Over the whole box the surrogate's maximum is 180.6 MW (congested) at a vertex where one solve finds $G = 180.8$ MW; the largest G over all 256 vertices is 180.8 MW (the same vertex). Intermittent: 74.1 against $G = 74.3$ at the vertex and 74.3 over the vertices (same vertex). For this G neither route is exact: the maximum of a non-convex function over a box need not sit at a vertex, so both are estimates, and the largest reference draws (157.2 and 50.6 MW) are the only unconditional lower bounds on hand.

4.9 What survives

Table 8: What survives on the non-convex model. For each question: what its answer rests on, and in each regime the label — **certified** (a theorem under a declared and tested hypothesis), *empirical* (a one-sided claim that rests on $L \le G$ and was observed to hold on every holdout and reference draw; no theorem), or estimate (a number with a standard error, no direction) — with the evidence from Section 4.

	rests on	congested		intermittent	
Q1	the gradients being unbiased at each draw	estimate	max err 0.94 vs FD 0.71 (scale 240)	estimate	max err 1.64 vs FD 1.78 (scale 62)
Q2	the surrogate's variance decomposition	estimate	max $\lvert\Delta S_T\rvert$ 0.021 vs Saltelli 0.043	estimate	max $\lvert\Delta S_T\rvert$ 0.027 vs Saltelli 0.060
Q3	exact gradients (the DGSM hypothesis)	**certified**	bound $\ge$ reference on 8/8; 1 excluded	**certified**	bound $\ge$ reference on 8/8; 0 excluded
Q4	the surrogate's variance decomposition	estimate	max $\lvert\Delta S_1\rvert$ 0.023 vs Saltelli 0.033	estimate	max $\lvert\Delta S_1\rvert$ 0.028 vs Saltelli 0.065
Q5	$L \le G$ under the new belief	*empirical*	$\mathbb{E}[L]$ = 39.45 vs ref 39.90; err 0.00 vs QMC 0.01	estimate	$\mathbb{E}[L]$ = 1.09 vs ref 1.11; err 0.01 vs QMC 0.00
Q6	$L \le G$ under the copula	*empirical*	$\mathbb{E}[L]$ = 73.40 vs ref 73.60; err 0.01 vs QMC 0.03	estimate	$\mathbb{E}[L]$ = 9.21 vs ref 9.36; err 0.02 vs QMC 0.03
Q7	$L \le G$ along the grid	*empirical*	within 0.28 MW of the reference; below everywhere: yes	estimate	within 0.38 MW of the reference; below everywhere: yes
Q8	$L \le G$ on the draws	*empirical*	floors hold at every level: yes; CVaR_{95} 127.2 vs 127.2	estimate	floors hold at every level: yes; CVaR_{95} 25.1 vs 25.3
Q9	$L \le G$ at the vertex	*empirical*	max L 180.6, G there 180.8, vertex max 180.8	estimate	max L 74.1, G there 74.3, vertex max 74.3

Two questions and one guarantee survive the loss of convexity without qualification: the gradients are exact, so the screening bound of Q3 holds by theorem in both regimes, and the estimates of Q1, Q2 and Q4 are as accurate as before. The one-sided claims of Q5–Q9 rest on $L \le G$; in the congested regime they were observed to hold on every one of 12,000 fresh draws and are labeled empirical; in the intermittent regime the inequality fails on 11% of the draws, by tiny amounts, and the same numbers are labeled estimates. What the data are consistent with is this: with every line active the consequence is the sum of the 20 line loadings, and a violation can only come from concavity of an *active* line's loading (a line that de-activates makes its plane fall below G, never above); the observed residual ratios (0.005 congested, 0.043 intermittent) and the absence of violations in the congested regime are consistent with the aggregate curvature being small and positive while individual loadings are not convex. The study does not establish this; it reports it.

4.10 A different domain: the duct heater

To show that nothing above is specific to power flow, Q1–Q3 were asked of a duct heater from the `thermo` plugin: air at mass flow $\dot{m}$ and inlet temperature T_{in} receives an electric input W and loses $Q = 0.5\,\mathrm{kW}$, so that $T_{\mathrm{out}} = T_{\mathrm{in}} + (W - Q)/(\dot{m}\, c_p)$. The three inputs are uniform on $\dot{m} \in [2.4, 3.6]\,\mathrm{kg/s}$, $T_{\mathrm{in}} \in [285.15, 295.15]\,\mathrm{K}$, $W \in [10, 20]\,\mathrm{kW}$; the consequence is T_{out}, whose Hessian in $(\dot{m}, W)$ is indefinite — not convex. The model is three nodes, two enthalpy-advection edges, a source and a loss; the engine solves it in one Newton step and the accessor returns the gradients from the same factorization, exact to machine precision against the closed form. The same census, holdout and queries were run with no change to any call.

Table 9: The duct heater, Q1–Q3. Q1: $\mathbb{E}[\partial T_{\text{out}}/\partial x_i]$ from the census (200 solves, zero further), its closed form over the box, and finite differences of the mean (12,000 solves). Q2: S_T from the surrogate, Saltelli on the model (5,000 solves), and the reference (40,000). Q3: the DGSM bound, a theorem here.

	Q1 (K per unit)			Q2, S_T			Q3
input	TERRA	closed form	FD	TERRA	Saltelli	ref.	bound
$\dot{m}$ (kg/s)	-1.65058	-1.66989	-1.67865	0.032	0.036	0.035	0.045
T_{in} (K)	1.00000	1.00000	1.00000	0.861	0.881	0.872	1.050
W (kW)	0.33201	0.33621	0.33685	0.110	0.098	0.096	0.117

The census gradients recover the closed-form shadow prices within 1.2% (Monte Carlo error of the 200-point mean; finite differences at 12,000 solves: 0.5%); the surrogate's total indices are within 0.013 of the reference (Saltelli on the model at 5,000 solves: 0.009); the DGSM bound is a theorem and holds (yes). The inequality $L \leq G$, by contrast, fails on 100% of the reference draws: a tangent plane of T_{out} in the $(\dot{m}, W)$ plane lies above the surface on one side and below it on the other, so the surrogate is an interpolant here and any floor built on it would be an estimate. The labels came out as they must: Q3 certified, Q1–Q2 estimates, nothing empirical.

5 Discussion

What the comparison establishes. Three things. First, the gradient condition of the companion study is not a property of linear programs but of solvers: at a converged solution of any smooth equation system the implicit-function theorem delivers every sensitivity from one factorization, so a census on a Newton-solved model costs what it cost on the LP. Second, the guarantee that rests on exact gradients alone — the derivative-based screening bound — survives the loss of convexity intact, in both regimes and in a second physical domain. Third, the guarantees that rest on $L \leq G$ do not survive as theorems, and the software's labels say so; what remains of them is an accurate estimate with a standard error, at or above the sampling alternative's accuracy at a fraction of its solves, plus an observation — no violation on 12,000 draws in the congested regime — that a planner may weigh but cannot sign.

What it rests on. The accessor needs a converged solution with a non-singular Jacobian and a consequence that is differentiable where it is evaluated; the active set of a relu is taken at the centre and the result is exact away from the boundary. The screening bound needs absolute continuity along each axis with the slopes as almost-everywhere gradients, which the smooth solution map and the Lipschitz relu provide; a discontinuous consequence, or a model whose operating regime switches inside the box, would forfeit it, and the evaluator's declaration is where that must be withheld. Every empirical label is a statement about the draws that were checked, at the tolerance stated.

Where sampling keeps its place. On the expectation alone a quasi-Monte Carlo study is as accurate as the surrogate route at equal solves, here as in the companion study; sampling needs no gradients and makes no claim it cannot keep. The surrogate's advantage on a non-convex model is narrower than on a convex one: the cost of every question after the first, and the one guarantee that convexity was never needed for.

6 Conclusions

- On the IEEE 14-bus AC power flow with 8 uncertain loads, a census of 300 points costs 300 solves and 300 Jacobian factorizations — one of each per point, from the engine's sensitivity accessor — against 17 solves per point by finite differences.
- Exact gradients survive the loss of convexity: the derivative-based screening bound holds by theorem in both regimes (8/8 and 8/8 above the reference), and the shadow prices and variance attribution are estimated as accurately as on the convex model.
- The one-sided claims that rest on $L \le G$ are not theorems on this model. In the congested regime they held on every one of 12,000 fresh draws and are reported as empirical; in the intermittent regime the inequality fails on 11% of the draws by at most 7.2×10^{-4} of sd$[G]$, and the same numbers are estimates.
- The construction is domain-neutral: the duct heater from the thermal plugin, through the same calls, gives machine-precision gradients, a certified screening bound, and a surrogate that is an interpolant rather than a bound — with the labels assigned by the same rule.

A The AC power flow and the accessor

Equations. With V_b, θ_b the voltage magnitude and angle at bus b, and $y_\ell = g_\ell + \mathrm{j}b_\ell$ the series admittance of branch ℓ from b to b' (with a shunt and a tap for transformers), the branch flows at the b end are

$$P_{\ell,b} = V_b^2 g_\ell - V_b V_{b'}\big(g_\ell \cos\theta_{bb'} + b_\ell \sin\theta_{bb'}\big),$$
$$Q_{\ell,b} = -V_b^2(b_\ell + b_\ell^{\mathrm{sh}}) - V_b V_{b'}\big(g_\ell \sin\theta_{bb'} - b_\ell \cos\theta_{bb'}\big),$$

$\theta_{bb'} = \theta_b - \theta_{b'}$, and symmetrically at the b' end. At every bus the active and reactive power balance: generation minus load equals the sum of the branch flows leaving the bus. The slack bus fixes $\theta = 0$ and V and supplies whatever active and reactive power balance requires; a voltage-controlled bus fixes V and its active generation and supplies reactive power; a load bus fixes both injections. In the engine each bus is a node balanced in the active and reactive domains, each branch two edges into a midpoint node with the flow closures above, and each load a pinned input u entering its bus's balance with a minus sign. The system has 114 unknowns and as many equations.

Accessor. At the converged z^*, $F(z^*, u) = 0$ with F smooth in (z, u) and J_z non-singular, so $\partial z^*/\partial u = -J_z^{-1} J_u$. The engine assembles J_u by keeping, in each residual row, the partial derivatives with respect to pinned inputs that the forward assembly discards: ∓ 1 for a pinned load in a balance row, the closures' partials with respect to a pinned argument, -1 on a row that pins a variable to a value. One sparse LU of J_z then serves all K_u right-hand sides. The consequence's gradient is $\nabla_u G = J_u$-weighted through the chain rule: with y the flows, $\partial y/\partial u$ the rows of S, $|S_\ell| = \sqrt{P^2 + Q^2}$ at the loaded end, and $G = 100 \sum_\ell (|S_\ell| - r_\ell)^+$, $\partial G/\partial u = 100 \sum_{\ell\ \mathrm{active}} (P_\ell\, \partial P_\ell/\partial u + Q_\ell\, \partial Q_\ell/\partial u)/|S_\ell|$; the multipliers enter through $u = Ad + b$, so $\nabla_d G = A^\top \nabla_u G$.

B Estimators

Write A and B for two independent $n \times K$ matrices of draws from the belief and $A_B^{(i)}$ for A with column i replaced by B's. For a function f (either G on the model or L on the surrogate) with sample variance V:

$$S_{T,i} \approx \frac{1}{2nV} \sum_{k=1}^{n} \big[f(A_k) - f(A_{B,k}^{(i)})\big]^2, \qquad S_{1,i} \approx 1 - \frac{1}{2nV} \sum_{k=1}^{n} \big[f(B_k) - f(A_{B,k}^{(i)})\big]^2,$$

Jansen's estimators. The derivative bound for $d_i \sim U(a_i, b_i)$ is $S_{T,i} \le (b_i - a_i)^2\, \mathbb{E}[(\partial f/\partial d_i)^2]/(\pi^2 V)$ (Sobol' and Kucherenko, 2009; Lamboni et al., 2013), valid when f is absolutely continuous along each axis with $\partial f/\partial d_i$ its almost-everywhere derivative in L^2 — the Poincaré inequality for the uniform law, whose constant is $1/\pi^2$; convexity plays no part. Morris's screening uses R trajectories that change one input at a time by a fixed step and reports μ_i^*, the mean absolute change of f per unit of input i. The control variate for a new belief is $\mathbb{E}[G] = \mathbb{E}[L] + \mathbb{E}[G - L]$, the first term over scrambled-Sobol' draws of L, the second over a residual sample of solves in R scrambles; the 95% half-width combines the two replicate spreads in quadrature. VaR_α is the α-quantile and CVaR_α the mean above it; both are monotone in the pointwise order, so their values on L bound their values on G from below exactly when $L \le G$. The worst case of L over the box is $\max_j \big[G_j - c_j \cdot d_j + \sum_k \max(c_{jk} a_k, c_{jk} b_k)\big]$, attained at the vertex that takes each coordinate to the end its slope favors; for the duct heater the closed-form shadow prices over the box are $\mathbb{E}[\partial T_{\text{out}}/\partial \dot{m}] = -(\mathbb{E}[W] - Q)/(c_p\, a\, b)$ for $\dot{m} \sim U(a, b)$, $\partial T_{\text{out}}/\partial T_{\text{in}} = 1$, and $\mathbb{E}[\partial T_{\text{out}}/\partial W] = \ln(b/a)/((b - a)\, c_p)$.

C The software calls

Table 10: The software call behind each step (`plugins.uncertainty` unless noted).

step	call	from
model	`make_ac_pf_model(spec, load_variables=True)`, `load_variable_names`	`plugins.power_flow`
gradients	`engine_gradient(model, inputs, outputs, input_map=, outer=)`	on `tge.sensitivity`
census	`harvest`, `holdout_check`	
$\mathbb{E}[G]$	`control_variate`	
Q1	`mean_gradient`	
Q2	`sobol_total`	
Q3	`dgsm_bound` (label from `Pieces.exact_gradients`)	
Q4	`sobol_first`	
Q5	`control_variate` with a second `uniform_box`	
Q6	`control_variate` with `gaussian_copula`	
Q7	`conditional_mean`	
Q8	`exceedance_floor` (label from `Pieces.convex_declared`)	
Q9	`worst_case_floor` (closed-form corner maximum)	
thermo	`make_duct_heater`	`plugins.thermo`

References

R. D. Christie, "Power Systems Test Case Archive: 14 Bus Power Flow Test Case," University of Washington, 1993 (data of the IEEE 14-bus system, 1962).

M. J. W. Jansen, "Analysis of variance designs for model output," *Computer Physics Communications*, vol. 117, pp. 35–43, 1999.

A. Saltelli et al., "Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index," *Computer Physics Communications*, vol. 181, pp. 259–270, 2010.

I. M. Sobol' and S. Kucherenko, "Derivative based global sensitivity measures and their link with global sensitivity indices," *Mathematics and Computers in Simulation*, vol. 79, pp. 3009–3017, 2009.

M. Lamboni, B. Iooss, A.-L. Popelin and F. Gamboa, "Derivative-based global sensitivity measures: general links with Sobol' indices and numerical tests," *Mathematics and Computers in Simulation*, vol. 87, pp. 45–54, 2013.

M. D. Morris, "Factorial sampling plans for preliminary computational experiments," *Technometrics*, vol. 33, no. 2, pp. 161–174, 1991.

R. T. Rockafellar and S. Uryasev, "Optimization of conditional value-at-risk," *Journal of Risk*, vol. 2, no. 3, pp. 21–41, 2000.

J. E. Kelley, "The cutting-plane method for solving convex programs," *Journal of the SIAM*, vol. 8, no. 4, pp. 703–712, 1960.

W. F. Tinney and C. E. Hart, "Power flow solution by Newton's method," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-86, no. 11, pp. 1449–1460, 1967.

Reproduction. With `PYTHONPATH=src`, run `run.py` then `make_report.py` with `--pdf`. Seed 20260909, 200,570 AC solves and 22,000 factorizations including references, 79,200 duct-heater solves, 0 fallback attempts. Every number above is read from `results.json`; the schematic figures carry no data.