One Census, Nine Questions

Planning queries from a certified surrogate of a load-shedding model
under uncertain transmission line ratings

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

September 2026

Abstract

Transmission line ratings — the maximum power each conductor may carry — are uncertain, and the load a network must shed when they turn out low is the quantity planners act on. Estimating its expectation is the cheap part of the analysis; the decisions rest on what follows: which line's added capacity relieves the most shed, which ratings drive the uncertainty, which can be ignored, what changes when the weather forecast changes, how heavy the tail is. With Monte Carlo sampling each of these needs a sample designed for it. This study shows that when the network solve returns its sensitivities as well as its value, all of them are answered from one census. On the RTS-GMLC test system with 14 uncertain ratings, a census of 3,000 solves is harvested into a cutting-plane surrogate $L \leq G$ of the minimum load shed G, certified on 2,000 holdout draws. Nine questions are answered from the surrogate at zero further solves (seven of them) or from a small residual sample (two), each against the standard sampling method for the same question and a large reference. The surrogate route spends 6,536 solves in total; the alternatives spend 113,876. Accuracy is at or above the alternatives' in every case, and where the surrogate makes a one-sided statement — a floor on the tail, on the worst case and on the mean under a new belief, a ceiling on every attribution — the statement holds on the reference. The capability rests on two conditions, stated and tested: the solve must return its sensitivities, and the consequence must be convex in the uncertain inputs.

1 Introduction

A transmission line's rating is the largest power flow its conductor may carry without overheating. Ratings depend on weather — ambient temperature, wind, solar heating — and are therefore uncertain at planning time. When ratings fall below what the dispatch needs, load must be shed. For a planner the shed is a random variable G, and the questions of interest are functionals of its distribution: its mean, its tail, its sensitivity to each line, its behavior under a different belief about the weather.

The standard approach is Monte Carlo: draw ratings, solve the network, average. It is correct and simple, and its cost does not depend on the number of uncertain inputs. Its limitation is that a sample answers the question it was drawn for. The expectation of G is one sample; the share of its variance due to line i requires a second sample constructed for that purpose; a different belief requires a third; a conditional curve a fourth. Each new question is a new study.

This note examines an alternative that becomes available when the solve returns not only G but ∇G, its gradient with respect to the uncertain inputs. For the load-shedding model used here the gradient is free: the model is a linear program and its dual variables are the sensitivities of its value to every rating bound. A set of (input, value, gradient) triples — a *census* — defines a piecewise-linear surrogate that, because the model is convex, lies below G everywhere. Nine planning questions are then posed to the surrogate. The study measures what each answer costs, how accurate it is against a large reference, and how that compares with the sampling method a practitioner would otherwise use.

The software. The network model and the surrogate machinery are components of TERRA, a modeling engine for networked physical systems whose solves return sensitivities as well as values; Section 3.1 describes it and the two plugins used. Nothing in the method is specific to that software; what it requires of any solver is stated there.

Outline. Section 2 describes the test system, the model, the uncertainty, and the nine questions. Section 3 gives the surrogate, its certificate, the estimator behind each question, the sampling alternatives, and how both are scored. Section 4 reports the answers. Section 5 states what the comparison establishes and the conditions it rests on. Appendices give the model's formulation, the estimators, and the software calls.

2 Problem statement

2.1 The test system

The RTS-GMLC is the 2019 update of the IEEE Reliability Test System, a synthetic transmission network maintained for reproducible reliability studies (Barrows et al., 2020). The instance used here has 73 buses, 120 branches (lines and transformers), 96 generators with 9,076 MW of capacity, and 51 loads totaling 8,550 MW (Figure 1). Every branch rating is derated to 50% of its nominal value so that the network is congested: at the derated ratings the dispatch must already shed 154.1 MW.

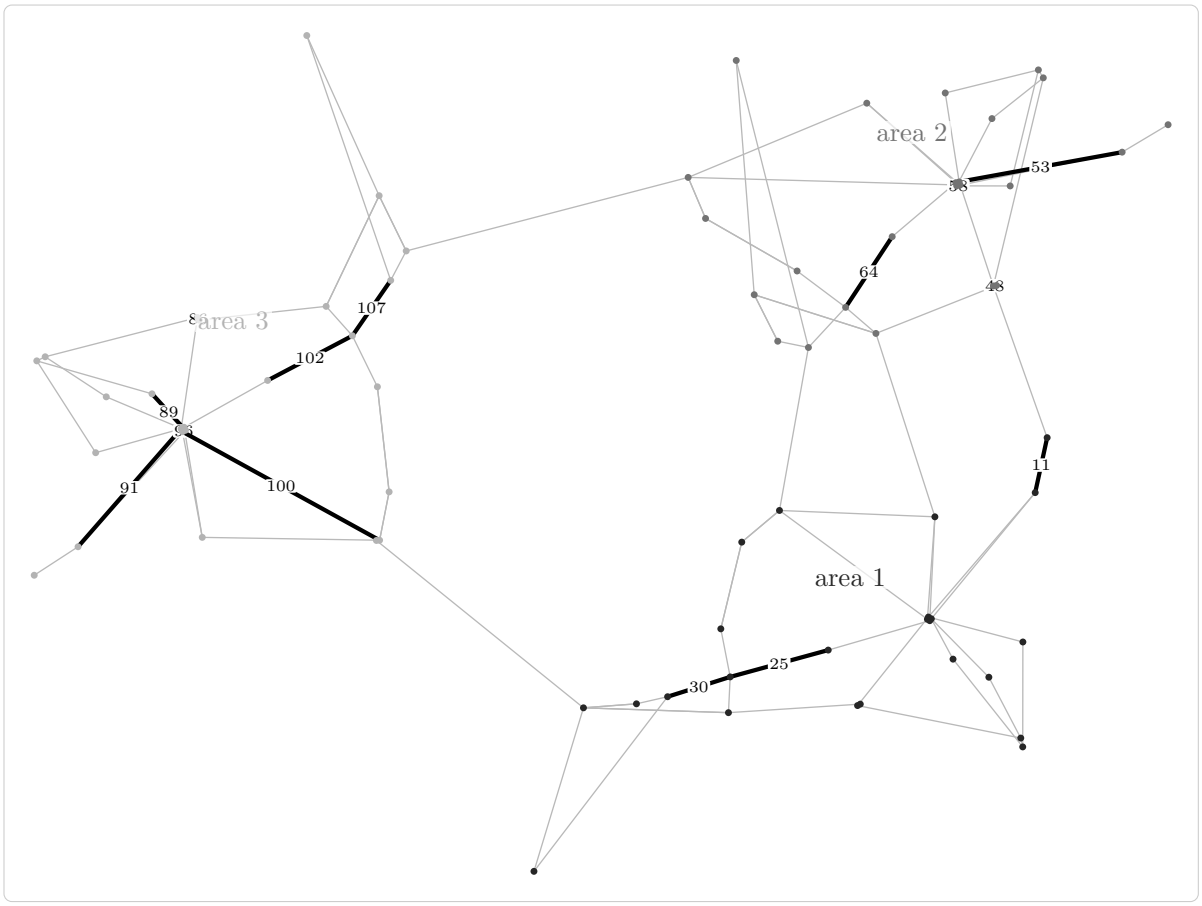


Figure 1: The test system: the 73 buses (dots, shaded by area) and 120 branches of the RTS-GMLC in their geographic layout (bus coordinates from the RTS-GMLC data set). The 14 branches whose ratings are uncertain are drawn heavy and labeled with their branch number; they are the lines whose rating bound is active at the derated operating point, and they lie in every area.

2.2 The load-shedding model

The consequence G is the minimum load shed, in MW, that a re-dispatch of the whole system can achieve under the DC power-flow approximation. It is the value of a linear program (Appendix A): generation is chosen within its limits, each load may be curtailed, power balances at every bus, flows follow the DC angle equations, and every branch flow is bounded by its rating. Two properties of this model are used throughout. First, its value is a *convex* function of the rating bounds — the value of a linear program is convex in its right-hand sides. Second, the dual variables on the flow bounds are, by linear-programming sensitivity, the derivatives of the value with respect to the ratings; the solver returns them with every solution at no extra cost.

2.3 The uncertainty

A single solve at the derated point identifies the branches whose rating bound is active: $K = 14$ of the 120 (branches 11, 25, 30, 48, 53, 58, 64, 86, 89, 91, 96, 100, 102, 107). These are the uncertain inputs. Under *belief 1* each is uniform on $[0.85, 1.15]$ times its derated rating, independently of the others. Two further beliefs are used in Section 4: *belief 2*, every box shifted to $[0.8, 1.1]$ (a hotter

season), and a *shared-weather* belief with belief 1's marginals coupled by a one-factor Gaussian copula of pairwise latent correlation $\rho = 0.6$ (Figure 2).

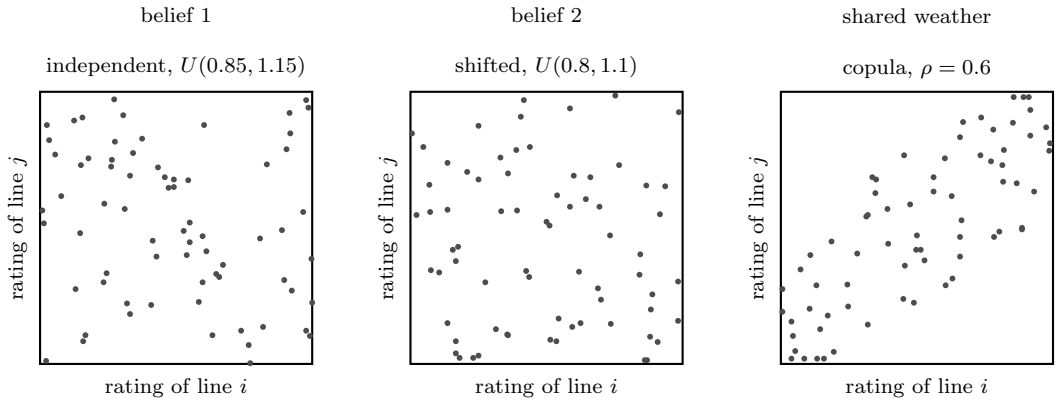


Figure 2: The three input distributions (schematic, two of the K ratings shown). The surrogate is harvested once under belief 1 and re-used under the other two: its planes are valid everywhere in the input space, so only the weighting changes.

2.4 The questions

Table 1 lists the nine questions with the quantity each asks for and the decision it informs. Q1 prices capacity; Q2–Q4 rank the lines by influence, exclude some with a guarantee, and value a measurement; Q5–Q6 re-plan under a different belief; Q7 is the response curve of one line; Q8–Q9 bound the tail and the worst case.

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

	quantity	decision
Q1	$\partial\,\mathbb{E}[G]/\partial r_i$	which line to reconductor; the value of one MW of rating
Q2	$S_{T,i} = \mathbb{E}[\mathrm{Var}(G \mid r_{-i})]/\mathrm{Var}[G]$	which ratings drive the uncertainty
Q3	certified upper bound on $S_{T,i}$	which lines can be ignored, with a guarantee
Q4	$S_{1,i}\,\mathrm{Var}[G] = \mathrm{Var}(\mathbb{E}[G \mid r_i])$	the variance removed by measuring line i
Q5	$\mathbb{E}[G]$ under a shifted belief	re-planning when the forecast changes
Q6	$\mathbb{E}[G]$ under correlated ratings	ratings that share weather
Q7	$\mathbb{E}[G \mid r_i = r]$ over r	what an uprate of line i buys
Q8	$P(G > g)$, VaR_α, CVaR_α	compliance: the tail, bounded
Q9	$\max_{\mathrm{box}} G$	the worst outcome the ratings admit

Every method is charged in *engine solves*: one solution of the load-shedding program for one vector of ratings. Wall-clock time is not used, because it depends on the solver and the machine while the solve count does not.

3 Method

3.1 TERRA, and what one solve returns

Readers need no prior knowledge of TERRA; this subsection states what it is and the two things this study takes from it.

TERRA is software for computing the state of a networked physical system — a power grid, a district heating loop, a data-center cooling plant — from a description of its parts. The description is a graph. Each *node* is a place where a conserved quantity balances (the power entering a bus equals the power leaving it; the mass entering a pipe junction equals the mass leaving it). Each *edge* is a flow between two nodes. Each edge and node carries *closures*: the constitutive relations that say how a flow depends on the state, such as the DC law that a line's flow is proportional to the angle difference across it, or a generator's output limits. Given the description, the engine assembles one system of equations — a balance equation per node, a closure equation per relation — and solves it. A *solve* is one such solution for one set of input values; its result is the value of every flow and every state variable in the network. For the model of this study the equations include an optimization (the dispatch chooses generation and curtailment to minimize shed), and the engine delegates that piece to a linear-programming solver.

Two features of the engine are used here and are worth stating plainly. First, *a solve returns sensitivities*. Because the closures are declared rather than buried in code, the engine knows how every output depends on every input, and at the solution it can report the derivative of any output with respect to any input — the Jacobian of the equations for a smooth model, and for a model that contains an optimization, the optimizer's *dual variables*. A dual variable is the rate at which the optimum changes when one constraint is relaxed by one unit; the dual of a line's flow limit is therefore exactly $\partial G/\partial r$, the change in minimum shed per unit of that line's rating. The optimizer computes the duals as a by-product of the optimum, so they come with the solve at no extra cost. Second, *the model is one object*. The same solve serves every question asked of the system; no question needs its own model, and the engine's plugins — packaged domain models and analysis methods that operate on any model — can be combined freely.

This study uses two of TERRA's plugins. The `power_flow` plugin provides the DC minimum-load-shed dispatch of Appendix A as a linear program (solved with HiGHS), together with an *evaluator* that presents it to the rest of the system as a function of the uncertain ratings: given a vector $r \in \mathbb{R}^K$ it writes the ratings into the model, solves once, and returns the pair

$$\big(G(r),\ \nabla G(r)\big) \in \mathbb{R} \times \mathbb{R}^K,$$

the shed in MW and its derivative with respect to each uncertain rating in MW per unit of rating, read from the duals of the flow bounds. The evaluator also declares two facts the downstream machinery relies on: that G is convex in r, and the numerical tolerance of its values. The `uncertainty` plugin takes any such evaluator — it never sees the network — and provides the census, the surrogate, the certificate and the nine queries as generic operations on $(r, G, \nabla G)$ triples. Table 8 lists the calls.

Concretely, for this system one solve returns $1 + 14$ numbers. The census of Section 3.3 is therefore a matrix of 3,000 rows and $2K + 1 = 29$ columns: the rating vector drawn, the shed, and the 14 duals. Everything the TERRA route does below is arithmetic on that matrix and on cheap draws from a belief; the model is not solved again except for the two residual samples of Q5 and Q6.

3.2 Monte Carlo and its designs

The alternative to which every answer is compared is sampling on the same model, with the evaluator used as a black box that returns G only. Its elements are standard and are stated here so that each alternative below is fully specified.

Plain Monte Carlo. Draw N rating vectors from the belief, solve each, average: $\bar{G} = \frac{1}{N}\sum G(r_k)$. The standard error is $\sigma_G/\sqrt{N}$ regardless of K. *Quasi-Monte Carlo* replaces the random draws by a scrambled Sobol' sequence, a point set that fills the input box more evenly; its error decays faster than $N^{-1/2}$ on smooth integrands, and independent scramblings give a replicate spread that serves as its interval. It is the strong baseline for any expectation.

Finite differences of the mean (Q1). To estimate $\partial\,\mathbb{E}[G]/\partial r_i$ without gradients, compute the mean at r_i shifted up and down by h and difference: $[\bar{G}(r + he_i) - \bar{G}(r - he_i)]/2h$. Using the same N draws for both shifts (common random numbers) cancels most of the sampling noise; the cost is $2N$ solves per input, $2KN$ in all.

Pick-freeze designs (Q2, Q4; Figure 5). A variance-based index asks how much of $\mathrm{Var}[G]$ is due to input i. Saltelli's design draws two independent $N \times K$ matrices A and B and forms, for each i, the hybrid $A_B^{(i)}$ that takes column i from B and the rest from A. Row k of A and of $A_B^{(i)}$ differ only in r_i, so the average of $[G(A_k) - G(A_{B,k}^{(i)})]^2$ measures the variance r_i moves on its own and with its interactions (the total index); the average of $[G(B_k) - G(A_{B,k}^{(i)})]^2$ measures what it moves by itself (the first-order index). All K hybrids plus A and B cost $(K + 2)N$ solves.

Elementary-effect screening (Q3). Morris's method walks R random trajectories through the input box, changing one input at a time by a step Δ and recording the change in G per unit of input; the mean absolute change μ_i^* over the trajectories ranks the inputs at $R(K + 1)$ solves. It ranks; it does not bound.

Conditional Monte Carlo (Q7). For each value r on a grid, draw the other $K - 1$ ratings N times, fix $r_i = r$, solve and average: N solves per grid point.

Empirical tails and vertex enumeration (Q8, Q9). The tail of G is estimated by the fraction of a fresh sample above each level, with its binomial standard error; the worst case over the box, for a convex G, is found exactly by solving all 2^K vertices.

3.3 The surrogate and its certificate

Let $\{(r_j, G_j, c_j)\}_{j=1}^{C}$ be the census, $C = 3,000$ draws from belief 1 with their values and gradients. The surrogate is

$$L(r) \;=\; \max\Big(0,\; \max_j \big[G_j + c_j \cdot (r - r_j)\big]\Big).$$

Each term is the tangent plane of G at r_j; since G is convex every plane lies below G, and since $G \geq 0$ the floor does too, so $L \leq G$ at every r — not only on the census, and not only under belief 1 (Figure 3). Evaluating L on a batch of points is one matrix product with the 3,000 slopes followed by a row maximum; it costs no solve.

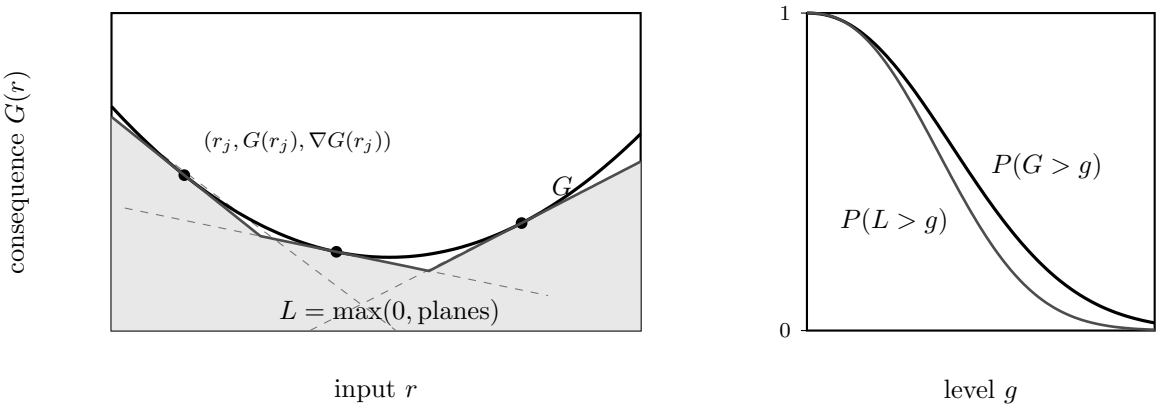


Figure 3: Left: the cutting-plane surrogate in one dimension. Each solve at an anchor r_j returns the value $G(r_j)$ and the gradient $\nabla G(r_j)$, which define a supporting plane; L (grey region) is the maximum of the planes and of the floor 0. Because G is convex, every plane lies below G and so does L. Right: the consequence for any tail statement — since $L \leq G$ pointwise, $P(L > g) \leq P(G > g)$ at every level, and the same holds for quantiles, conditional tail means and the maximum over any set.

The inequality is a theorem given convexity; the *certificate* is its empirical check on fresh draws, which guards against an incorrectly declared convexity or a wrong sign in the gradient. On 2,000 holdout draws the number of violations of $L \leq G$ beyond the solver's tolerance was 0 (largest excess 4.8e-10 MW), and on the 100,000 reference draws of Section 4 it was 0. The residual $G - L$ has mean 0.61 MW and standard deviation 1.18 MW, against a standard deviation of G of 50.3 MW: the surrogate captures all but a small, non-negative remainder.

3.4 One census, nine queries: both routes, question by question

Figure 4 shows the TERRA route. For each question the paragraphs below state what is asked, exactly how the TERRA route answers it from the census matrix and the surrogate, and exactly how the sampling route answers it; solve counts are the budgets used in Section 4. The estimators are collected in Appendix B.

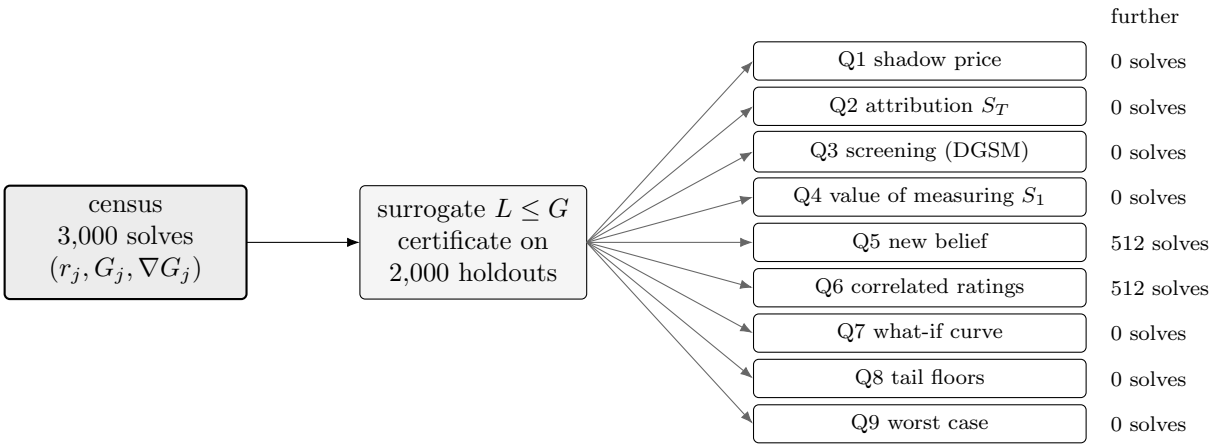
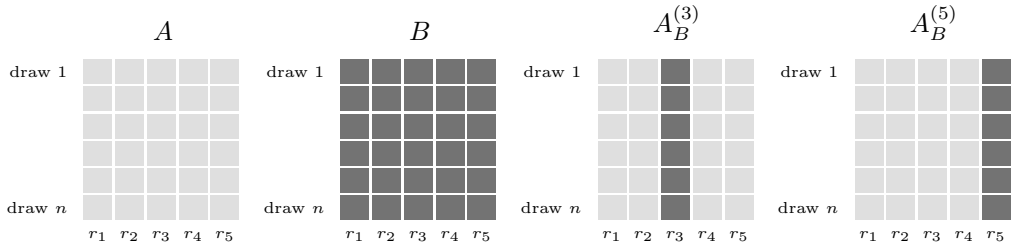


Figure 4: One census, nine questions. The census and its certificate are paid once; each question is then a function of the surrogate and of cheap draws from the input distribution. Only the two questions that change the input distribution (Q5, Q6) solve again, for a small residual sample.



each row is one rating vector to solve; light: from A, dark: from B. Pairing row k of A with row k of $A_B^{(i)}$ holds every rating but r_i fixed.

Figure 5: 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 $A_B^{(i)}$ therefore agree on all inputs except r_i, and the average squared difference of f across such pairs measures how much r_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.

Q1, shadow price of capacity. *Asked:* $\partial\,\mathbb{E}[G]/\partial r_i$ for each line — how much expected shed one more MW of rating on line i removes. *TERRA:* because differentiation and expectation commute, this is $\mathbb{E}[\partial G/\partial r_i]$, and the census already holds 3,000 draws of $\partial G/\partial r_i$ from the belief: the answer is the column mean of the census gradient matrix, with the column's standard error; divided by the system base it is in MW per MW. Zero solves. *Sampling:* central differences of the mean with common random numbers, $N = 2,000$ draws per shift, $2KN = 56,000$ solves.

Q2, variance attribution. *Asked:* the total Sobol' index $S_{T,i}$ of each rating. *TERRA:* the pick-freeze design of Figure 5 applied to L instead of G: A, B and the K hybrids are drawn from belief 1 with $n = 50,000$ rows each and evaluated on the surrogate — $(K+1)n$ matrix products, zero solves — and Jansen's estimator is applied. A bootstrap over the rows gives the standard error. *Sampling:* the same design on the model, $N = 1,000$, $(K+2)N = 16,000$ solves.

Q3, screening. *Asked:* which lines cannot matter, with a guarantee. *TERRA:* the derivative-based bound of Sobol' and Kucherenko, $S_{T,i} \le w_i^2\,\mathbb{E}[(\partial G/\partial r_i)^2]/(\pi^2\,\mathrm{Var}[G])$ for r_i uniform on an interval of width w_i; the second moment is the column mean of the squared census gradients, so the bound is arithmetic on the same matrix as Q1. A line whose bound is below 1% is excluded. Zero solves. *Sampling:* Morris screening with $R = 20$ trajectories, $R(K+1) = 300$ solves, which yields μ_i^* and a ranking.

Q4, value of measuring a line. *Asked:* the variance of G that would disappear if rating i were known exactly, $\mathrm{Var}(\mathbb{E}[G \mid r_i]) = S_{1,i}\,\mathrm{Var}[G]$. *TERRA:* Jansen's first-order estimator on the surrogate with the same matrices as Q2; zero solves. *Sampling:* the same Saltelli run as Q2 gives the first-order index at no further solves, which is why the alternative's cost for Q4 is zero in the ledger.

Q5, Q6, a new belief. *Asked:* $\mathbb{E}[G]$ under a belief other than the one the census was drawn from — belief 2 (every box shifted) and the shared-weather copula. *TERRA:* the surrogate is distribution-free, so $\mathbb{E}[L]$ under the new belief is the average of L over 2^{21} scrambled-Sobol' draws from it (no solves; $R = 16$ scrambles give its spread), and it is a certified lower bound on $\mathbb{E}[G]$ on its own because $G - L \ge 0$. The residual $\mathbb{E}[G - L]$ is then estimated from 512 solves at fresh draws from the new belief, arranged in 8 scrambled replicates, and the two terms are added; the 95% interval combines the two replicate spreads. *Sampling:* a fresh quasi-Monte Carlo study under the new belief, $N = 4,096$ solves in four scrambles. The reference is the same at $N = 20,000$.

Q7, what-if curve. *Asked:* $\mathbb{E}[G \mid r_i = r]$ across the rating range of the most influential line. *TERRA:* draw the other $K - 1$ ratings once from belief 1 ($n = 50,000$ rows), overwrite column i with each grid value r in turn, evaluate L and average: one matrix product per grid point, zero solves, a lower bound at every point. *Sampling:* conditional Monte Carlo, $N = 1,000$ solves at each of the 7 grid points, 7,000 in all.

Q8, tail floors. *Asked:* $P(G > g)$ at a set of levels, and the value-at-risk and conditional value-at-risk at $\alpha = 0.95$. *TERRA:* evaluate L on the 100,000 reference draws (zero solves; any large draw from the belief would do) and read off the fraction above each level, the α-quantile and the mean above it. Because $L \le G$ pointwise, each of these is a lower bound on the same quantity of G. *Sampling:* the same statistics on a fresh sample of 10,000 solves, with binomial standard errors and no direction to the error.

Q9, worst case over the box. *Asked:* the largest shed any rating vector in the box can produce. *TERRA:* L is convex, so its maximum over the box is at a vertex; all $2^{14} = 16,384$ vertices are evaluated on the surrogate in chunks (zero solves), and the maximum is a lower bound on $\max_{\mathrm{box}} G$, attained at a candidate vertex. *Sampling:* the same enumeration on the model, 16,384 solves, which for a convex G is exact and serves as the reference.

3.5 References and scoring

Each answer is scored against a reference computed independently of both routes and at a larger budget: the mean gradient over 100,000 reference draws (Q1, Q8), Saltelli at $N = 4,000$ (Q2–Q4), fresh quasi-Monte Carlo studies of 20,000 solves (Q5, Q6), conditional Monte Carlo at 4,000 solves per grid point (Q7), and the exact vertex enumeration (Q9). The reference draws themselves

return gradients, so the Q1 reference is the same estimator as the TERRA answer at thirty-three times the sample.

4 Results

4.1 Cost

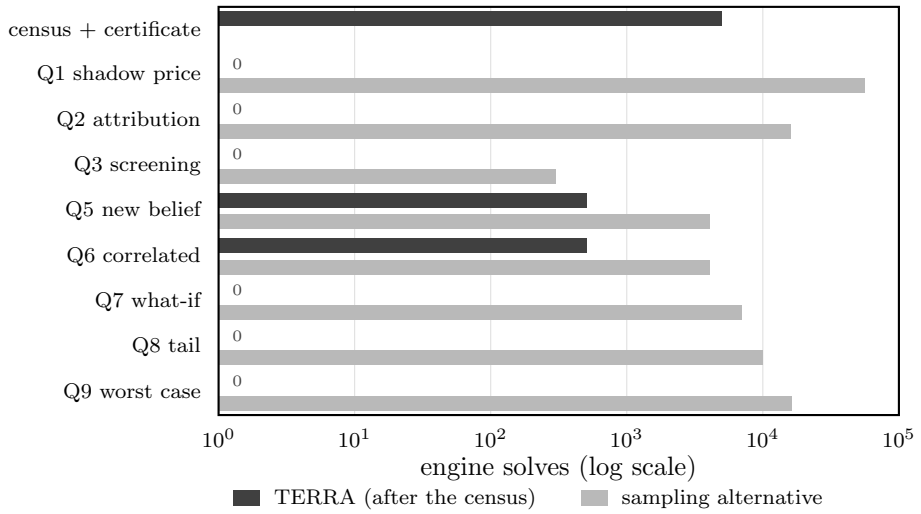


Figure 6: Engine solves per question: the surrogate route (after the census) against the standard sampling method at the budgets of Section 4. Q4 is answered from the same run as Q2 on both sides. Totals: 6,536 (TERRA, census included) and 113,876 (alternatives).

Table 2: Solve ledger. References are excluded.

question	TERRA solves	alternative	alternative solves
census (values and duals)	3,000	—	—
holdout certificate	2,000	—	—
Q1 shadow price of capacity	0	finite differences of the mean	56,000
Q2 variance attribution	0	Saltelli on the LP	16,000
Q3 screening	0	Morris elementary effects	300
Q4 value of measuring a line	0	the same Saltelli run	0
Q5 a new belief	512	fresh QMC study	4,096
Q6 correlated ratings	512	fresh QMC study	4,096
Q7 what-if curve	0	conditional Monte Carlo	7,000
Q8 tail floors	0	empirical tail	10,000
Q9 worst case over the box	0	solve every vertex	16,384
total	6,536		113,876

4.2 Expected shed under belief 1

For completeness, the question the census was paid for: the control variate $\mathbb{E}[G] = \mathbb{E}[L] + \mathbb{E}[G - L]$ with a 512-solve residual sample gives 214.19 ± 0.09 MW (95%) with the certified lower bound 213.53 MW, against the reference 214.28 ± 0.16 MW. Everything below is answered from the same census.

4.3 Q1: shadow price of capacity

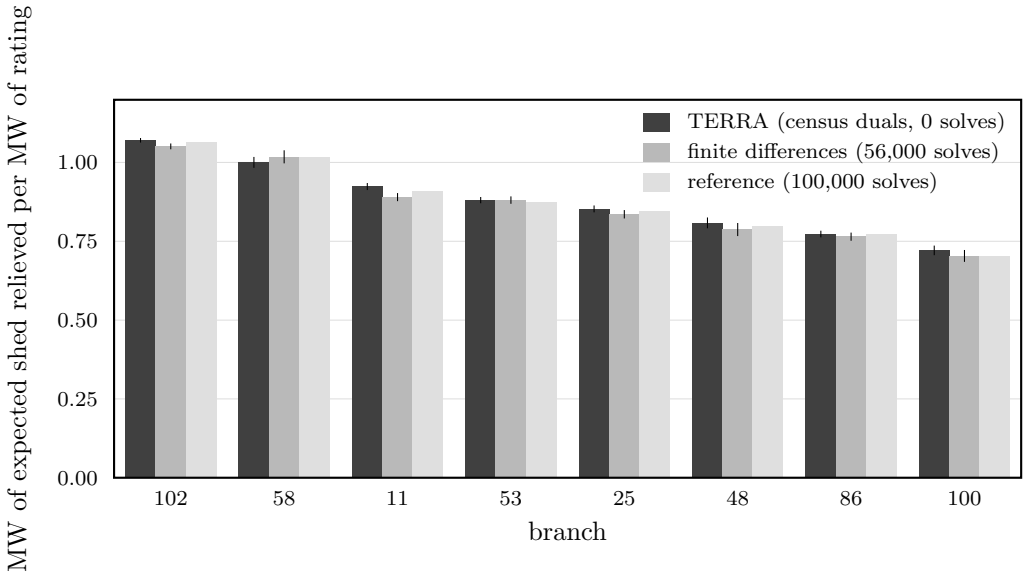


Figure 7: Q1: shadow price of capacity, $-\partial\,\mathbb{E}[G]/\partial r_i$, for the eight lines with the largest relief. Error bars are one standard error.

Table 3: Q1: shadow price of capacity, $\partial\,\mathbb{E}[G]/\partial r_i$ (MW of expected shed per MW of rating), for all 14 lines with their nominal (derated) ratings. TERRA: mean of the census duals, zero solves. Alternative: central differences of $\mathbb{E}[G]$ with common random numbers, 56,000 solves. Reference: mean dual over 100,000 draws.

branch	rating (MW)	TERRA	s.e.	alternative	s.e.	reference
102	250	-1.0704	0.0072	-1.0510	0.0092	-1.0656
58	200	-1.0003	0.0172	-1.0179	0.0208	-1.0157
11	88	-0.9237	0.0111	-0.8902	0.0128	-0.9094
53	88	-0.8805	0.0099	-0.8806	0.0117	-0.8728
25	250	-0.8528	0.0109	-0.8358	0.0133	-0.8464
48	200	-0.8083	0.0171	-0.7873	0.0207	-0.7965
86	200	-0.7730	0.0106	-0.7646	0.0130	-0.7734
100	250	-0.7211	0.0153	-0.7035	0.0188	-0.7036
64	250	-0.5785	0.0115	-0.5707	0.0140	-0.5815
96	200	-0.5151	0.0119	-0.5347	0.0149	-0.5194
91	88	-0.3209	0.0058	-0.3309	0.0072	-0.3281
89	88	-0.1021	0.0030	-0.1070	0.0037	-0.1073
30	250	-0.0469	0.0024	-0.0465	0.0028	-0.0481
107	250	-0.0291	0.0015	-0.0325	0.0021	-0.0295

The largest error of the zero-solve answer over the 14 lines is 0.0175 MW/MW; the finite-difference alternative, at 56,000 solves, reaches 0.0192. Both identify the same three lines as the most valuable to uprate (102, 58, 11). A shadow price of $-s$ means that one megawatt of additional rating on that line removes s megawatts of expected curtailment; multiplied by the hours of exposure and the value of lost load it prices a reconductoring, and the ranking across lines is the order in which such investments pay.

4.4 Q2–Q4: attribution, screening, value of measurement

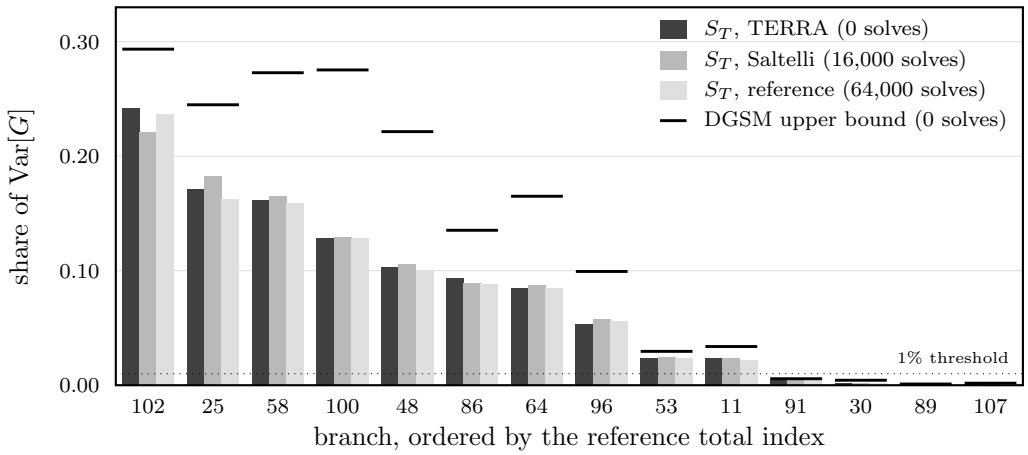


Figure 8: Q2 and Q3: total Sobol' index of every uncertain rating by the three routes, with the DGSM upper bound from the census duals as a cap on each group. A line whose cap lies below the 1% line is excluded with a guarantee.

Table 4: Q2–Q4 for all 14 lines, ordered by the reference total index. S_T: TERRA from the surrogate (zero solves), the alternative from Saltelli on the LP (16,000 solves), the reference from Saltelli at 64,000 solves. S_1 and the variance removed by measuring the line (MW^2): TERRA and reference. DGSM: the certified upper bound on S_T from the census; μ^*: the Morris mean elementary effect (300 solves). Negative first-order values are estimator noise around zero.

	S_T			S_1		value	DGSM	Morris
branch	TERRA	Saltelli	ref.	TERRA	ref.	(MW^2)	bound	μ^*
102	0.242	0.221	0.237	0.226	0.236	570	0.293	103.8
25	0.171	0.183	0.163	0.147	0.131	370	0.245	112.2
58	0.161	0.165	0.159	0.137	0.146	345	0.273	86.2
100	0.128	0.129	0.128	0.090	0.096	227	0.275	68.4
48	0.103	0.106	0.099	0.075	0.104	188	0.221	69.1
86	0.093	0.089	0.088	0.080	0.068	203	0.135	69.2
64	0.085	0.088	0.085	0.062	0.060	156	0.165	50.8
96	0.054	0.058	0.056	0.027	0.017	68	0.099	72.3
53	0.023	0.025	0.024	0.016	0.013	39	0.030	99.4
11	0.023	0.024	0.022	0.013	0.007	32	0.034	62.3
91	0.004	0.005	0.005	-0.003	-0.005	-7	0.006	47.1
30	0.002	0.002	0.002	-0.005	-0.009	-12	0.004	2.4
89	0.001	0.001	0.001	-0.005	-0.008	-12	0.001	10.6
107	0.001	0.001	0.001	-0.005	-0.009	-13	0.002	4.5

Total indices from the surrogate are within 0.009 of the reference on every line; Saltelli on the model at 16,000 solves is within 0.020. The three most influential lines are 102, 25, 58 by every route. First-order indices are within 0.029 (surrogate) and 0.044 (Saltelli) of the reference; their reference sum is 0.85, so about 15% of the variance is interaction among the ratings, which no single measurement removes. The DGSM bound lies above the reference total index on every line and excludes 4 lines (30, 89, 91, 107) below 1% with a guarantee. Morris, at 300 solves, places 30,

107, 89, 91 at the bottom against the reference's 107, 89, 30, 91: the same set, ranked, with no statement of how much any of them could matter.

4.5 Q5–Q6: a new belief, correlated ratings

Table 5: Q5–Q6: the census re-marginalized under a different input distribution. TERRA: $\mathbb{E}[L]$ under the new distribution plus a residual sample, with the 95% half-width and the certified lower bound; alternative: a fresh scrambled-Sobol' study (one standard error over four scrambles); reference: a larger fresh study. Belief 1 gave $\mathbb{E}[G] = 214.28$ MW.

	TERRA			alternative		reference
belief	$\mathbb{E}[G]$ (MW)	lower bound	solves	$\mathbb{E}[G]$ (MW)	solves	$\mathbb{E}[G]$ (MW)
belief 2, $U(0.80, 1.10)$	302.18 ± 0.10	301.40	512	302.05 ± 0.07	4,096	302.09 ± 0.02
shared weather, $\rho = 0.6$	202.28 ± 0.06	201.87	512	202.31 ± 0.08	4,096	202.19 ± 0.05

Under belief 2 the expected shed rises to 302.1 MW; the surrogate harvested under belief 1 recovers it within 0.09 MW from 512 residual solves, and a fresh quasi-Monte Carlo study of 4,096 solves is within 0.04 MW. Correlating the ratings lowers the mean to 202.2 MW — with a shared driver the low-rating draws coincide less often with the network's most congested combinations — and the census is reused within 0.08 MW (fresh study: 0.12). The certified lower bounds hold in both cases. Two points deserve emphasis. On the mean alone a fresh quasi-Monte Carlo study is comparably accurate, so the advantage of reuse here is the eight-fold saving in solves and the one-sided guarantee, not accuracy. And no solve made under belief 1 is wasted when the belief changes: convexity makes the planes valid everywhere, so the census is a permanent asset of the network, not of one forecast.

4.6 Q7: what-if curve

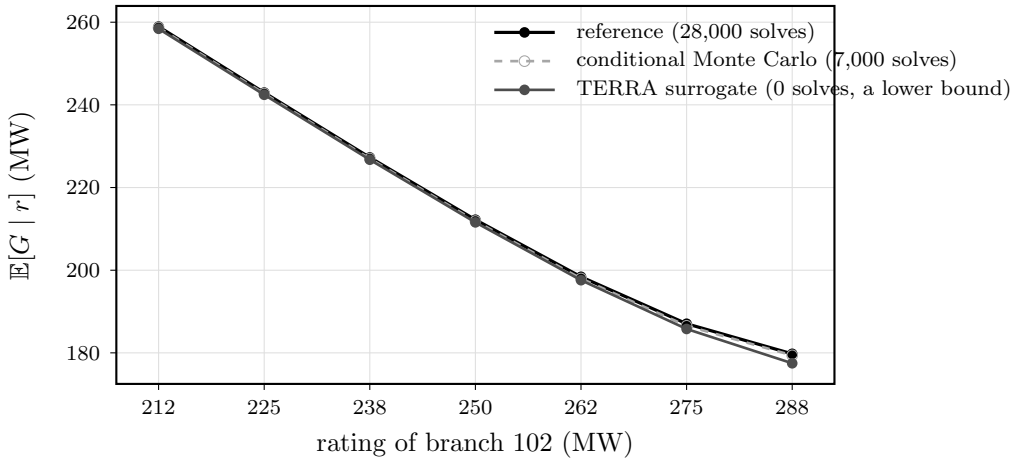


Figure 9: Q7: expected shed as a function of one line's rating, the other 13 ratings averaged over belief 1. The surrogate curve lies below the reference at every point by construction; its slope is the shadow price of that line.

Table 6 gives the values. The surrogate curve lies below the reference by at most 2.27 MW across the range, with slope -1.101 MW/MW against the reference's -1.081; conditional Monte Carlo at 7,000 solves deviates by up to 0.49 MW and is not one-sided. The slope is the shadow price of Q1 for this line, now seen across its whole rating range rather than at the mean.

Table 6: Q7: $\mathbb{E}[G \mid r_{102} = r]$ across the rating range of branch 102.

rating (MW)	TERRA (MW)	conditional MC (MW)	reference (MW)
212.5	258.41	258.72	258.93
225.0	242.45	242.74	242.97
237.5	226.73	227.01	227.27
250.0	211.57	211.89	212.21
262.5	197.58	198.01	198.40
275.0	185.78	186.56	187.04
287.5	177.49	179.39	179.76

4.7 Q8–Q9: tail and worst-case floors

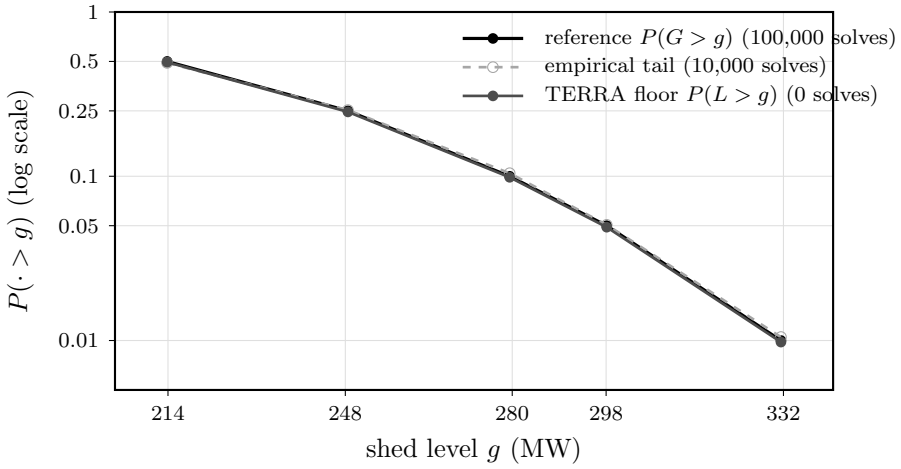


Figure 10: Q8: exceedance probability of the shed at the reference quantiles. The surrogate floor is below the reference at every level, as the certificate guarantees, and within its sampling noise of the empirical tail.

Table 7: Q8: exceedance probabilities at the reference quantiles. TERRA: $P(L > g)$, a lower bound with zero solves; alternative: the empirical tail of 10,000 solves with its binomial standard error; reference: 100,000 solves.

quantile	g (MW)	TERRA floor	empirical	reference
0.50	213.8	0.4950	0.4902 ± 0.0050	0.5000
0.75	248.5	0.2464	0.2526 ± 0.0043	0.2500
0.90	279.5	0.0985	0.1043 ± 0.0031	0.1000
0.95	298.1	0.0489	0.0506 ± 0.0022	0.0500
0.99	331.5	0.0098	0.0105 ± 0.0010	0.0100

The floors hold at every level. At $\alpha = 0.95$ the surrogate gives VaR ≥ 297.6 and CVaR ≥ 318.0 MW against the reference 298.1 and 318.4; the 10,000-solve empirical tail estimates 298.3 and 319.1 with no direction to its error. Over the whole rating box, the surrogate's maximum — 16,384 cheap evaluations, zero solves — is 471.7 MW at the same vertex as the exact worst case of 472.0 MW found by solving all 16,384 vertices; the gap is 0.3 MW. For comparison, the average over the vertices is 295.0 MW, well above the mean of 214.3 MW: the box's corners are not representative of its interior, which is why a two-point grid is not a substitute for either route.

5 Discussion

What the comparison establishes. Two things. First, the questions a planner asks after the mean are answered from the object that computed the mean, at zero or near-zero further solves, with accuracy at or above the sampling alternative's at budgets of thousands of solves each (Figure 6). Second, roughly half of those answers come with a direction: floors on the mean under a new belief, on the tail and on the worst case, and a ceiling on every attribution. Sampling produces intervals; the model produces inequalities. For the instrumentation decision, "line 89 cannot account for more than 0.1% of the variance" is a different kind of statement from "our sample suggests about 0.1%", and for a compliance conversation "at least" is what can be signed.

What it rests on. Three conditions, each a boundary of the claim. The solve must return its sensitivities: here the linear program's duals, at no cost; a model that returns only a value would pay $2K + 1$ solves per plane by finite differences and, at a kink, would not even obtain a subgradient. The consequence must be convex in the inputs for the planes to lie below it: true of the load-shedding program's value, not of an AC power flow, where the same machinery yields estimates without floors. And every certified statement is one-sided: the surrogate never overestimates, so upper tails and worst cases are bounded from below only, and the upper direction remains statistical.

Where sampling keeps its place. On the expectation itself a well-constructed quasi-Monte Carlo study is as accurate as the surrogate route at equal solves (Q5, Q6 show this directly), and sampling needs neither gradients nor convexity. The surrogate's advantage is not the first number; it is every number after the first, and the guarantees attached to them.

What generalizes. Nothing in the construction refers to ratings or to this network. A census is a set of (input, value, gradient) triples from any convex consequence of any model whose solver exposes a gradient, and the nine questions are functionals of the surrogate. The same calls (Appendix C) apply to a thermal or a cooling model once its solve returns sensitivities.

6 Conclusions

- One census of 3,000 solves, certified on 2,000 holdout draws, answers nine planning questions on a 73-bus test system with 14 uncertain ratings at 1,536 further solves in total; the sampling alternatives spend 113,876 for the same questions at comparable or lower accuracy.
- Every one-sided claim the surrogate makes — 4 lines certified below 1% of the variance, floors on the tail, on the worst case, and on the mean under two new beliefs — holds on the reference.

- The capability requires the solve to return its sensitivities and the consequence to be convex; outside those conditions the same calls return estimates without certificates, and on the expectation alone quasi-Monte Carlo sampling is an equal.

A The load-shedding program

With p_g the generation, p_ℓ the branch flows, θ_b the bus angles and $x_i \in [0,1]$ the served fraction of load d_i, all in per unit on a 100 MVA base,

$$
\begin{aligned}
G(r) \;=\; \min\ & \textstyle\sum_i d_i\,(1-x_i) \\
\text{s.t. }\ & \textstyle\sum_{\ell:\text{from}=b} p_\ell - \sum_{\ell:\text{to}=b} p_\ell - \sum_{g\in b} p_g + \sum_{i\in b} d_i x_i = 0 && \forall \text{ bus } b,\\
& X_\ell\, p_\ell - \theta_{\text{from}(\ell)} + \theta_{\text{to}(\ell)} = 0 && \forall \text{ branch } \ell,\\
& \theta_{\text{ref}} = 0, \quad p_g^{\min} \le p_g \le p_g^{\max}, \quad 0 \le x_i \le 1, \\
& -r_\ell \le p_\ell \le r_\ell && \forall \text{ branch } \ell.
\end{aligned}
$$

The ratings enter only as the flow bounds r_ℓ. By linear-programming duality the value is convex and piecewise linear in r, and $\partial G/\partial r_\ell$ is the dual of the active flow bound (with the sign reversed at the lower bound). The uncertain vector is the sub-vector of r on the K branches whose bounds are active at the derated point; the remaining ratings are held at their derated values.

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, $(K+1)n$ and $(K+2)n$ evaluations. The derivative bound for $r_i \sim U(a_i, b_i)$ is $S_{T,i} \le (b_i - a_i)^2\, \mathbb{E}[(\partial f/\partial r_i)^2]/(\pi^2 V)$ (Sobol' and Kucherenko, 2009), whose constant $1/\pi^2$ is the Poincaré constant of the uniform law; the second moment is the mean of the squared census gradients. Morris's screening (Morris, 1991) uses R trajectories that change one input at a time by a fixed step Δ 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 in R independent scrambles, the second over a residual sample of solves in R scrambles; the 95% half-width combines the two replicate spreads in quadrature, and the lower bound is $\mathbb{E}[L]$'s one-sided 95% limit since $\mathbb{E}[G - L] \ge 0$. For the tail, VaR_α is the α-quantile and CVaR_α the mean above it (Rockafellar and Uryasev, 2000); both are monotone in the pointwise order, so their values on L bound their values on G from below.

C The software calls

Table 8: The software call behind each question (`plugins.uncertainty` unless noted).

question	call	from `plugins.power_flow`
census	`harvest`, `holdout_check`	`MinShedRatingEvaluator`, `rating_sensitive_lines`
$\mathbb{E}[G]$	`control_variate`	
Q1	`mean_gradient`	
Q2	`sobol_total`	
Q3	`dgsm_bound`	
Q4	`sobol_first`	
Q5	`control_variate` with a second `uniform_box`	
Q6	`control_variate` with `gaussian_copula`	
Q7	`conditional_mean`	
Q8	`exceedance_floor`	
Q9	`worst_case_floor`	

References

C. Barrows et al., "The IEEE Reliability Test System: A Proposed 2019 Update," *IEEE Transactions on Power Systems*, vol. 35, no. 1, pp. 119–127, 2020.

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. 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.

Reproduction. With `PYTHONPATH=src`, run `run.py` then `make_report.py` with `--pdf`. Seed 20260908, backend `highs`, 352,413 engine solves including references. Every number above is read from `results.json`; the schematic figures carry no data.