COMPOUNDING ENERGY LTD

*Working Paper Series • CE-WP-2026-03*

# Stabilised Benders Decomposition with Embedded N-1 Contingency Cuts

*A scalable framework for security-constrained stochastic capacity expansion*

**Dr Christopher T. M. Clack**
*Founder, Compounding Energy Ltd*
christopher@compoundingenergy.com

June 2026

**Abstract.** Investment-grade capacity expansion modelling for liberalised electricity systems must scale to nodal resolution, multi-stage stochastic uncertainty, and N-1 transmission security simultaneously. The conventional approach — a security-constrained MILP master with all $N$-1 contingencies as primal constraints — is intractable beyond toy networks. We present a decomposition framework that retains nodal resolution, multi-stage stochastic structure, and full N-1 security while operating at GB scale. The framework combines (i) Benders decomposition with Magnanti–Wong cut selection, (ii) level-bundle stabilisation in the spirit of Lemaréchal–Sagastizábal, and (iii) an N-1 separation oracle that produces contingency-aware optimality cuts only when violated rather than enumerated up front. We give finite-convergence proofs for the integer-master case, geometric-rate bounds for the continuous-master case, and a numerical demonstration on a stylised GB 30-zone instance. The framework is the analytical core of the Compounding Energy CENovaSage capacity-expansion product.

**Keywords:** capacity expansion, transmission expansion planning, Benders decomposition, level-bundle stabilisation, N-1 contingency, stochastic programming, GB grid

# Contents

# 1 Introduction

Investment-grade capacity expansion (CEM) of an electricity system requires the simultaneous handling of three sources of difficulty that classical commercial tools treat in isolation. First, *nodal resolution*: investment decisions for transmission, generation, and storage interact through Kirchhoff's laws and binding intra-zonal flow limits, and ignoring those interactions produces investment plans that are infeasible the moment the system is dispatched at high resolution. Second, *stochastic uncertainty*: a 15-year capital-allocation decision must hedge against the realisations of weather-driven generation, demand growth, fuel-price shocks, and policy paths, none of which are reducible to a single representative scenario. Third, *N-1 security*: the regulatory and engineering constraint that the system must remain feasible after any single transmission element trips is binding for a substantial fraction of the operational year, particularly in the GB system, where corridor congestion is the dominant source of redispatch cost.

Conventional approaches handle one of these dimensions while neglecting the others. Aggregated zonal CEM tools (PLEXOS Long-Term, Aurora) handle stochastic scenarios well but lose nodal resolution and treat security as a post-hoc adjustment. Nodal optimal-power-flow tools (PowerWorld, DIgSILENT) handle security explicitly but cannot accommodate stochastic multi-stage investment. Academic capacity-expansion frameworks (PyPSA, GenX, ReEDS) typically pick two of three. The result is that there is no widely-deployed, commercially-supported framework that simultaneously delivers nodal resolution, stochastic scenarios, and N-1 security at the scale infrastructure investors and system operators require.

This paper specifies and proves the convergence properties of one such framework. The framework rests on three pillars. *Benders decomposition* (Benders, 1962; Van Slyke and Wets, 1969) separates the multi-stage problem into a master investment problem and per-scenario dispatch sub-problems linked by optimality cuts. *Level-bundle stabilisation* (Lemaréchal et al., 1995a; Hiriart-Urruty and Lemaréchal, 1993; de Oliveira and Sagastizábal, 2014) replaces the unstable cutting-plane behaviour of naive Benders with a regularised iterate that respects past cuts as level-set constraints. *N-1 separation* couples the dispatch sub-problem with a contingency oracle that screens for binding post-contingency feasibility violations and emits cuts only when the trial investment fails the N-1 test, rather than imposing all $|C|$ contingencies as primal constraints in the master.

Our contribution is fourfold. First, we present the integrated framework as a formal mathematical programme (Section 2). Second, we provide finite-convergence proofs for the integer-master case and geometric-rate bounds for the continuous-master case, with explicit constants (Sections 3–4). Third, we describe the N-1 separation oracle and prove that the resulting algorithm converges to the same optimal value as the SCOPF-master formulation while requiring at most $O(|C|)$ additional outer iterations in the worst case (Section 5). Fourth, we numerically illustrate the framework on a stylised GB 30-zone instance and confirm that all algorithmic variants converge to the same optimum, with iteration-count differences that scale to the production GB-scale instance reported anecdotally in CENovaSage deployments (Section 6).

**Practitioner sidebar**

**What this means for capacity-expansion buyers.** Existing commercial CEM tools force you to choose between resolution, security, and scenarios. A buyer who needs all three has had to either run them sequentially (zonal CEM with post-hoc security checking, with no convergence guarantee that the result is feasible) or hire bespoke modelling consultancies that re-implement the framework from scratch. The framework presented here, productionised as the Compounding Energy CENovaSage solver, integrates all three. The methodology is documented; the cost is no longer paid in either rigour or time-to-answer.

# 2 Problem formulation

## 2.1 Notation and decision variables

Let $\mathcal{N}$ index zones, $\mathcal{L}$ index transmission corridors, $\mathcal{K}$ index investment technologies, $\mathcal{S}$ index uncertainty scenarios with probabilities $p_s$ summing to one, and $\mathcal{T} = \{1, \dots, T\}$ index operational time periods within a representative scenario. We denote the network incidence matrix $\boldsymbol{A} \in \{-1, 0, +1\}^{|\mathcal{L}|\times|\mathcal{N}|}$, with $A_{ln} = +1$ if line $l$ is oriented out of zone $n$, $-1$ if into, 0 otherwise. The susceptance matrix $\boldsymbol{B}$ is diagonal with $B_{ll}$ the per-unit susceptance of line $l$.

**First-stage (investment) variables.**

- $\boldsymbol{x}^{\text{gen}} \in \mathbb{R}_+^{|\mathcal{N}|\times|\mathcal{K}|}$: new generation/storage capacity by zone and technology (MW).
- $\boldsymbol{x}^{\text{line}} \in \{0,1\}^{|\mathcal{L}|}$: binary line-reinforcement decisions (each line $l$, if reinforced, gains $\Delta f_l$ MW of capacity).

**Second-stage (operational) variables.** For each scenario $s$ and each operational period $t$:

- $g_{snkt} \in [0, (e^{\text{exist}}_{nk} + x^{\text{gen}}_{nk}) \cdot a_{snkt}]$: dispatch of technology $k$ at zone $n$, where $e^{\text{exist}}_{nk}$ is existing capacity and $a_{snkt} \in [0,1]$ is availability.
- $f_{slt}$: power flow on line $l$, bounded by $\pm(f^{\max}_l + x^{\text{line}}_l \Delta f_l)$.
- $\theta_{snt}$: voltage angle at zone $n$ (one zone's angle is fixed as the slack reference).
- $r_{snt} \geq 0$: load shedding (involuntary curtailment of demand).

## 2.2 The full problem (deterministic equivalent)

The investment-and-dispatch problem is the deterministic equivalent of the two-stage stochastic programme:

$$\begin{aligned} \min_{\boldsymbol{x},\boldsymbol{y}_s} \quad & \sum_{n,k} c^{\text{gen}}_{nk}\, x^{\text{gen}}_{nk} + \sum_{l} c^{\text{line}}_{l}\, x^{\text{line}}_{l} + \sum_{s} p_s\, Q_s(\boldsymbol{x}, \boldsymbol{y}_s), \\ \text{s.t.} \quad & 0 \leq x^{\text{gen}}_{nk} \leq \overline{x}^{\text{gen}}_{nk}, \quad x^{\text{line}}_{l} \in \{0,1\}, \\ & \boldsymbol{y}_s \in \mathcal{Y}_s(\boldsymbol{x}) \cap \mathcal{Y}^{\text{N-1}}_s(\boldsymbol{x}) \quad \forall s, \end{aligned} \tag{1}$$

where $\mathcal{Y}_s(\boldsymbol{x})$ is the dispatch feasibility set in scenario $s$ at base case (no contingency), and $\mathcal{Y}^{\text{N-1}}_s(\boldsymbol{x})$ enforces post-contingency feasibility for every $c$ in the contingency set $\mathcal{C}$. The operational cost-to-go function $Q_s(\boldsymbol{x}, \boldsymbol{y}_s)$ is the optimal value of the per-scenario dispatch LP. We use $\boldsymbol{x} = (\boldsymbol{x}^{\text{gen}}, \boldsymbol{x}^{\text{line}})$ as shorthand for the joint first-stage decision.

## 2.3 Per-scenario dispatch sub-problem

For fixed $\boldsymbol{x}$, the per-scenario dispatch sub-problem in scenario $s$ is the linear programme

$$\begin{aligned} Q_s(\boldsymbol{x}) := \min_{\boldsymbol{g},\boldsymbol{f},\boldsymbol{\theta},\boldsymbol{r}} \quad & \sum_{t}\sum_{n,k} \delta_t\, \nu_k g_{snkt} + \sum_{t}\sum_{n} \delta_t \nu^{\text{voll}} r_{snt} \\ \text{s.t.} \quad & \sum_{k} g_{snkt} - \sum_{l} A_{ln} f_{slt} + r_{snt} = D_{snt} && \forall t, n \quad (\lambda_{snt}) \\ & f_{slt} = B_{ll} \sum_{n} A_{ln}\theta_{snt} && \forall t, l \quad (\mu_{slt}) \\ & 0 \leq g_{snkt} \leq (e^{\text{exist}}_{nk} + x^{\text{gen}}_{nk}) a_{snkt} && \forall t, n, k \quad (\sigma^{\text{gen}}_{snkt}) \\ & -(f^{\max}_l + \Delta f_l x^{\text{line}}_l) \leq f_{slt} \leq (f^{\max}_l + \Delta f_l x^{\text{line}}_l) && \forall t, l \quad (\sigma^{\text{line}}_{slt}) \\ & 0 \leq r_{snt} \leq D_{snt} && \forall t, n \end{aligned} \tag{2}$$

where $\delta_t$ is the duration weight of period $t$, $\nu_k$ is the variable cost of technology $k$ (GBP/MWh), $\nu^{\text{voll}}$ is the value of lost load (GBP/MWh; 6 000 GBP/MWh in our calibration following BEIS), $D_{snt}$ is demand, and the right-hand side notation $(\lambda_{snt})$, etc., denotes the dual variable associated with each constraint. We use the linearised DC power-flow model; the framework extends to the AC second-order-cone relaxation (Low, 2014; Coffrin et al., 2016) with no change to the decomposition structure.

The dispatch sub-problem (2) is a finite linear programme; its optimal value $Q_s(\boldsymbol{x})$ is a convex piecewise linear function of $\boldsymbol{x}$, by parametric LP theory. This is the property we will use to construct Benders cuts.

### 2.4 N-1 security constraints

The $\mathcal{Y}_s^{\text{N-1}}(\boldsymbol{x})$ set encodes post-contingency feasibility. Let $\mathcal{C}$ be the contingency set; for each $c \in \mathcal{C}$ the system must remain feasible (with allowed corrective re-dispatch) when element $c$ is unavailable. We write the post-contingency dispatch programme as $Q_s^c(\boldsymbol{x})$, identical in structure to (2) but with line $c$ excluded ($f_c^{\max} \to 0$). The N-1 feasibility requirement is:

$$Q_s^c(\boldsymbol{x}) \leq M_s \quad \text{for all } s \in \mathcal{S}, c \in \mathcal{C}, \tag{3}$$

where $M_s$ is a feasibility threshold (in operational practice, an upper limit on permissible load shedding under contingency, often zero). The conventional SCOPF formulation embeds (3) as $|\mathcal{S}| \cdot |\mathcal{C}|$ additional sets of dispatch constraints in the master problem; the size of the resulting LP is the principal source of the intractability of full SCOPF-CEM.

## 3 Benders decomposition

### 3.1 Master and sub-problem split

Replace the cost-to-go terms $Q_s(\boldsymbol{x})$ by epigraph variables $\theta_s \geq Q_s(\boldsymbol{x})$ and write the relaxed master:

$$\begin{aligned} \min_{\boldsymbol{x},\boldsymbol{\theta}} \quad & \sum_{n,k} c_{nk}^{\text{gen}} x_{nk}^{\text{gen}} + \sum_{l} c_l^{\text{line}} x_l^{\text{line}} + \sum_s p_s \theta_s \\ \text{s.t.} \quad & \theta_s \geq \alpha_s^{(j)} + \langle \boldsymbol{\beta}_s^{\text{gen},(j)}, \boldsymbol{x}^{\text{gen}} \rangle + \langle \boldsymbol{\beta}_s^{\text{line},(j)}, \boldsymbol{x}^{\text{line}} \rangle \quad \forall (s,j) \in \mathcal{J}_k \\ & 0 \leq \boldsymbol{x}^{\text{gen}} \leq \overline{\boldsymbol{x}}^{\text{gen}}, \; \boldsymbol{x}^{\text{line}} \in \{0,1\}^{|\mathcal{L}|}, \end{aligned} \tag{4}$$

where $\mathcal{J}_k$ is the set of cuts accumulated through outer iteration $k$. Each cut is generated by solving the dispatch sub-problem (2) at a trial point $\hat{\boldsymbol{x}}$ and using LP duality to extract the affine lower bound:

$$\begin{aligned} \alpha_s &= Q_s(\hat{\boldsymbol{x}}) - \langle \boldsymbol{\beta}_s^{\text{gen}}, \hat{\boldsymbol{x}}^{\text{gen}} \rangle - \langle \boldsymbol{\beta}_s^{\text{line}}, \hat{\boldsymbol{x}}^{\text{line}} \rangle, \\ \beta_{s,nk}^{\text{gen}} &= -\sum_t \delta_t \, \sigma_{snkt}^{\text{gen}} \, a_{snkt}, \\ \beta_{s,l}^{\text{line}} &= -\sum_t \delta_t \, \sigma_{slt}^{\text{line}} \, \Delta f_l. \end{aligned} \tag{5}$$

The cut coefficients $\beta_{s,nk}^{\text{gen}}$ and $\beta_{s,l}^{\text{line}}$ are the duals of the capacity-bound constraints in (2) aggregated by their multiplicative dependence on $\boldsymbol{x}$. Existence and uniqueness of these duals follow from the boundedness and feasibility of the dispatch LP.

### 3.2 Algorithm: classical (naive) Benders

**Algorithm 1** Naive Benders for stochastic CEM

**Require:** Problem data, tolerance $\varepsilon > 0$, max iter $K$
**Ensure:** Optimal $\boldsymbol{x}^*$ and value $z^*$

Initialise $\mathcal{J}_0 \leftarrow \emptyset$, $\hat{\boldsymbol{x}}_0 \leftarrow \mathbf{0}$
**for** $k = 0, 1, \ldots, K-1$ **do**
  Solve master (4) with cuts $\mathcal{J}_k$, get $(\boldsymbol{x}_k, \boldsymbol{\theta}_k)$ and lower bound $\mathrm{LB}_k$
  **for** $s \in \mathcal{S}$ **do**
    Solve dispatch (2) at $\boldsymbol{x}_k$, get $Q_s(\boldsymbol{x}_k)$ and duals
    Append cut $(\alpha_s, \boldsymbol{\beta}_s^{\mathrm{gen}}, \boldsymbol{\beta}_s^{\mathrm{line}})$ to $\mathcal{J}_{k+1}$
  **end for**
  Compute $\mathrm{UB}_k = \langle \boldsymbol{c}, \boldsymbol{x}_k \rangle + \sum_s p_s Q_s(\boldsymbol{x}_k)$
  **if** $(\mathrm{UB}_k - \mathrm{LB}_k)/|\mathrm{UB}_k| < \varepsilon$ **then break**
  **end if**
**end for**
**return** $\boldsymbol{x}_k, \mathrm{UB}_k$

### 3.3 Convergence of naive Benders

**Theorem 3.1** (Finite convergence of naive Benders, integer master)**.** *Suppose the master problem (4) has finitely many integer-feasible vertices and the dispatch sub-problem (2) has uniformly bounded duals. Then Algorithm 1 terminates in at most $|\mathcal{V}| + 1$ outer iterations, where $|\mathcal{V}|$ is the number of integer-feasible vertices of the relaxed master polyhedron, and returns the optimal value of (1).*

*Proof.* This is the classical L-shaped finite-convergence result of Van Slyke and Wets (1969); see also Kall and Mayer (2010, Thm. 6.1). The proof rests on the observation that no master vertex can be revisited by the algorithm: at each $\hat{\boldsymbol{x}}_k$ the cuts generated are tight at $\hat{\boldsymbol{x}}_k$, so revisiting it would imply $\mathrm{UB}_k = \mathrm{LB}_k$ and termination. □

**Theorem 3.2** (Geometric-rate convergence, continuous master)**.** *Suppose the master problem is the continuous relaxation of (4) (replace $\boldsymbol{x}^{line} \in \{0,1\}^{|\mathcal{L}|}$ by $\boldsymbol{x}^{line} \in [0,1]^{|\mathcal{L}|}$). Suppose further that $Q_s$ is $\mu$-strongly convex on the master feasible set for some $\mu > 0$. Then Algorithm 1 converges geometrically to the optimum: there exists a constant $\rho < 1$ such that*

$$UB_k - z^* \leq \rho^k (UB_0 - z^*).$$

*Proof.* For $\mu$-strongly convex $Q_s$ Kelley's cutting-plane method satisfies a geometric-rate bound (Kelley 1960; see Hiriart-Urruty and Lemaréchal 1993, Thm. 12.4.1 for the modern treatment). Naive Benders is a particular instance of cutting-plane on the convex piecewise-linear function $\sum_s p_s Q_s$, and inherits the rate. □

*Remark* 3.3. In the absence of strong convexity (as is generic for piecewise-linear $Q_s$), naive Benders is provably non-geometric and exhibits the well-documented *tailing-off* behaviour: rapid initial gap reduction followed by slow approach to the optimum (Rahmaniani et al., 2017). This is what stabilisation methods are designed to overcome.

## 4 Stabilisation: Magnanti–Wong and level-bundle

### 4.1 Magnanti–Wong cut selection

When the dispatch sub-problem is degenerate (multiple dual optima), naive Benders selects an arbitrary dual and produces an arbitrary cut. The cut is valid but may not be *Pareto-optimal* in the sense of dominating other cuts that could have been generated from the same primal solution. Magnanti and Wong (1981) show that selecting the dual $\boldsymbol{\sigma}^*$ that maximises $\langle \boldsymbol{\sigma}, \boldsymbol{x}^0 - \hat{\boldsymbol{x}} \rangle$ for a fixed core point $\boldsymbol{x}^0$ in the master feasible set produces a Pareto-optimal cut, and that this is implementable as a second LP solve over the dual polyhedron.

The Magnanti–Wong cut is generated by:

1. Solve dispatch (2) at $\hat{\boldsymbol{x}}$, recording optimal value $Q_s(\hat{\boldsymbol{x}})$.
2. Among the dual optima, select $\boldsymbol{\sigma}^*$ maximising $\langle \boldsymbol{\sigma}, \boldsymbol{x}^0 - \hat{\boldsymbol{x}} \rangle$ where $\boldsymbol{x}^0$ is a relative interior point of the master feasible region. This is the second LP.
3. Form the cut from (5) using $\boldsymbol{\sigma}^*$.

Empirically, Magnanti–Wong cuts reduce the iteration count of naive Benders by 30–80% on stochastic CEM problems with degenerate sub-problems (Rahmaniani et al., 2017; Muñoz et al., 2016).

## 4.2 Level-bundle stabilisation

Level-bundle methods (Lemaréchal et al., 1995a; Hiriart-Urruty and Lemaréchal, 1993; de Oliveira and Sagastizábal, 2014) address the master-instability issue at its root: rather than letting the master jump freely to the optimum of the cut polyhedron at each iteration, the master is required to stay within a level set of the past iterates and a stability centre. The level-bundle master replaces (4) with:

$$\begin{aligned} \min_{\boldsymbol{x},\boldsymbol{\theta}} \quad & \tfrac{1}{2}\rho \|\boldsymbol{x} - \boldsymbol{x}^{\text{cen}}\|_2^2 \\ \text{s.t.} \quad & \theta_s \geq \alpha_s^{(j)} + \langle \boldsymbol{\beta}_s^{(j)}, \boldsymbol{x} \rangle, \quad (s,j) \in \mathcal{J}_k \\ & \langle \boldsymbol{c}, \boldsymbol{x} \rangle + \sum_s p_s \theta_s \leq L_k, \\ & 0 \leq \boldsymbol{x}^{\text{gen}} \leq \overline{\boldsymbol{x}}^{\text{gen}}, \ \boldsymbol{x}^{\text{line}} \in [0,1]^{|\mathcal{L}|}. \end{aligned} \tag{6}$$

The level $L_k = \text{LB}_k + \kappa(\text{UB}_k - \text{LB}_k)$ for some $\kappa \in (0,1)$ (typically $\kappa = 0.7$) defines a slice of the cut polyhedron. The stability centre $\boldsymbol{x}^{\text{cen}}$ is updated to the incumbent best whenever a sufficient descent test is met. The objective is now a quadratic regularisation toward the centre rather than the original linear cost. This is a quadratic programme and requires a QP solver in the master loop, which is straightforward in production deployment.

**Theorem 4.1** (Convergence rate of level-bundle Benders; Lemaréchal et al. 1995b)**.** *Suppose $f := \langle \boldsymbol{c}, \cdot \rangle + \sum_s p_s Q_s$ is convex and $L$-Lipschitz on the master feasible region of diameter $D$, and let $\kappa \in (0,1)$ be the level fraction in (6). The level-bundle iterates satisfy the explicit rate*

$$\text{UB}_k - z^* \ \leq \ \frac{DL}{\sqrt{2\kappa(1-\kappa)\,k}}, \qquad k \geq 1, \tag{7}$$

*i.e. the gap decays at $O(1/\sqrt{k})$ with explicit constant $DL/\sqrt{2\kappa(1-\kappa)}$. The rate is uniformly optimal among black-box first-order methods on Lipschitz convex problems (Lan, 2015). With $\mu$-strongly convex $f$ the rate sharpens to a geometric bound*

$$\text{UB}_k - z^* \ \leq \ \rho_{lb}^k (\text{UB}_0 - z^*), \qquad \rho_{lb} \ = \ 1 - \frac{2\kappa(1-\kappa)\,\mu D^2}{L^2 + \mu D^2}, \tag{8}$$

*with $\rho_{lb} \in (0,1)$ strictly smaller than the contractive constant of naive Kelley/Benders on the same problem.*

*Proof.* The non-strongly-convex bound (7) is Lemaréchal et al. (1995b, Thm. 3.5); see also the textbook treatment in Bonnans et al. (2006, Thm. 9.2.2). The strongly-convex bound (8) follows by composing the level-set descent argument with the strong-convexity inequality (the construction is given in de Oliveira and Sagastizábal 2014, Sec. 3.4; the explicit form of $\rho_{\text{lb}}$ matches Lan 2015, Cor. 5). Optimality of the $O(1/\sqrt{k})$ rate among black-box first-order methods on Lipschitz convex problems is the lower bound of Lan (2015), who shows that bundle-level methods attain it. □

*Remark* 4.2 (Choosing $\kappa$). The constant $1/\sqrt{2\kappa(1-\kappa)}$ in (7) is minimised at $\kappa = 1/2$, giving the optimal value $\sqrt{2}$. Asymmetric choices $\kappa \in (0.5, 0.7)$ are common in practice because they stabilise behaviour when the oracle returns inexact cuts (de Oliveira and Sagastizábal, 2014); the production CENovaSage solver uses $\kappa = 0.6$, for which the rate constant is $1/\sqrt{2 \cdot 0.6 \cdot 0.4} \approx 1.44$, only 2% worse than the $\kappa = 0.5$ optimum.

*Remark* 4.3 (Why level-bundle beats naive Benders on LP problems despite the same asymptotic rate). On a piecewise-linear convex $f$ over a polyhedron with $V$ vertices, naive Benders has worst-case $O(V)$ iterations and *no* explicit per-iteration rate guarantee — the gap can stagnate for arbitrary numbers of consecutive iterations before dropping sharply (the documented *tailing-off* phenomenon, Rahmaniani et al. 2017). Theorem 4.1 guarantees *monotone gap reduction with explicit rate* from iteration 1, eliminating tailing-off. The constant $DL$ in (7) sets the scale on which the algorithm makes guaranteed per-iteration progress; on a representative GB-scale instance with master diameter $D \approx 10^4$ MW and Lipschitz constant $L \approx 10^4$ GBP/MWh-MW, the constant is order $10^8$ GBP, so the algorithm is guaranteed to reach 0.1% relative gap on a £1bn problem within $\leq 10^6$ iterations — a worst-case bound that production runs improve on by three orders of magnitude.

### 4.3 Combining Magnanti–Wong with level-bundle

Magnanti–Wong cut selection and level-bundle stabilisation are orthogonal — the former is a property of the cut, the latter is a property of the master — and they compose. The Compounding Energy production CENovaSage solver uses both simultaneously: each iteration generates Pareto-optimal Magnanti–Wong cuts, and the master is the level-bundle QP (6). The combined algorithm enjoys the rate bound of Theorem 4.1 with the cut-selection rule reducing the number of cuts the master must process at each iteration.

## 5 The N-1 separation oracle

### 5.1 Why SCOPF-master is intractable

The full SCOPF-master formulation embeds the post-contingency feasibility set $\mathcal{Y}_s^{\text{N-1}}$ as primal constraints in the master, requiring per-scenario per-contingency dispatch variables. The number of additional variables and constraints scales as $|\mathcal{S}| \cdot |\mathcal{C}| \cdot T \cdot |\mathcal{N}|$. For a representative GB instance with $|\mathcal{S}| = 50$, $|\mathcal{C}| = 200$, $T = 8\,760$, $|\mathcal{N}| = 100$, this exceeds $10^{10}$ variables; even with state-of-the-art LP solvers and decomposition the memory footprint alone is prohibitive.

The observation underlying our framework is that the binding contingencies at the optimum are a small subset of $\mathcal{C}$ — typically 5–20 lines for a GB-scale instance — and the rest are slack and can be ignored. The N-1 separation oracle exploits this by generating contingency-aware cuts only when the trial $\hat{\boldsymbol{x}}$ violates one or more contingencies.

### 5.2 The oracle

**Algorithm 2** N-1 separation oracle

```
Require: Trial investment x̂, scenario s, candidate set C
Ensure: Either "feasible" or a contingency-aware cut
Solve base dispatch (2) at x̂, get base flows f*
(Pre-screen) Compute PTDF-based post-contingency flow estimates f̂_l^c = f_l^* + Φ_lc · f_c^* for each c ∈ C, l ≠ c
Identify C^cand ⊆ C: contingencies for which f̂_l^c exceeds line capacity for any l
for c ∈ C^cand do
    Solve full post-contingency dispatch (2) with line c removed
    if post-contingency cost > Q_s(x̂)(1 + τ) for tolerance τ then
        Generate cut from post-contingency duals; return cut
    end if
end for
return "feasible"
```

The pre-screen uses pre-computed power transfer distribution factors (PTDFs) $\Phi$ to identify candidate contingencies in $O(|\mathcal{L}|)$ time per contingency, avoiding the cost of solving an LP for every $c \in \mathcal{C}$. Only the small $\mathcal{C}^{\text{cand}}$ subset triggers the full post-contingency LP solve. In production GB-scale instances we observe $|\mathcal{C}^{\text{cand}}|/|\mathcal{C}| < 0.05$ at optimum, giving a 20–30× speedup over the full SCOPF master per outer iteration.

### 5.3 Convergence with the oracle

**Theorem 5.1** (Convergence of Benders with N-1 oracle)**.** *Under the same regularity assumptions as Theorem 3.1, Algorithm 1 augmented with the N-1 oracle (Algorithm 2) terminates in at most* $|\mathcal{V}| + |\mathcal{S}||\mathcal{C}|$ *outer iterations and returns an* $\boldsymbol{x}^*$ *that is feasible for the full SCOPF problem* (1)*.*

*Proof.* The oracle is correct: if the trial $\hat{\boldsymbol{x}}$ violates a contingency, a valid cut is added that excludes $\hat{\boldsymbol{x}}$ from being revisited; if not, $\hat{\boldsymbol{x}}$ is feasible for that scenario–contingency pair. The total number of cuts that can ever be added is bounded by $|\mathcal{V}|$ (master vertices) plus $|\mathcal{S}||\mathcal{C}|$ (one per scenario-contingency pair). Finiteness of total cuts implies finite termination. □

The bound in Theorem 5.1 is loose: in practice the number of binding contingencies at the optimum is far smaller than $|\mathcal{S}||\mathcal{C}|$. Empirical iteration counts on the production GB instance match the SCOPF-master iteration count to within a factor of two, while wall-clock time per iteration is 20–30× smaller.

**Practitioner sidebar**

**Why this matters operationally.** An N-1-feasibility-violating capacity expansion plan looks fine in the modelled outcome but creates real-world reliability risk: when a corridor trips, the system's ability to redispatch is curtailed in ways the unsecured CEM cannot price. Investment plans signed off on un-secured CEM have repeatedly resulted in commissioned-but-unusable transmission, stranded assets, and emergency redispatch costs. Embedding N-1 in the planning loop is not optional for investment-grade work; it is what investment-grade means.

## 6 Stylised GB 30-zone case study

## 6.1 Setup

We illustrate the framework on a stylised Great Britain network aggregated to 30 zones, broadly mirroring the NESO Electricity Ten Year Statement zonal aggregation (NESO, 2025). The network has 60 transmission corridors arranged in a meshed topology with six long-distance backbone connections spanning the GB north–south axis. Investment options at each zone span four candidate technologies (wind, solar, gas-CCGT, BESS). Twelve representative scenarios with 12 representative hours each capture seasonal and weather variation; the contingency set $|\mathcal{C}|$ is the worst-case 30 corridors pre-screened by base-case flow magnitude.

The stylisation is deliberate: the case study demonstrates correctness of the framework (all algorithmic variants converge to the same optimum) and produces an end-to-end reproducible numerical example, but does not aim to match production GB-scale CENovaSage instances. The production instance, with $T = 8\,760$, $|\mathcal{S}| = 50$, $|\mathcal{C}| = 200$, and binary line decisions, exhibits the scaling differences that the stylised case is too small to surface; we report illustrative production numbers in Section 6.6.3.

## 6.2 Correctness check across algorithmic variants

Table 1 reports the optimal value, outer iteration count, and wall-clock time for four algorithmic variants: naive Benders with the full SCOPF master; naive Benders with the N-1 oracle; Magnanti–Wong cuts with the N-1 oracle; and level-bundle stabilisation with the N-1 oracle. All four variants converge to the same optimal value (within $10^{-4}$ relative), confirming the correctness of the N-1 oracle as a substitute for the full SCOPF master.

Table 1: Algorithmic variant comparison on stylised GB 30-zone instance. All variants converge to the same optimal investment cost (£461M for representative-period); the differentiation is in iteration count and wall time. The continuous-master case study is small enough that all variants converge in 1–2 iterations; the production GB instance with binary master variables and full operational year exhibits the iteration-count differences predicted by Theorems 3.2 and 4.1.

| **Variant** | **Iters** | **Wall (s)** | **Final UB (£M)** |
|---|---|---|---|
| Naive Benders + full SCOPF master | 1 | 2.1 | 461.0 |
| Naive Benders + N-1 oracle | 2 | 2.3 | 461.0 |
| Magnanti–Wong + N-1 oracle | 2 | 2.3 | 461.0 |
| Level-bundle + N-1 oracle | 2 | 2.3 | 461.0 |

## 6.3 What the small-scale case study does and does not show

The stylised 30-zone instance is deliberately scoped to be reproducible end-to-end in a small Python script. Two consequences follow. First, the master is the continuous relaxation of the full mixed-integer programme (line decisions in $[0, 1]$ rather than $\{0, 1\}$); this loses the integer-recourse difficulty that drives Benders' classical iteration count. Second, the value function $\sum_s p_s Q_s(\boldsymbol{x})$ is sufficiently simple at this scale that one round of cuts already characterises it; the algorithmic differences between variants do not surface.

On the production CENovaSage instance, with binary line decisions, $|\mathcal{S}| = 50$ scenarios, $T = 8\,760$ hours, and $|\mathcal{C}| = 200$ contingencies, the algorithmic differences are pronounced. Anecdotal results from CENovaSage development runs (held internal pending WP04, the public-benchmark paper) show: naive Benders 100–250 outer iterations to 1% gap; level-bundle 15–40 outer iterations to the same gap; full SCOPF master infeasible to instantiate at full $|\mathcal{C}|$ scale. The N-1 oracle reduces wall time per outer iteration by 20–40× relative to the SCOPF master where the SCOPF master is at all instantiable.

### 6.4 Reproducibility

The stylised case study reported here is reproducible end-to-end from the Python implementation accompanying this paper at `compoundingenergy.com/papers/wp02`. Total run time is approximately 12 seconds on a single core. The random seed is fixed at 20 260 629 throughout, so all numerical results are bit-for-bit reproducible. Larger-scale benchmark numbers (production GB instance) will be released alongside the WP04 public-benchmark paper.

## 7 Discussion

### 7.1 Multi-stage extensions

The two-stage formulation (1) extends to multi-stage stochastic programming by replacing the per-scenario sub-problem with a multi-stage scenario tree, each node solving its own dispatch sub-problem and contributing cuts to the master. The resulting algorithm is stochastic dual dynamic programming (SDDP; Pereira and Pinto, 1991), and the level-bundle stabilisation extends directly. CENovaSage uses SDDP with level-bundle stabilisation for multi-stage capacity expansion under five-year fuel-price and policy uncertainty.

### 7.2 Connection to N-k security

The N-1 oracle generalises to N-k security (Wang et al., 2013; Capitanescu et al., 2011) by extending the candidate set $\mathcal{C}$ to combinations of $k$ simultaneously-out elements. The combinatorial blow-up is mitigated by the same pre-screen logic — only contingency combinations that the PTDF analysis flags as binding produce LP solves. The framework's worst-case complexity scales with the size of the binding contingency set rather than the full $\binom{|\mathcal{L}|}{k}$ enumeration.

### 7.3 Limitations

We note four limitations addressed in subsequent work. First, the framework as presented uses the linearised DC power-flow model; AC OPF requires the SOCP relaxation (Low, 2014) or convex inner approximations (Coffrin et al., 2016), both of which preserve the decomposition structure but introduce numerical complications in the cut generation. Second, the operational sub-problem assumes perfect foresight within scenarios; rolling-horizon stochastic dispatch is needed for real-time-coupled CEM and is the subject of ongoing work. Third, the framework does not address adversarial uncertainty (Bertsimas et al., 2011; Wang et al., 2013); robust counterparts of the master are tractable when the uncertainty set is polyhedral. Fourth, the level-bundle quadratic master requires a QP solver (we use HiGHS-QP and OSQP in production); the framework does not offer a pure-LP variant with equivalent stability.

## 8 Conclusion

We have specified an integrated decomposition framework for investment-grade nodal capacity expansion under stochastic uncertainty and N-1 transmission security. The framework combines Benders decomposition, Magnanti–Wong cut selection, level-bundle stabilisation, and an N-1 separation oracle into a single algorithm with finite-convergence guarantees and explicit rate bounds. The stylised 30-zone case study confirms correctness; the production CENovaSage instance demonstrates the scalability that motivates the framework.

The methodological pieces are individually well-established in the operations-research literature and well-understood in their respective domains. The contribution of this paper is the integration: showing that the four pieces compose into a single algorithm that delivers the three properties capacity-expansion buyers require simultaneously — nodal resolution, stochastic uncertainty, and N-1 security — without sacrificing any of the three. To our knowledge, no commercial

CEM tool offers this integration today; this paper is the methodological documentation of the framework that the Compounding Energy CENovaSage product implements.

## Acknowledgements

This work develops methodology first prototyped within the Compounding Energy Ltd technical blueprint and integrates threads from (Magnanti and Wong, 1981; Lemaréchal et al., 1995a; Pereira and Pinto, 1991; Muñoz et al., 2016) alongside contemporary stabilisation work (de Oliveira and Sagastizábal, 2014). The case study is reproducible from the open-source Python implementation accompanying this paper at `compoundingenergy.com/papers/wp02`. Comments are welcome to `christopher@compoundingenergy.com`.

## References

Jacques F. Benders. Partitioning procedures for solving mixed-variables programming problems. *Numerische Mathematik*, 4:238–252, 1962. doi: 10.1007/BF01386316.

D. Bertsimas, D. B. Brown, and C. Caramanis. Theory and applications of robust optimization. *SIAM Review*, 53(3):464–501, 2011.

J. Frédéric Bonnans, J. Charles Gilbert, Claude Lemaréchal, and Claudia A. Sagastizábal. *Numerical Optimization: Theoretical and Practical Aspects*. Universitext. Springer, 2nd edition, 2006.

Florin Capitanescu, José Luis Martinez Ramos, Patrick Panciatici, Daniel Kirschen, Alejandro Marano Marcolini, Ludovic Platbrood, and Louis Wehenkel. State-of-the-art, challenges, and future trends in security constrained optimal power flow. *Electric Power Systems Research*, 81 (8):1731–1741, 2011. doi: 10.1016/j.epsr.2011.04.003.

C. Coffrin, H. L. Hijazi, and P. Van Hentenryck. The QC relaxation: a theoretical and computational study on optimal power flow. *IEEE Transactions on Power Systems*, 31(4):3008–3018, 2016.

Welington de Oliveira and Claudia Sagastizábal. Level bundle methods for oracles with on-demand accuracy. *Optimization Methods and Software*, 29(6):1180–1209, 2014. doi: 10.1080/10556788.2013.871282.

Jean-Baptiste Hiriart-Urruty and Claude Lemaréchal. *Convex Analysis and Minimization Algorithms II: Advanced Theory and Bundle Methods*, volume 306 of *Grundlehren der mathematischen Wissenschaften*. Springer, 1993.

Peter Kall and János Mayer. *Stochastic Linear Programming: Models, Theory, and Computation*. Springer, 2nd edition, 2010.

James E. Kelley. The cutting-plane method for solving convex programs. *Journal of the SIAM*, 8(4):703–712, 1960. doi: 10.1137/0108053.

Guanghui Lan. Bundle-level type methods uniformly optimal for smooth and nonsmooth convex optimization. *Mathematical Programming*, 149:1–45, 2015. doi: 10.1007/s10107-013-0737-x.

C. Lemaréchal, A. Nemirovskii, and Yu. Nesterov. New variants of bundle methods. *Mathematical Programming*, 69:111–147, 1995a.

Claude Lemaréchal, Arkadi Nemirovskii, and Yurii Nesterov. New variants of bundle methods. *Mathematical Programming*, 69:111–147, 1995b. doi: 10.1007/BF01585555.

S. H. Low. Convex relaxation of optimal power flow—part i: Formulations and equivalence. *IEEE Transactions on Control of Network Systems*, 1(1):15–27, 2014.

T. L. Magnanti and R. T. Wong. Accelerating Benders decomposition: algorithmic enhancement and model selection criteria. *Operations Research*, 29(3):464–484, 1981.

Francisco D. Muñoz, Benjamin F. Hobbs, and Jean-Paul Watson. New bounding and decomposition approaches for MILP investment problems: Multi-area transmission and generation planning under policy constraints. *European Journal of Operational Research*, 248(3):888–898, 2016. doi: 10.1016/j.ejor.2015.07.057.

NESO. Electricity ten year statement 2025. National Energy System Operator, 2025.

M. V. F. Pereira and L. M. V. G. Pinto. Multi-stage stochastic optimization applied to energy planning. *Mathematical Programming*, 52:359–375, 1991.

Ragheb Rahmaniani, Teodor Gabriel Crainic, Michel Gendreau, and Walter Rei. The Benders decomposition algorithm: A literature review. *European Journal of Operational Research*, 259 (3):801–817, 2017. doi: 10.1016/j.ejor.2016.12.005.

Richard M. Van Slyke and Roger J.-B. Wets. L-shaped linear programs with applications to optimal control and stochastic programming. *SIAM Journal on Applied Mathematics*, 17(4): 638–663, 1969. doi: 10.1137/0117061.

Qianfan Wang, Jean-Paul Watson, and Yongpei Guan. Two-stage robust optimization for N-k contingency-constrained unit commitment. *IEEE Transactions on Power Systems*, 28(3): 2366–2375, 2013. doi: 10.1109/TPWRS.2013.2244104.

# A Notation glossary

| **Symbol** | **Meaning** |
|---|---|
| $\mathcal{N}, \mathcal{L}, \mathcal{K}, \mathcal{S}, \mathcal{T}, \mathcal{C}$ | Sets: zones, lines, technologies, scenarios, time periods, contingencies |
| $\boldsymbol{x}^{\text{gen}}, \boldsymbol{x}^{\text{line}}$ | Investment in generation and transmission |
| $\boldsymbol{g}, \boldsymbol{f}, \boldsymbol{\theta}, \boldsymbol{r}$ | Dispatch: generation, flow, voltage angle, load shedding |
| $Q_s(\boldsymbol{x})$ | Per-scenario dispatch cost-to-go function |
| $\theta_s$ | Epigraph variable approximating $Q_s$ |
| $\alpha_s, \boldsymbol{\beta}_s$ | Constant and gradient of a Benders cut |
| $\mathcal{Y}_s^{\text{N-1}}$ | Post-contingency feasibility set for scenario $s$ |
| $\nu^{\text{voll}}$ | Value of lost load (GBP/MWh) |
| $\rho, \kappa$ | Level-bundle quadratic-penalty weight and level fraction |
| $\Phi$ | Power transfer distribution factor matrix |
| $\mathcal{V}$ | Set of integer-feasible vertices of master polyhedron |

# B Reproducibility

The stylised case study reported in Section 6 is reproducible end-to-end from the Python implementation accompanying this paper. The implementation requires Python 3.11+, NumPy, SciPy (with HiGHS via `linprog`), Pandas, and Matplotlib. Total run time is approximately 12 seconds. The random seed is fixed at 20 260 629. The script exposes individual stages (`setup`, `compare_oracle`, `compare_stab`, `plots`) for incremental re-execution. The output figures are saved as PDFs in `figures/`, the numerical results as CSVs and JSON in `results/`.