OXFORD JOURNAL
OF STUDENT
SCHOLARSHIP

Reliability-Constrained Capacity Planning for School Organic-Waste Networks Under Uncertainty

Christopher Dylan Suruadji
christophersuruadji@gmail.com

ABSTRACT

School organic-waste networks face uncertain collection and treatment disruptions. This study compares 231 capacity configurations in a hypothetical eight-school network using 50,000 simulated training weeks and an independent 100,000-week Monte Carlo evaluation. Binomial screening limits failure probability, while tail and conditional-loss measures distinguish infrequent shortfalls from severe failed weeks.

The baseline-selected hybrid places 20% of rated capacity at a shared facility and installs 25% reserve. Its held-out failure frequency is 2.35%, and it wins all 20 independent training replications. At equal capacity, this split minimizes $CVaR_{0.95}$ at 25% reserve, but conditional mean shortfall increases with the shared fraction. The original 100-kg $CVaR_{0.95}$ screen permits large losses during failures; adding a conditional-mean limit selects a different hybrid, and combining it with a worst-1% limit increases reserve.

The baseline hybrid exceeds its original severity limit in three of five stress tests. Its small economic advantage and sensitivity to service criteria and facility charges do not justify prescribing an exact split generally. The contribution is a reproducible comparison of reserve placement, failure severity, and school-level service; operational use requires measured loads, technology-specific availability, costs, and backup capacity.

Keywords: school organic waste; capacity pooling; stochastic planning; reliability; conditional value-at-risk; resource recovery

1. INTRODUCTION

A school network can have enough treatment capacity for average collection and still leave waste untreated. Spare capacity at one school may be unavailable to another. A shared facility can accept excess material from several schools, but its disruption affects all schools that depend on it. Capacity planning must therefore consider where reserves are placed and how much untreated material the network can tolerate.

School canteens concentrate collection at identifiable sites and operating times, making reserve sharing a relevant planning question. Green Soldier provides the motivating school waste-management context; the eight schools and simulated facilities here are hypothetical, not an evaluation of that initiative. A shortfall requires a destination outside the modeled network, so backup mass matters alongside failure frequency. School food-waste audits also show that meal provision and measurement boundaries differ substantially across settings (Bilqis et al., 2023; Wilkie et al., 2015).

The study asks how dividing equal total rated capacity between local and shared facilities changes the probability and severity of untreated collection under correlated loads and facility disruptions. The analysis also examines whether designs meeting the same failure-frequency criterion can differ substantially in the amount of untreated waste during failed weeks. The numerical experiments use disclosed inputs, held-out Monte Carlo scenarios, and stress tests.

Two comparisons are kept distinct. Equal-capacity experiments hold reserve and sampled weeks constant while varying the shared fraction, isolating capacity placement within the model. Selected-design comparisons evaluate configurations chosen by specified service screens and the planning objective; their capacities may differ. Environmental indicators describe the consequences of assumed coefficients and are not a life-cycle assessment.

2. RELATED RESEARCH AND SCOPE OF THE CONTRIBUTION

2.1 Established approaches to waste-system design

Previous studies have examined hybrid waste infrastructure and planning under uncertainty. Xiong et al. (2016) used a two-stage stochastic approach to assess decentralized waste-to-energy systems, including the joint probability that installed units meet financial targets. Kuznetsova et al. (2019) developed a method for comparing combined centralized and decentralized waste-to-energy systems. Both studies provide a basis for examining capacity placement under uncertain operating conditions.

Eppen (1979) formalized how pooling uncertain demands can reduce expected inventory costs, with the benefit depending on correlation. Restricted local treatment and accessible shared reserve provide a related capacity-pooling mechanism, although waste treatment differs from the inventory setting. Schmitt et al. (2015) distinguish demand pooling from disruption diversification. Snyder and Daskin (2005) explicitly incorporate facility failures and backup service into location planning. Here locations are fixed; partial throughput losses expose the tension between accessible reserve and concentrated disruption risk.

Gambella et al. (2019) studied tactical waste-flow allocation with a two-stage, multi-period stochastic model and historical data from an Italian waste operator. Their separation of prior decisions from subsequent flow adjustments informs the capacity-and-allocation structure used here. Saif et al. (2019) selected technology and capacity in an organic municipal-waste network before optimizing transport and power production under uncertainty. Capacity selection followed by uncertain waste allocation is therefore an established approach.

The present model fixes facility locations and treatment coefficients, represents one school week, and has no interweek inventory. It uses synthetic inputs to compare reserve placement and untreated mass; it does not reproduce the operator or regional case studies.

Rathore et al. (2022) combined bin allocation, composting-plant location, and collection routing to compare centralized and decentralized organic-waste systems in Bilaspur, India. Xie et al. (2024) studied recycling-center locations and capacities under uncertain construction-waste supply and recycled-material demand. The latter reported an improvement over a model based on mean quantities. These studies motivate comparisons under uncertain loads, although their numerical parameters are not used in the school network.

2.2 Risk, environmental accounting, and remaining design questions

Rockafellar and Uryasev (2000) developed the optimization representation of conditional value-at-risk (CVaR), which measures loss in the upper tail of a distribution. Lotfi et al. (2022) incorporated CVaR and robustness into medical-waste network design. Zhao et al. (2019), in a preprint on multi-facility capacity expansion, applied risk aversion to a waste-to-energy setting. The present study uses these established risk concepts to measure untreated mass. Sample-based chance-constrained optimization is developed by Luedtke and Ahmed (2008) and Pagnoncelli et al. (2009). The present method instead exhaustively evaluates a fixed 231-design grid and applies simultaneous binomial upper bounds; it does not solve a continuous chance-constrained problem or claim a new algorithm.

Environmental results depend on the treatment process and system boundary. Qadar et al. (2024) compared food-waste management alternatives in Semarang using life-cycle assessment. Their treatment-specific estimates cannot be assigned directly to the generic facilities modeled here. Malashin et al. (2026) considered economic, environmental, and resilience criteria for a regional biogas supply chain. Thus, joint assessment of cost, environmental effects, and service is already part of waste-network research.

This study isolates a capacity-placement question within an established waste-network planning framework. Facility locations and treatment coefficients are fixed, allowing 231 capacity configurations to be compared under common simulated loads and disruptions. The experiment examines how frequency and severity criteria select reserve levels and how a cost-minimizing allocation rule distributes shortfalls among schools. Its contribution is the controlled comparison and its planning implications, rather than a new optimization algorithm.

3. MATERIALS AND METHODS

3.1 System boundary and input provenance

The unit of analysis is one five-day school week. Eight hypothetical schools may process material locally or send residual material to one shared facility. Local units cannot accept waste from other schools. The shared location and school-to-facility distances are fixed. The modeled stream consists only of separately

collected organic material after rejection of unsuitable material. Material not collected or rejected before entry is outside the capacity-service denominator. Untreated material means accepted collection not processed by this network during that week; the model neither assumes that it remains physically stored nor credits it with recovery.

School populations, generation rates, collection and rejection fractions, distances, disruption probabilities, costs, emissions factors, and product yields are synthetic inputs disclosed for replication and sensitivity analysis. None is estimated from a school survey or presented as a current market price. Published research informs the model structure and comparisons, but does not calibrate these values.

Eight schools provide a transparent benchmark with heterogeneous loads and distances, enough collection points to distinguish restricted local reserve from pooled reserve and to inspect individual service. Eight is not a representative sample size or an estimated optimal network size. The generation distribution enforces positive loads; beta fractions remain bounded; the common factor introduces shared variation. Their numerical dispersion parameters define a baseline rather than a fitted distribution. The parameter-rationale table in the supplement records each input group, its modeling role, empirical status, and sensitivity treatment. Published school measurements provide context rather than calibration. Bilqis et al. (2023) report 33.6 kg/day of total waste for 265 boarding-school residents, with 41.8% organic material, implying about 0.053 kg organic waste/person/day. Wilkie et al. (2015) report school food-waste rates of 0.025–0.065 kg/student/day, while Antón-Peset et al. (2021) report mean plate waste of about 0.140 kg/student/day before an intervention. The assumed 0.105–0.119 kg/person/day lies within this broad contextual span but exceeds the first two studies. Residential status, meals, population denominators, and stream definitions prevent treating these comparisons as a fitted Indonesian day-school rate.

Table 1. Synthetic school inputs and implied mean processable collection.

School	People N	g (kg/person/day)	p	c	d (km)	μ (kg/week)
S1	430	0.108	0.68	0.080	6	145.264
S2	510	0.112	0.72	0.090	8	187.125
S3	590	0.105	0.70	0.100	9	195.142
S4	660	0.119	0.74	0.075	11	268.803
S5	720	0.110	0.69	0.095	13	247.282
S6	810	0.116	0.73	0.085	10	313.803
S7	920	0.113	0.71	0.105	7	330.307
S8	1,050	0.118	0.75	0.090	12	422.809
Total	5,690	—	—	—	—	2,110.536

Note. N is contributing population; g is mean organic generation (kg/person/day); p is mean collection fraction; c is mean rejected fraction; d is distance to the shared facility (km); μ is mean accepted processable collection (kg/week). All inputs are synthetic. Totals use unrounded school values.

For school i, let N_i be the contributing population, $\bar{g}_i$ the mean organic generation rate, $\bar{p}_i$ the mean collection fraction, and $\bar{c}_i$ the mean rejected fraction. Mean processable collection is calculated as

$$\mu_i = 5N_i\bar{g}_i\bar{p}_i(1 - \bar{c}_i), \quad \mu = \sum_{i=1}^{8} \mu_i = 2{,}110.536 \text{ kg/week}. \quad (1)$$

The product of means equals expected collection under the assumed independence between generation, collection, and rejection groups. This would not hold automatically if these groups were correlated in field data. A smaller processable load can result from poor participation or higher rejection, so lower shortfall alone must not be interpreted as better waste prevention or collection performance.

3.2 First-stage capacity decisions

Let h denote the fraction of rated capacity installed at the shared facility and s the reserve above mean processable collection. Local capacity L_i and shared capacity H are determined before weekly conditions are known:

$$L_i = (1 - h)(1 + s)\mu_i, \quad H = h(1 + s)\mu. \quad (2)$$

The grid uses h = 0, 0.1, …, 1.0 and s = 0, 0.05, …, 1.00, giving 231 designs. A decentralized design has h = 0; a centralized design has h = 1; intermediate values are hybrid. For every h at fixed s, total rated

capacity equals (1 + s)μ. Local capacity is proportional to school mean collection. The method therefore identifies a best candidate in a restricted grid; it does not optimize facility locations, individual local sizes, vehicle routes, or all possible network structures.

Every hybrid configuration installs eight local units and one shared facility, regardless of the fraction of capacity assigned to each. Expenditure includes the fixed charges for all nine facilities. At full centralization, no local units are installed; at full decentralization, no shared facility is installed. Consequently, fixed facility charges change at the two architecture endpoints. Thus, even a small shared unit pays the full IDR 92,000 weekly charge. The discontinuities can favor endpoints or particular splits; Section 3.8 tests a partially capacity-dependent alternative.

3.3 Weekly loads and dependence

Weekly generation is lognormal with log standard deviation 0.14. Collection and rejection fractions are beta-distributed with their school means and concentration parameters 75 and 120, respectively. For each variable group $k \in \{g, p, c\}$, school i and week ω have latent score

$$z_{ki\omega} = aZ_{k0\omega} + \sqrt{(1 - a^2)}\,\varepsilon_{ki\omega}, \quad a = 0.55. \quad (3)$$

All common factors Z and individual factors ε are independent standard normal draws; different variable groups are independent. Within each group, two schools' latent scores have correlation $a^2 = 0.3025$. This is the latent Gaussian correlation, not a claim that transformed waste quantities have exactly that Pearson correlation. Common factors represent simultaneous influences across schools without making their loads identical.

$$g_{i\omega} = \bar{g}_i \exp(0.14z_{gi\omega} - 0.14^2/2), \quad (4a)$$

$$p_{i\omega} = B^{-1}(\Phi(z_{pi\omega}); 75\bar{p}_i, 75(1 - \bar{p}_i)), \quad (4b)$$

$$c_{i\omega} = B^{-1}(\Phi(z_{ci\omega}); 120\bar{c}_i, 120(1 - \bar{c}_i)), \quad (4c)$$

$$Q_{i\omega} = 5N_i g_{i\omega} p_{i\omega}(1 - c_{i\omega}). \quad (4d)$$

Here Φ is the standard normal cumulative distribution function, and B^{-1} is the inverse beta cumulative distribution function with the two stated shape parameters. The lognormal correction preserves mean generation, and the beta transformations keep fractions in [0, 1]. Weeks are independently generated; persistent seasonal or serial dependence is not included.

3.4 Facility availability and operational allocation

Each local unit independently has a 0.025 probability of a disrupted week, and the shared facility has probability 0.035. A disrupted facility retains 55% of rated weekly throughput; otherwise availability equals one. These assumptions represent infrequent partial interruptions. Neither the one-percentage-point

probability difference nor 55% residual availability has empirical support in this benchmark; both are examined in additional sensitivity analyses. Availability variables $A_i\omega$ and $A_0\omega$ are independent of load and of one another in the baseline. Rated throughput is aggregate service capacity, not completion of biological treatment within one week. Technology-specific residence time, space, moisture, and inventory requirements must be established before interpreting it as implementable physical capacity.

For each realized week, the network uses available local capacity first, then sends excess material to the shared facility in ascending distance order. Let $x_i\omega$ denote local treatment, $y_i\omega$ shared treatment, and $u_i\omega$ untreated collection. The allocation obeys

$$x_{i\omega} + y_{i\omega} + u_{i\omega} = Q_{i\omega}; \quad 0 \le x_{i\omega} \le A_{i\omega}L_i; \quad (5a)$$

$$\sum_i y_{i\omega} \le A_{0\omega}H; \quad y_{i\omega} \ge 0, \quad u_{i\omega} \ge 0. \quad (5b)$$

With the baseline coefficients in Table 2, local processing costs IDR 1,380/kg. Shared processing plus transport costs IDR 1,410–1,900/kg. Both cost less than the IDR 5,000/kg shortfall penalty. Using local capacity first and allocating shared capacity in ascending distance order therefore minimizes variable cost plus shortage penalty and maximizes treated mass. Distance priority is an economic benchmark with no per-school equity constraint, not a normative service recommendation. An additional comparison allocates scarce shared throughput in proportion to residual school loads after local treatment; this preserves total treated mass while changing school-level losses and transport expenditure.

Define $[v]^+ = \max(v, 0)$. Total untreated mass follows directly from the residual load presented to the shared facility:

$$U_\omega = \sum_i u_{i\omega} = [\sum_i [Q_{i\omega} - A_{i\omega}L_i]^+ - A_{0\omega}H]^+. \quad (6)$$

Equation (6) includes both endpoints: $H = 0$ gives the sum of local excesses, and $L_i = 0$ pools all loads at the shared facility. Shared capacity accepts residual loads from any school; unused local capacity remains restricted to its own school.

3.5 Economic and environmental accounting

Table 2. Assumed cost, emissions, and recovery coefficients.

Quantity	Local	Shared / transport
Capacity charge per week per kg/week rated capacity	IDR 215	IDR 125
Fixed charge per installed facility per week	IDR 23,000	IDR 92,000
Treatment cost per kg input	IDR 1,380	IDR 990
Transport cost per kg-km	—	IDR 70
Treatment emissions (kg CO_2e/kg input)	0.095	0.075
Transport emissions (kg CO_2e/kg-km)	—	0.0045
Avoided-disposal factor (kg CO_2e/kg treated)	0.400	0.400
Recovered wet product (kg/kg input)	0.250	0.230
Shortfall planning penalty (IDR/kg)	5,000 across the network	Not counted as expenditure

Weekly expenditure K includes charges for installed capacity, fixed operation, treatment, and transport. Let n^L be the number of installed local units and I^H indicate whether the shared facility is installed. Then

$$K_\omega = 215\sum_i L_i + 125H + 23{,}000n^L + 92{,}000I^H + 1{,}380\sum_i x_{i\omega} + 990\sum_i y_{i\omega} + 70\sum_i d_i y_{i\omega}. \quad (7)$$

Capacity charges are synthetic weekly equivalents, not supplier quotations or a discounted investment appraisal. Lower shared capacity and treatment coefficients represent possible scale economies, while the distance-dependent transport charge creates a countervailing cost. Their numerical values define a testable trade-off rather than an estimate of actual savings. The planning objective J is mean expenditure plus 5,000 times mean untreated kilograms. The penalty is an analyst preference for avoiding shortfall, not an observed disposal invoice. The system boundary excludes upstream bin collection, land purchase, financing, product revenue, and additional emergency disposal costs. Indonesian rupiah (IDR) is the modeling currency. Relative differences are more informative than the absolute amounts; no exchange-rate conversion or market-price claim is intended.

$$J = E[K_\omega] + 5{,}000E[U_\omega]. \quad (8)$$

Illustrative net avoided emissions Eω credit only treated mass relative to a common disposal counterfactual. Untreated material receives no avoided-disposal credit. Recovered wet product Rω is a yield indicator, distinct from input mass treated:

$$E_{\omega} = 0.305\sum_i x_{i\omega} + 0.325\sum_i y_{i\omega} - 0.0045\sum_i d_i y_{i\omega}, \quad (9a)$$

$$R_{\omega} = 0.25\sum_i x_{i\omega} + 0.23\sum_i y_{i\omega}. \quad (9b)$$

The model does not distinguish compost, digestate, or another biological product and does not represent process kinetics, infrastructure emissions, methane leakage, moisture conversion, product quality, or substitution credits. Equations (9a–b) are therefore scenario accounting, not a full life-cycle assessment or evidence of saleable output. Their purpose is to make the consequences of assumed coefficients visible. The baseline transport coefficient multiplies the tabulated school-to-hub distance; empty-return trips and vehicle loading are not separately modeled.

3.6 Reliability screening, severity, and held-out Monte Carlo evaluation

A failure occurs if any accepted processable collection is untreated: $U\omega > 10^{-8}$ kg, with the small tolerance preventing floating-point noise. This strict service definition treats 1 kg and hundreds of kilograms as equally frequent failure events; severity is reported separately. The tolerance is numerical, not an operational allowance. The service target is P(failure) < 0.10. It limits the frequency of any shortfall, not the fraction of mass untreated. All 231 designs are evaluated on the same 50,000 training weeks using NumPy PCG64 seed 240926. Using common scenarios permits direct paired comparisons without giving one design a more favorable sequence of weeks.

A design passes training screening only if a one-sided exact binomial upper bound for its failure probability is below 0.10. For k failures in n weeks, the bound is $B^{-1}(1 - \alpha; k + 1, n - k)$, or one when k = n (Clopper & Pearson, 1934). Training uses $\alpha = 0.05/231$, providing simultaneous coverage of at least 95% over the grid by the Bonferroni inequality. This controls simulation uncertainty under the assumed distribution; its relevance to an actual school network depends on how well that distribution represents measured conditions.

Among screened candidates, the smallest sample J is selected overall and within each architecture family. A second selection adds an illustrative $CVaR_{0.95}(U) \leq 100$ kg condition. This condition constrains the mean loss in the worst 5% of weeks, including zero-loss weeks when failures occur less often than 5%. It is not the conditional mean given any failure. The 100-kg limit is an analyst-chosen service preference, approximately 4.74% of mean weekly collection, and is not a regulatory or empirically established threshold. The 95% tail averages approximately the worst one week in twenty, a useful measure of recurring backup demand, but does not cap any particular failed week. CVaR follows the standard representation

$$CVaR_{0.95}(U) = \min_t \{t + 20E[(U - t)^+]\}. \quad (10)$$

For equally weighted samples whose size is divisible by 20, empirical $CVaR_{0.95}$ is the mean of the largest 5% of U values. Because U is nonnegative, if $P(U > 0) \leq 0.05$ then $VaR_{0.95}$ is zero and $CVaR_{0.95}(U) = E[U]/0.05 = 20E[U]$. The analogous identity holds for an empirical sample with at most 5% positive losses. In this regime the 100-kg screen is equivalent to mean weekly shortfall ≤ 5 kg. It neither limits individual failures to 100 kg nor bounds the mean loss conditional on failure. We therefore report conditional severity alongside $CVaR_{0.95}$ and add $CVaR_{0.99}$, the mean of the largest 1% of held-out losses, alongside conditional mean $E[U \mid U > 0]$. $CVaR_{0.99}$ also dilutes losses with zeros when failure probability is below 1%; the conditional mean avoids that dilution but still does not bound individual losses. Section 3.8 adds explicit selection sensitivities using both measures.

The severity screen uses a point estimate. Supplementary uncertainty assessment resamples the 100,000 held-out weekly losses with replacement in 2,000 replicates per fixed candidate, recomputes both CVaRs, and reports 2.5th–97.5th percentile intervals. PCG64 seeds 181026–181031 follow the candidate order in review_extensions.py. These are pointwise bootstrap diagnostics under the synthetic distribution, not simultaneous bounds, finite-sample feasibility certification, or uncertainty about the assumed inputs.

Six distinct candidates are nominated before held-out evaluation: the three architecture winners, the centralized severity-screened winner, the objective runner-up, and a preselected decentralized 80%-reserve benchmark. Duplicate nominations from the severity screen are counted once. These fixed designs are evaluated on 100,000 independent weeks with seed 121026 and $\alpha = 0.05/6$ for simultaneous one-sided binomial bounds. The evaluation does not reselect winners. Mean-expenditure and mean-shortfall standard errors use independent weeks; the runner-up comparison uses paired differences in $K + 5{,}000U$ and a normal-approximation 95% interval. Independence refers to new random draws from the same model: this tests Monte Carlo generalization within the assumed distribution and is not validation against observed school operations.

3.7 Stress tests, coefficient sensitivity, and computational verification

Five stress tests each use 50,000 weeks with seed 131026 and retain the selected capacities. They remove cross-school dependence ($a = 0$); increase it ($a = 0.8$); increase load deviations by 50% through $\max[0, \mu_i + 1.5(Q_i\omega - \mu_i)]$; remove all facility disruptions; or add a common disruption in 2% of weeks. The common disruption sets all facility availability to 0.55 in affected weeks. These are hypothetical distribution shifts, not estimated future frequencies.

Additional availability sensitivities use the same 50,000 loads and underlying availability uniforms with seed 131026. Local and shared disruption probabilities are separately varied to 1%, 5%, and 10%, holding the other probability at baseline. Residual availability for disrupted facilities is separately varied to 25%, 40%, 70%, and 85%. A joint adverse case uses 10% disruption probabilities and 25% residual availability. Each is compared with a matched baseline on those same draws, without reselecting capacities. The proportional-allocation comparison uses the original 100,000 held-out weeks. These additions diagnose assumption sensitivity; they are not a robust optimization or a new confirmatory selection.

Additional training-sample analyses vary the shortfall penalty to IDR 2,500 or 10,000/kg, halve or double the capacity-charge coefficients, and multiply transport cost by 1.25 or 2. The remaining coefficients and the dispatch rule are held fixed. These penalty values still exceed every baseline unit processing-and-transport cost. An exploratory environmental screen requires mean illustrative avoided emissions of at least 600, 620, or 640 kg CO_2e/week. These additional designs are sensitivity outputs and are not represented as independently validated selections.

Every design evaluation checks mass balance, nonnegative allocations, and shared capacity. Forty allocations match separately formulated linear programs to numerical precision. The archived implementation reproduces the original training grid, held-out results, and school-level allocation calculations. Complete precision and diagnostic tolerances remain in the supplement.

3.8 Additional selection and uncertainty analyses

In addition to the baseline selection criteria, the analysis evaluates two additional illustrative service requirements: $CVaR_{0.99} \leq 200$ kg, representing average backup demand in the worst 1% of weeks (about 9.5% of analytical mean collection), and $E[U \mid U > 0] \leq 100$ kg, representing average backup demand when any failure occurs (about 4.7%). These are distinct preferences, not regulatory limits or maximum losses. Threshold sweeps use 100–400 kg for $CVaR_{0.99}$ and 50–250 kg for conditional mean. All 231 designs are screened on the original training sample. Seven unique family/overall nominations are then fixed and evaluated on a new 100,000-week sample (seed 221026), with binomial upper bounds using 0.05/7. Severity remains point-estimate screening.

Equal-capacity comparisons evaluate every h at each s on the original 100,000 held-out weeks. Table 7 details s = 0.25; the supplement includes the full grid and five reserve slices. Separately, 20 independent training samples of 50,000 weeks (PCG64 seeds 221100–221119) repeat the complete original selection, with and without the $CVaR_{0.95}$ screen. This assesses simulation selection stability under fixed assumptions.

A fixed-charge sensitivity replaces the shared charge by $F^H(h) = 46{,}000(1 + h)$ IDR/week for $h > 0$, and zero at $h = 0$. Half of the full shared charge is overhead and half scales with h; local charges remain unchanged. This synthetic specification retains a discontinuity at zero and is not a fitted scale-cost function. An operational-tolerance sensitivity defines failure as $U\omega > 0.01\sum_i Q_i\omega$, repeats the binomial screen, and leaves loss penalties and severity measures unchanged. The 1% allowance is illustrative.

For all stress and availability scenarios, two-sided exact binomial 95% intervals describe failure frequency. Severity intervals use 2,000 nonparametric bootstrap replicates of whole simulated weeks, with percentile endpoints at 2.5% and 97.5%. Identical resampling weights preserve paired scenario-minus-matched-baseline differences. Seeds 222001–222054 and all intervals are recorded in stress-availability-uncertainty.csv. These intervals are pointwise Monte Carlo diagnostics, not simultaneous feasibility guarantees or uncertainty about actual school parameters.

4. RESULTS

4.1 Demand variation and the value of accessible reserve

Mean total collection in the training sample is 2,112.1 kg/week, with standard deviation 214.1 kg/week and 5th and 95th percentiles of 1,777.5 and 2,479.6 kg/week. The simulated mean is close to the analytical mean of 2,110.5 kg/week. At zero reserve, decentralized, reference-hybrid (h = 0.4), and centralized designs fail in 94.69%, 52.89%, and 50.40% of training weeks, respectively. Total capacity equal to mean demand is therefore insufficient for the stated service target.

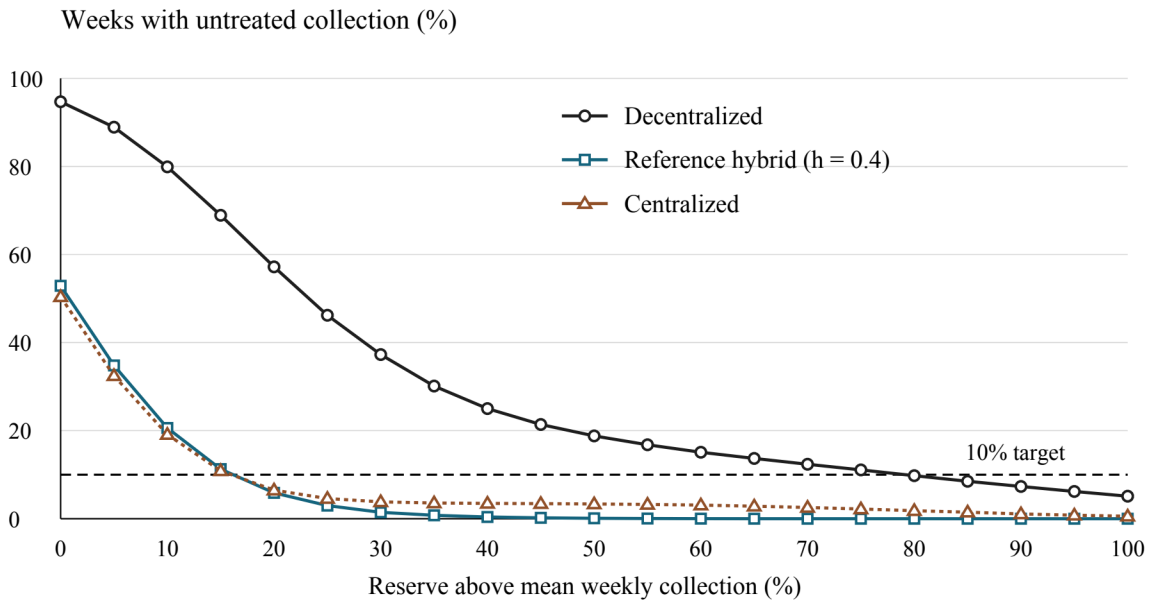


Figure 1. Training failure frequencies at equal reserve levels for decentralized, reference-hybrid (h = 0.4), and centralized designs. Each reserve level has identical total capacity across curves. The dashed 10% line is the target; formal screening uses simultaneous upper bounds, not the plotted point estimates.

At 20% reserve, the same three architectures have failure frequencies of 57.19%, 5.88%, and 6.49%. Equal total capacity does not imply equal usable protection: a decentralized network cannot share idle capacity, while the shared facility absorbs residual loads. The reference hybrid's frequency falls below the centralized frequency at higher reserves, where concentrating disruption exposure increasingly offsets the benefit of complete pooling.

4.2 Selected designs and held-out performance

Table 3a. Original six nominations: capacity, frequency, and expenditure on 100,000 held-out weeks (seed 121026).

Design (h, s)	Capacity kg/week	Failure %	Upper bound %	Mean U kg/week	Expenditure million IDR/week
Decentralized (0, 0.85)	3,904.49	8.20	8.41	2.97	3.9326
Hybrid (0.2, 0.25)	2,638.17	2.35	2.47	2.38	3.7511
Centralized (1, 0.20)	2,532.64	6.53	6.72	28.38	3.9037
Hybrid runner-up (0.1, 0.20)	2,532.64	8.35	8.57	7.98	3.7236
Centralized + severity (1, 0.75)	3,693.44	2.23	2.34	4.53	4.0927
Decentralized benchmark (0, 0.80)	3,798.96	9.45	9.67	3.64	3.9089

Note. Upper bounds provide at least 95% simultaneous one-sided coverage across these six fixed nominations. U is accepted mass left untreated. Expenditure excludes the shortfall penalty. Rounded capacities are calculated from Equation (2), not transferred between design rows.

Table 3b. Severity of the same six nominations (kg; seed 121026).

Design (h, s)	$CVaR_{0.95}$	$CVaR_{0.99}$	Mean U given failure
Decentralized (0, 0.85)	53.46	106.09	36.23
Hybrid (0.2, 0.25)	47.64	186.00	101.36
Centralized (1, 0.20)	557.38	975.90	434.83
Hybrid runner-up (0.1, 0.20)	144.30	292.44	95.53
Centralized + severity (1, 0.75)	90.69	337.47	203.51
Decentralized benchmark (0, 0.80)	61.78	116.41	38.56

The selected hybrid has h = 0.2 and s = 0.25. Equation (2) gives $L_i = \mu_i$ and $H = 0.25\mu \approx 527.6$ kg/week: local units cover school means and the hub holds the reserve. The reported 7.24% is the ratio of mean shared-treated mass (152.87 kg/week) to mean accepted mass (2,111.01 kg/week). The mean of weekly ratios, $n^{-1}\sum\omega[(\sum_i y_i\omega)/(\sum_i Q_i\omega)]$, is 6.79%. The former weights weeks by accepted mass; the latter weights weeks equally.

The hybrid's mean expenditure has Monte Carlo standard error IDR 1,018/week, and mean untreated mass has standard error 0.067 kg/week. Its planning objective is IDR 3.7630 million/week. The runner-up has lower expenditure but greater expected shortfall. The paired difference in objective, runner-up minus hybrid, is IDR 539/week, with an approximate 95% interval of IDR 102–975/week. This is about 0.014% of the hybrid objective: statistically detectable under the model but economically small. Selection stability is examined separately in Section 4.7; statistical precision does not make this difference economically large. However, the frequency-screen runner-up has held-out $CVaR_{0.95}$ of 144.30 kg, so economic proximity does not make it interchangeable under the 100-kg severity criterion.

The first decentralized design to pass the training screen has 85% reserve. At 80% reserve, the training failure estimate is 9.78%, but its simultaneous upper bound is 10.25%. The separately nominated 80% benchmark passes held-out evaluation with an upper bound of 9.67%. The selected 85% reserve therefore reflects the conservative screening rule; it does not establish that 80% reserve is physically inadequate.

4.3 Failure frequency and shortfall severity

Table 4. Shortfall severity for the three primary architecture candidates.

Metric (kg unless stated)	Decentralized	Hybrid	Centralized
Mean untreated mass per week	2.97	2.38	28.38
95th-percentile untreated mass	19.46	0.00	75.57
$CVaR_{0.95}$: mean of worst 5%	53.46	47.64	557.38
$CVaR_{0.99}$: mean of worst 1%	106.09	186.00	975.90
Mean untreated mass given failure	36.23	101.36	434.83
Illustrative avoided emissions (kg CO_2e/week)	642.95	639.44	584.72
Illustrative recovered wet mass (kg/week)	527.01	524.10	479.01

All three candidates pass the held-out frequency screen, but centralized $CVaR_{0.95}$ is approximately 11.7 times the hybrid value. The hybrid fails less often than the decentralized design, yet its conditional mean shortfall is larger. Its 2.35% failure frequency puts zero-loss weeks into the worst-5% average: 2.382/0.05 ≈ 47.64 kg, while mean shortfall given failure is 101.36 kg. Thus, passing the 100-kg screen does not mean that failed weeks remain below 100 kg.

Held-out $CVaR_{0.99}$ values are 106.09, 186.00, and 975.90 kg for the decentralized, hybrid, and centralized candidates, respectively. Pointwise 95% bootstrap intervals for their $CVaR_{0.95}$ values are 52.17–54.73, 45.11–50.26, and 541.11–573.23 kg. The hybrid $CVaR_{0.99}$ interval is 178.04–194.64 kg. These diagnostics quantify simulation sampling uncertainty and show the higher severity revealed by a narrower tail; the original selection still uses $CVaR_{0.95}$. These selected designs have different capacities; Section 4.6 isolates architecture at equal capacity.

For full centralization, let Q be total weekly collection, p_0 the shared disruption probability, and ρ the residual availability. Independence between load and disruption gives $P(U > 0) = (1 - p_0)P(Q > H) + p_0P(Q > \rho H)$, apart from the numerical tolerance. With $p_0 = 0.035$, $\rho = 0.55$, and $H = 1.20\mu$, held-out load exceedance frequencies are 3.08% in the undisrupted state and approximately 100% at disrupted capacity. Weighting these gives 6.47% failure probability, close to the direct joint-simulation estimate of 6.53%. This integrates out the random disruption using the same load sample and provides a structural consistency check.

During a disruption, capacity is $0.55 \times 1.20\mu = 0.66\mu$. At mean load, the shortfall is 0.34μ, approximately 718 kg. The reserve needed merely to preserve mean-load capacity at 55% availability is $1/0.55 - 1 =$

81.8%; this is an algebraic reference, not a guarantee under unbounded random loads. A frequency constraint can tolerate large disrupted-week losses because disruptions are relatively infrequent.

The 100-kg $CVaR_{0.95}$ screen leaves the selected hybrid and decentralized designs unchanged and raises the selected centralized reserve from 20% to 75%. The latter has held-out $CVaR_{0.95}$ of 90.69 kg (pointwise bootstrap interval 86.06–95.48 kg) and mean expenditure IDR 4.0927 million/week. Yet its conditional failed-week mean is 203.51 kg and $CVaR_{0.99}$ is 337.47 kg. A passed severity screen must therefore be interpreted at its stated tail level. Only the binomial bounds provide the specified simultaneous failure-probability coverage. Section 4.7 reports how additional severe-week criteria alter selection.

4.4 Dependence and disruption stress tests

Table 5. Fixed-design stress tests: failure frequency (%) / $CVaR_{0.95}$ (kg).

Scenario	Decentralized	Hybrid	Centralized
Baseline validation	8.20 / 53.46	2.35 / 47.64	6.53 / 557.38
Independent school factors (a = 0)	8.17 / 51.53	0.47 / 5.56	3.69 / 519.38
Stronger common factors (a = 0.8)	8.01 / 52.16	5.30 / 157.83	10.77 / 624.64
50% larger load deviations	9.24 / 83.85	8.71 / 246.38	13.08 / 656.72
No facility disruptions	0.03 / 0.15	1.20 / 20.34	3.05 / 60.59
Additional 2% common disruption	9.62 / 89.19	4.27 / 305.39	8.39 / 754.84

Note. Baseline validation uses 100,000 weeks (seed 121026); the five shifts use 50,000 weeks (seed 131026). The complete pointwise 95% binomial and bootstrap intervals for all three architectures, plus paired differences from a 50,000-week matched baseline, are in stress-availability-uncertainty.csv. Original validation CVaR intervals are in severity-uncertainty.csv. Capacities remain fixed; these are not newly certified designs.

The selected hybrid breaches the 100-kg $CVaR_{0.95}$ limit in three of the five stress tests: stronger common factors (157.83 kg), 50% larger load deviations (246.38 kg), and an additional common disruption (305.39 kg). Its failure-frequency point estimates remain below 10% in all three. This directly illustrates that acceptable failure frequency can coexist with unacceptable shortfall severity. The design is feasible under the baseline screens; these tests do not demonstrate robust feasibility. Pointwise 95% bootstrap intervals for the three hybrid $CVaR_{0.95}$ estimates are 150.1–165.3, 238.9–254.0, and 288.1–323.0 kg. Paired increases over the matched baseline have intervals 107.8–118.3, 196.8–206.6, and 244.3–277.7 kg. Their distance above 100 kg is not explained by Monte Carlo variation at this sample size.

Removing shared load variation reduces hybrid $CVaR_{0.95}$ to 5.56 kg. Shared reserve provides less protection when local and shared capacity decline together. Removing all disruptions reduces decentralized failure frequency to 0.03%, but that design has more installed capacity than the selected hybrid and centralized designs. Because capacities differ, this comparison cannot isolate network structure. Figure 2 reports the hybrid stress results.

Table 6 summarizes the selected hybrid's sensitivity to facility availability. Higher local disruption probability and lower residual throughput breach the 100-kg severity limit, while the joint adverse case exceeds both the frequency and severity targets. Full results for all three architectures are provided in the supplement.

Table 6. Fixed-hybrid availability sensitivity with pointwise 95% intervals.

Scenario	Failure % [95% interval]	$CVaR_{0.95}$ kg [95% interval]
Matched baseline	2.35 [2.22, 2.48]	44.8 [41.3, 48.4]
Local disruption 10%	5.88 [5.67, 6.09]	133.4 [127.4, 139.0]
Shared disruption 10%	3.27 [3.11, 3.43]	65.5 [61.3, 69.8]
Residual availability 25%	4.83 [4.64, 5.02]	118.0 [112.0, 124.4]
Joint adverse case	18.82 [18.48, 19.17]	406.2 [398.3, 414.1]

Note. Entries are estimates [95% intervals]. All rows use 50,000 matched scenarios (seed 131026). Frequency intervals are exact binomial; CVaR intervals use 2,000 weekly percentile bootstrap replicates. The joint case combines 10% local and shared disruption probabilities with 25% residual availability. Other parameters and capacities remain fixed. Table 3 uses a separate 100,000-week sample. Complete paired differences and other architectures appear in stress-availability-uncertainty.csv.

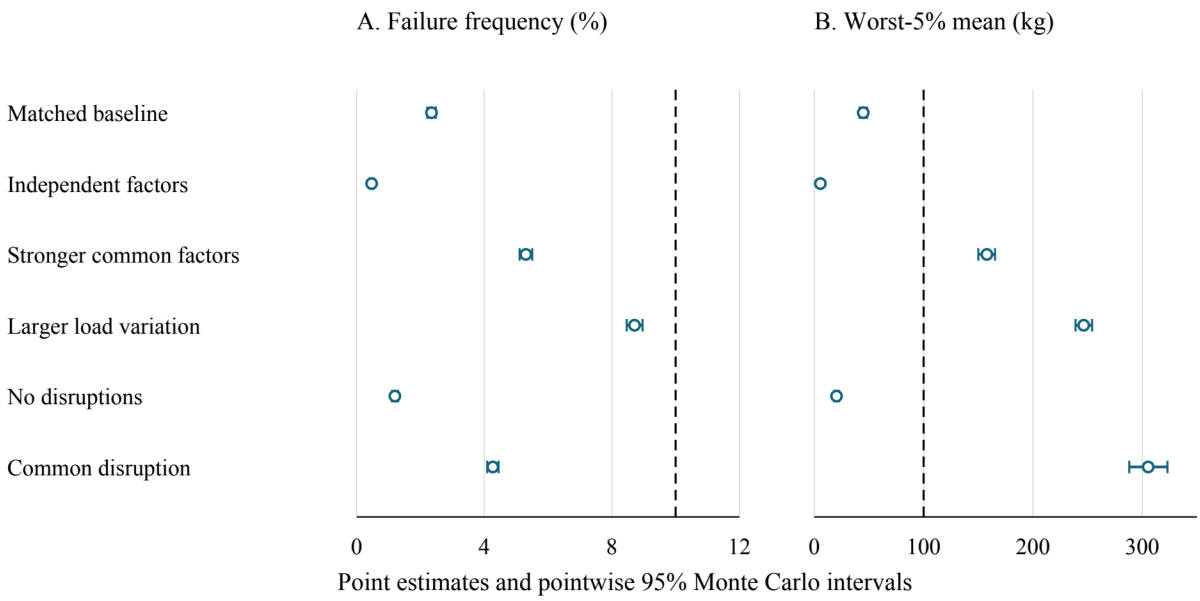


Figure 2. Fixed-hybrid stress results with pointwise 95% binomial frequency intervals and 2,000-replicate weekly bootstrap $CVaR_{0.95}$ intervals. Every point uses 50,000 weeks (seed 131026), including the matched baseline. Dashed lines mark the 10% frequency and 100-kg severity targets. This baseline differs from Table 5's 100,000-week validation row.

4.5 Cost sensitivity, environmental indicators, and service distribution

The selected capacity split changes with costs and the shortfall penalty. Reducing the penalty to IDR 2,500/kg selects h = 0.1, s = 0.20; raising it to IDR 10,000/kg selects h = 0.2, s = 0.30. Halving capacity charges selects h = 0.1, s = 0.30, whereas doubling them selects h = 0.2, s = 0.20. Increasing transport cost by 25% selects h = 0.1, s = 0.25, and doubling it selects h = 0.1, s = 0.30. These training-sample sensitivity results favor several nearby hybrid designs; these cost scenarios are training sensitivities rather than independently certified economic choices.

Under the illustrative environmental coefficients, the decentralized candidate avoids 642.95 kg CO_2e/week, the hybrid 639.44 kg, and the centralized candidate 584.72 kg. A 640-kg training-sample emissions floor changes the selected design to h = 0.1, s = 0.25; floors of 600 or 620 kg retain the baseline hybrid. These results depend on accounting assumptions. At equal treated mass, shared treatment has lower process emissions by 0.020 kg CO_2e/kg, but transport removes that advantage after 0.020/0.0045 = 4.44 km. Every assumed distance is longer than this break-even distance, so the coefficients structurally favor local processing per kilogram. This is not an independent empirical finding about the technologies.

A common avoided-disposal factor also controls the sign of the modeled benefit. For the hybrid, mean treatment-and-transport emissions per kilogram treated are approximately 0.097 kg CO_2e. A disposal counterfactual below that value would make its net benefit negative within this limited boundary; the

assumed baseline is 0.400. Similarly, modeled recovered wet mass depends directly on unvalidated yield coefficients and should not be read as nutrient recovery or verified product sales.

Distance-priority dispatch creates unequal service exposure. In the hybrid, the farthest school S5 has shortfall in 2.16% of held-out weeks and loses 0.45% of accepted mass, whereas no S1 shortfall is observed. Under centralization, S5 loses 4.75%. These outcomes follow from cost-minimizing dispatch, not poorer sorting or lower participation at S5; an observed zero is not zero population risk.

With proportional allocation of scarce shared capacity to residual loads, total shortfall is unchanged in every held-out week. For the hybrid, S5's untreated fraction falls to 0.11%, while S1's rises from zero to 0.09%; mean expenditure rises by IDR 388/week (paired Monte Carlo standard error IDR 10/week). This redistributes losses rather than removing them and is not a guarantee of equal service. It shows that aggregate frequency and CVaR can remain unchanged while school-level service changes materially.

4.6 Equal-capacity architecture comparison

At $s = 0.25$, all eleven shared fractions have 2,638.17 kg/week of rated capacity and face identical held-out loads and disruptions (Table 7). $CVaR_{0.95}$ is U-shaped, with its minimum at $h = 0.2$. Conditional mean shortfall instead increases across h: decentralization fails frequently with relatively smaller failures, whereas concentration exposes the network to larger failed-week losses. Neither ranking substitutes for the other.

Table 7. Equal-capacity outcomes at s = 0.25 (100,000 weeks; seed 121026).

h	Failure %	Mean U kg/week	$CVaR_{0.95}$ kg	$CVaR_{0.99}$ kg	Conditional mean kg
0.0	46.30	29.66	207.53	309.38	64.06
0.1	4.23	3.62	72.42	208.29	85.71
0.2	2.35	2.38	47.64	186.00	101.36
0.3	2.43	2.73	54.68	210.09	112.60
0.4	2.94	3.89	77.71	267.46	132.03
0.5	3.60	5.92	118.49	354.28	164.39
0.6	4.17	8.81	176.23	459.05	211.26
0.7	4.54	12.35	246.94	571.96	272.08
0.8	4.67	16.23	324.64	686.71	347.87
0.9	4.67	20.26	405.27	802.07	433.44
1.0	4.64	24.33	486.62	917.86	523.81

Note. Every row has a capacity of 2,638.17 kg/week. Mean U is kg/week; both CVaRs and the conditional mean are kg. Failure means $U > 10^{-8}$ kg. Full values for all 231 designs and additional reserve slices are in final-grid-heldout.csv and equal-capacity.csv.

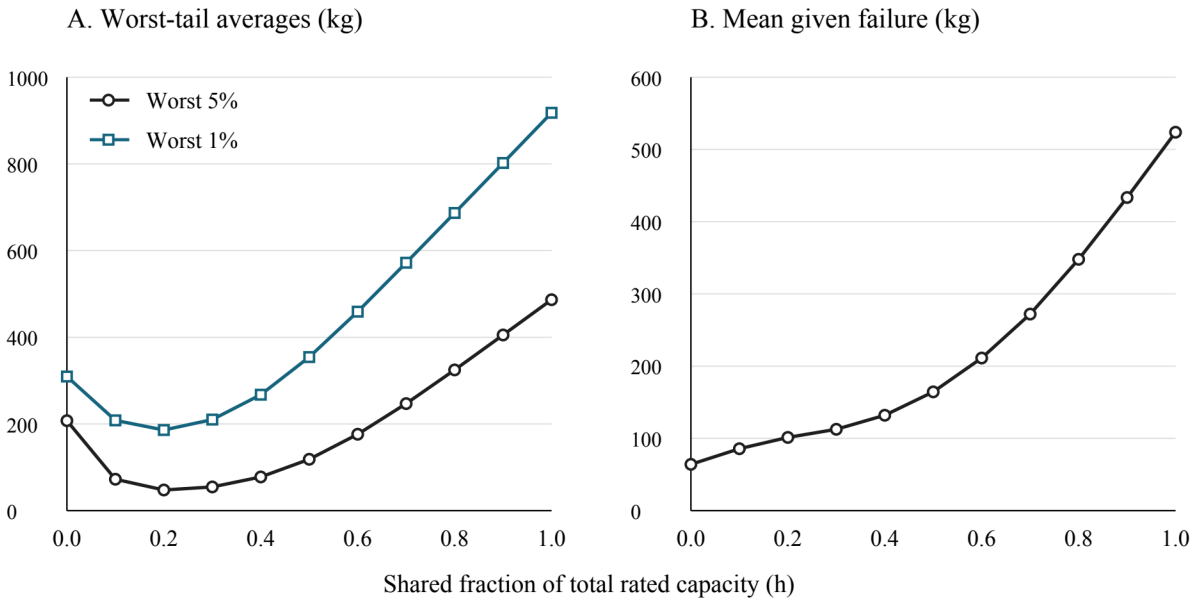


Figure 3. Severity versus shared-capacity fraction at constant 25% reserve. Left: worst-5% and worst-1% averages. Right: mean shortfall conditional on failure. Identical 100,000 held-out weeks isolate capacity placement; conditional mean increases even where $CVaR_{0.95}$ decreases.

4.7 Alternative severity screens and selection stability

Adding $CVaR_{0.99} \leq 200$ kg to the original screens retains the hybrid (0.2, 0.25) but increases the centralized reserve from 0.75 to 0.90. Adding conditional mean ≤ 100 kg instead selects hybrid (0.1, 0.25); imposing both selects (0.2, 0.30). The decentralized winner remains (0, 0.85) in all three cases. No centralized design meets the conditional-mean limit within the evaluated grid, even at 100% reserve. These are new sensitivity selections, not components of the original optimization.

Table 8. New fixed nominations on 100,000 independent weeks (seed 221026).

(h, s)	Failure %	Expenditure	$CVaR_{0.95}$	$CVaR_{0.99}$	Conditional mean
(0, 0.85)	8.25	3.9311	52.9	106.1	35.7
(0.1, 0.25)	4.26	3.7443	75.0	213.0	87.9
(0.2, 0.25)	2.38	3.7492	49.7	191.7	104.3
(0.2, 0.30)	1.05	3.7620	19.9	99.2	94.8
(1, 0.20)	6.45	3.9029	547.2	981.2	431.7
(1, 0.75)	2.13	4.0911	89.5	342.8	209.8
(1, 0.90)	1.07	4.1359	33.8	168.7	157.8

Note. All seven distinct nominations are shown. Expenditure excludes the penalty and is in million IDR/week; severity is kg. The supplement supplies mean shortfall, objective, capacity, and simultaneous binomial upper bounds for these seven. This sample differs from Tables 3–4.

On the fresh sample, the original hybrid's conditional mean is 104.3 kg and $CVaR_{0.99}$ is 191.7 kg. The conditional-only winner reduces conditional mean to 87.9 kg but has $CVaR_{0.99}$ of 213.0 kg. The combined-screen winner has 94.8 kg conditional mean and 99.2 kg $CVaR_{0.99}$, with higher expenditure. Thus, smaller average failed-week loss need not imply a smaller extreme tail.

Thresholds affect both feasibility and ranking. For $CVaR_{0.99}$ limits 100, 150, 200, 300, and 400 kg, respectively, 104, 111, 118, 130, and 134 designs pass the combined training screens; limits 100–150 select (0.2, 0.30), while 200–400 retain (0.2, 0.25). Conditional limits 50, 75, 100, 150, 200, and 250 kg leave 36, 44, 61, 112, 131, and 134 eligible designs, selecting (0.2, 0.60), (0.1, 0.35), (0.1, 0.25), then (0.2, 0.25) for the last three. Centralized feasibility under the conditional screen begins among these tested limits at 150 kg (reserve 0.95). Every design's membership and family winner are recorded in severity-feasibility.csv and severity-threshold-selections.csv.

Table 9. Original selection repeated across 20 independent training samples.

Original screening criterion	Selected (h, s)	Wins / replications
Frequency only	(0.2, 0.25)	20 / 20
Frequency + $CVaR_{0.95} \le 100$ kg	(0.2, 0.25)	20 / 20

The original hybrid wins all 20 replications under both original screening procedures; no neighboring hybrid wins. Across these samples, the runner-up (0.1, 0.20) minus original-winner objective difference ranges from IDR 108 to 1,029/week (median 702). This supports finite-model selection stability, while the very small advantage and sensitivity to service preferences limit an exact practical recommendation.

4.8 Expenditure structure, facility charges, and failure tolerance

Table 10 decomposes the objective on the original held-out sample. Treatment is about 76% of hybrid expenditure and 77% for the runner-up, so their similar processed volumes leave limited scope for differences in that large component. The percentage is not universal: full centralization substitutes lower treatment cost and much greater transport expense.

Table 10. Objective components for original nominations (million IDR/week).

Component	D	H	C	H′
Capacity charges	0.8395	0.5197	0.3166	0.5217
Fixed charges	0.1840	0.2760	0.0920	0.2760
Treatment	2.9091	2.8503	2.0618	2.8699
Transport	0.0000	0.1051	1.4333	0.0560
Shortfall penalty	0.0149	0.0119	0.1419	0.0399
Planning objective J	3.9474	3.7630	4.0456	3.7635

Note. D, H, C, and H′ denote decentralized, hybrid, centralized, and hybrid runner-up in Table 3. Expenditure is the first four rows; the objective also includes the fifth. Full precision and all six designs are in objective-components.csv.

Subtracting the common reference 1,380E[Q] from J leaves capacity and fixed charges − 390E[X] + 70E[$\sum_i d_i y_i$] + 3,620E[U], where Q = $\sum_i Q_i$ and X = $\sum_i y_i$. This algebra separates design-sensitive terms without changing rankings; actual treatment expenditure itself is not constant. The original J remains the selection objective.

Under the alternative shared-facility charge, frequency-only screening selects (0.1, 0.20), and frequency plus $CVaR_{0.95}$ selects (0.1, 0.25). Decentralized and centralized family winners are unchanged. Hybrids remain competitive, but the exact shared fraction depends on the assumed fixed-charge structure.

Using the 1% operational allowance retains the overall hybrid (0.2, 0.25), reduces the frequency-screened decentralized reserve from 0.85 to 0.70, and leaves the centralized frequency winner at 0.20 (0.75 with the original severity screen). For the original three designs, training strict/1%-allowance failure frequencies are 8.50/4.85%, 2.41/1.90%, and 6.49/5.89%, respectively. The hybrid retains the lowest frequency, but

the decentralized and centralized rankings reverse under the allowance. Feasibility expands: the newly selected decentralized design has an allowance-based frequency of 8.27%. The allowance does not remove untreated mass from the penalty or severity calculations.

5. DISCUSSION AND IMPLICATIONS FOR SCHOOL WASTE SYSTEMS

5.1 Reserve placement and service targets

At equal capacity with 25% reserve, moderate sharing reduces the worst-tail average shortfall relative to either endpoint, but the average failed-week shortfall increases with centralization. Shared reserve pools local excesses; shared outages concentrate losses. The selected-design comparison addresses a different decision: economical configurations meeting specified screens, with decentralized reserve substantially larger than hybrid reserve. Its approximately 557-versus-48 kg centralized/hybrid $CVaR_{0.95}$ contrast cannot by itself isolate architecture.

Twenty independent replications select the same baseline hybrid, supporting simulation stability under fixed assumptions. The IDR 539/week held-out advantage remains economically small. Alternative severe-week requirements select different nearby hybrids or additional reserve, and the partially capacity-dependent facility charge shifts the shared fraction. Stable numerical selection is therefore compatible with sensitivity to the decision maker's service requirements and cost model.

A low failure frequency does not imply small backup needs. Below 5% failures, $CVaR_{0.95}$ becomes a scaled mean weekly shortfall; the centralized 75%-reserve candidate passes it while averaging about 204 kg untreated during failures. Conditional means and narrower tails expose different dimensions of this risk, and neither bounds every loss. The hybrid's three stress-test breaches remain clear after Monte Carlo uncertainty is quantified. Backup capacity should be assessed against both failure occurrence and loss magnitude, with service targets agreed before selection.

5.2 Application to school waste-management systems

The model is a distributed-capacity framework motivated by schools. Shared off-site handling has an international precedent: New York City's school collection program serves multiple schools, and the city's 2026 waste plan identifies school organics among the inputs to the Staten Island Compost Facility (New York City Department of Sanitation, n.d., 2026). This documents shared municipal infrastructure, not the specific eight-school hybrid modeled here or an existing facility in Semarang or Jakarta.

An operating record should identify the school, date, and material stream, with masses for gross collection, rejection, local treatment, transfer to shared treatment, and material sent elsewhere. Accepted organic mass must equal the sum of its recorded destinations. Records should also include attendance or contributing population, collection participation, disruption duration, available throughput, transport distance, cost, and recovered product, with units stated. Plastics and other recyclables need separate records.

Weekly reports should distinguish collection coverage, rejection, untreated accepted mass, failure frequency, severe-week losses, and service by school. The proportional-allocation experiment shows why aggregate reliability is insufficient: identical network shortfall can conceal different school-level burdens. Lower collection should not be mistaken for better treatment service; collection coverage remains unknown without potential-generation measurements.

Before a capacity purchase, the synthetic inputs would need to be replaced with measurements taken across schools over the same periods, observed facility availability, and supplier-specific throughput and cost information. Later observations, including peak periods and disruptions, should be reserved for prospective validation. Reserve options should be tested against available backup capacity and an agreed allocation rule. The proportional residual-load comparison here illustrates one alternative. Minimum guaranteed service or other priority rules require separately specified constraints and evaluation.

The common-disruption results also identify a need to test backup arrangements that do not share the same failure source. An independent treatment provider or alternative collection arrangement is one possible option. Such arrangements are outside the present model. Their capacity and availability would need to be measured before assessing whether they can cover the simulated shortfalls.

5.3 Limitations and validity

The principal limitation is input validity. School populations and collection boundaries must be measured consistently; larger simulations cannot calibrate assumed costs, probabilities, or yields. Term breaks, holidays, attendance, examinations, and canteen schedules can change both load and available staffing. Independent five-day weeks omit these calendar effects and temporal persistence, potentially misrepresenting peak overlap and backup needs. No calendar scenario is fitted without supporting observations. Common events involving weather, transport, power, or staffing may also differ from the hypothetical stress tests.

The model represents aggregate weekly flows. It omits spoilage, within-week timing, serially correlated loads, storage, biological residence time, routing capacity, trip consolidation, facility-specific minimum scale, alternative hubs, and transfers between schools. It assumes that both local and shared units can treat every accepted organic kilogram. These simplifications limit operational use; a pilot would require technology-specific definitions of capacity and recovery.

Cost coefficients influence both the dispatch rule and the selected design. Under the assumed distances, environmental coefficients favor local treatment per kilogram. Product yield measures wet mass and gives no information about dry matter, nutrients, or safe end use. The calculations do not establish actual emissions savings, educational outcomes, or social effects. The school-level results also show that minimizing network cost can produce unequal service.

Statistical screening is conditional on independent weeks from the assumed distribution. Bonferroni bounds are conservative, as the 80%-reserved decentralized benchmark illustrates. Stability across 20 samples and the paired runner-up interval concern simulation variation, not a continuous global optimum

or parameter uncertainty. Severity screens and bootstrap intervals remain pointwise; they do not give simultaneous severity certification after searching the grid. The full-grid held-out comparison is descriptive, while new severity nominations use a separate evaluation sample. Availability and allocation experiments keep capacities fixed and do not establish performance across all parameters or fairness criteria.

6. CONCLUSION

The controlled experiment shows how local and shared reserve placement changes failure frequency, shortfall severity, and service distribution. At equal rated capacity, moderate sharing lowers the worst-5% average at 25% reserve, while conditional failed-week losses increase as capacity is concentrated. These measures answer different backup-planning questions.

The original hybrid is stable across 20 independent training samples, but its economic advantage is small. A worst-1% screen retains that choice, a conditional-mean screen selects a different hybrid, and combining both adds reserve. Alternative facility charges also change the preferred split. The original $\mathrm{CVaR}_{0.95}$ screen therefore cannot stand alone as a guarantee of manageable failed-week shortfalls.

The hybrid still breaches its original severity limit in three stress tests. Practical capacity decisions require measured school loads and calendars, technology-specific availability, realistic facility charges, assessed backup capacity, and an agreed allocation rule. The framework supports comparing those decisions; the simulation results do not establish operational readiness.

DATA AND CODE AVAILABILITY

The versioned reproducibility package (22 October 2026) contains analysis.py for the original grid and validation, review_extensions.py for the earlier severity, availability, and allocation checks, and final_extensions.py for the additional selections, equal-capacity comparisons, 20-replication stability study, cost decomposition, operational tolerance, and uncertainty analyses. It includes synthetic inputs, exact seeds and dependency versions, machine-readable results, high-resolution figures, and execution instructions. No individual student records or field measurements are included.

The reproducibility package is publicly available through Zenodo at https://doi.org/10.5281/zenodo.23570440.

REFERENCES

Antón-Peset, A., Fernandez-Zamudio, M.-A., & Pina, T. (2021). Promoting food waste reduction at primary schools. A case study. Sustainability, 13(2), Article 600. https://doi.org/10.3390/su13020600

Bilqis, A. T., Abdillah, A., Jahroh, S., Abdullah, A., Dewi, F., Tjahjono, B., Rakhmani, I., & Priadi, C. R. (2023). Anaerobic digestion potential in traditional boarding school. International Journal of Technology, 14(7), 1496–1505. https://doi.org/10.14716/ijtech.v14i7.6699

Clopper, C. J., & Pearson, E. S. (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrika, 26(4), 404–413. https://doi.org/10.1093/biomet/26.4.404

Eppen, G. D. (1979). Note—Effects of centralization on expected costs in a multi-location newsboy problem. Management Science, 25(5), 498–501. https://doi.org/10.1287/mnsc.25.5.498

Gambella, C., Maggioni, F., & Vigo, D. (2019). A stochastic programming model for a tactical solid waste management problem. European Journal of Operational Research, 273(2), 684–694. https://doi.org/10.1016/j.ejor.2018.08.005

Kuznetsova, E., Cardin, M.-A., Diao, M., & Zhang, S. (2019). Integrated decision-support methodology for combined centralized-decentralized waste-to-energy management systems design. Renewable and Sustainable Energy Reviews, 103, 477–500. https://doi.org/10.1016/j.rser.2018.12.020

Lotfi, R., Kargar, B., Gharehbaghi, A., & Weber, G.-W. (2022). Viable medical waste chain network design by considering risk and robustness. Environmental Science and Pollution Research, 29(53), 79702–79717. https://doi.org/10.1007/s11356-021-16727-9

Luedtke, J., & Ahmed, S. (2008). A sample approximation approach for optimization with probabilistic constraints. SIAM Journal on Optimization, 19(2), 674–699. https://doi.org/10.1137/070702928

Malashin, I. P., Martysyuk, D., Nelyub, V., Borodulin, A., Gantimurov, A., & Tynchenko, V. (2026). Multi-objective optimization of a regional biogas supply chain using organic waste. Scientific Reports, 16, Article 12593. https://doi.org/10.1038/s41598-026-42963-5

New York City Department of Sanitation. (2026). Revised draft solid waste management plan 2026: Attachments. https://www.nyc.gov/assets/dsny/downloads/resources/reports/solid-waste-management/2026-swmp/draft-swmp-2026-attachments-revised.pdf

New York City Department of Sanitation. (n.d.). Schools curbside composting. Retrieved October 22, 2026, from https://www.nyc.gov/site/dsny/collection/agencies-nonprofits/school-curbside-composting.page

Pagnoncelli, B. K., Ahmed, S., & Shapiro, A. (2009). Sample average approximation method for chance constrained programming: Theory and applications. Journal of Optimization Theory and Applications, 142(2), 399–416. https://doi.org/10.1007/s10957-009-9523-6

Qadar, S. A., Budihardjo, M. A., & Priyambada, I. B. (2024). Life cycle assessment to compare the environmental of food waste management system in Semarang City. Jurnal Presipitasi: Media Komunikasi dan Pengembangan Teknik Lingkungan, 21(1), 154–164. https://doi.org/10.14710/presipitasi.v21i1.154-164

Rathore, P., Chakraborty, S., Gupta, M., & Sarmah, S. P. (2022). Towards a sustainable organic waste supply chain: A comparison of centralized and decentralized systems. Journal of Environmental Management, 315, Article 115141. https://doi.org/10.1016/j.jenvman.2022.115141

Rockafellar, R. T., & Uryasev, S. (2000). Optimization of conditional value-at-risk. Journal of Risk, 2(3), 21–41. https://doi.org/10.21314/jor.2000.038

Saif, Y., Rizwan, M., Almansoori, A., & Elkamel, A. (2019). Municipality solid waste supply chain optimization to power production under uncertainty. Computers and Chemical Engineering, 121, 338–353. https://doi.org/10.1016/j.compchemeng.2018.11.003

Schmitt, A. J., Sun, S. A., Snyder, L. V., & Shen, Z.-J. M. (2015). Centralization versus decentralization: Risk pooling, risk diversification, and supply chain disruptions. Omega, 52, 201–212. https://doi.org/10.1016/j.omega.2014.06.002

Snyder, L. V., & Daskin, M. S. (2005). Reliability models for facility location: The expected failure cost case. Transportation Science, 39(3), 400–416. https://doi.org/10.1287/trsc.1040.0107

Wilkie, A. C., Graunke, R. E., & Cornejo, C. (2015). Food waste auditing at three Florida schools. Sustainability, 7(2), 1370–1387. https://doi.org/10.3390/su7021370

Xie, S., Lim, Y. T., Wang, H., Yi, W., & Antwi-Afari, M. F. (2024). Location and capacity optimization of waste recycling centers: Mathematical models and numerical experiments. Applied Sciences, 14(16), Article 7039. https://doi.org/10.3390/app14167039

Xiong, J., Ng, T. S. A., & Wang, S. (2016). An optimization model for economic feasibility analysis and design of decentralized waste-to-energy systems. Energy, 101, 239–251. https://doi.org/10.1016/j.energy.2016.01.080

Zhao, S., Haskell, W. B., & Cardin, M.-A. (2019). A flexible multi-facility capacity expansion problem with risk aversion [Preprint]. arXiv. https://doi.org/10.48550/arXiv.1905.05328