Computational Engineering and Technology Innovations

www.ceti.reapress.com

Comput. Eng. Technol. Innov. Vol. 3, No. 3 (2026) 226–248.

Paper Type: Original Article

Latency Aware Metaheuristics for Earthquake Early Warning Network Design

Mahmoud Alimoradi[1,*], Eisa Abdolmaleki[2], Hashem Saberi Najafi[3], Hadi Faghihmaleki[4]

[1] Department of Computer Engineering, Shafagh Institute of Higher Education, Tonekabon, Iran; mahmoud.alimoradi@shafagh.ac.ir.

[2] Department of Mathematics, To. C., Islamic Azad University, Tonekabon, Iran; eisa.abdolmaleki@gmail.com.

[3] Department of Applied Mathematics and Computer Science, Faculty of Applied Mathematics, University of Guilan, Guilan, Iran; hnajafi@aihe.ac.ir.

[4] Department of Civil Engineering, Ayandegan Institute of Higher Education, Tonekabon, Mazandaran, Iran; h.faghihmaleki@gmail.com.

Citation:

Received: 05 January 2026 *Revised: 16 February 2026* *Accepted: 03 May 2026*	Alimoradi, M., Abdolmaleki, E., Saberi Najafi, H., & Faghihmaleki, H. (2026). Latency aware metaheuristics for earthquake early warning network design. *Computational Engineering and Technology Innovations, 3*(3), 226-248.

Abstract

Earthquake Early Warning (EEW) network design couples a physical deadline with uncertain sensing, processing, and communication delays. An improved optimization score is useful only if it increases the probability that an actionable alert arrives before the relevant shaking. This paper develops a latency based interpretation of metaheuristic network design through a critical literature synthesis, a precise timing model, and a reproducible placement experiment. The model treats station reports as concurrent arrivals, represents confirmation by an order statistic, and includes central processing, delivery, action time, and station outages. We establish monotonicity under a fixed exogenous latency model, exhibit the failure of submodularity under multiple station confirmation, and specialize a finite family concentration bound to distinguish optimization error from sampling error. Five search methods are evaluated in six constructed planning cases using 900 runs and 1.62 million charged objective evaluations. Complete enumeration of 405,273 feasible placements supplies training optima for every case, and independent scenario sets assess generalization. Simulated Annealing (SA) and restarted swap search attain every small case optimum in all 90 runs per method, whereas performance becomes more variable in the larger cases. For prespecified placements, adding four seconds to the event path reduces weighted timely coverage by 9.33–10.71 percentage points, exceeding the largest median training gap among the tested solvers. The findings support deadline based evaluation and offline optimization, while remaining conditional on the stated geometry and delay distributions. They establish neither a universal algorithm ranking nor operational performance for a deployed warning network.

Keywords: Earthquake early warning, Sensor placement, Latency, Metaheuristic optimization, Scenario analysis, Reproducibility.

1 | Introduction

Earthquake Early Warning (EEW) begins after an earthquake has started. Its opportunity comes from detecting seismic motion and communicating information more quickly than damaging shaking reaches an

Corresponding Author: mahmoud.alimoradi@shafagh.ac.ir

https://doi.org/10.48314/ceti.v3i3.94

Licensee System Analytics. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0).

exposed location. The relevant outcome is therefore a timely and sufficiently reliable alert at the recipient, followed by an action that can still be completed. Allen and Melgar [1] situated this problem within the joint development of physical models, sensing systems, and user needs. The distinction matters for optimization. A design that predicts a source parameter accurately, covers a large geographical region, or minimizes an abstract network cost may still provide little usable warning where it is needed.

The available time is uneven. A nearby station can observe a compressional P wave before a more distant recipient experiences strong motion, but the station and recipient generally have different source distances. The point source P–S separation at a common location is only a simplified reference. Rupture evolution, the definition of damaging shaking, and the information needed to issue an alert introduce additional restrictions. Minson et al. [2] analyzed the timeliness imposed by earthquake growth, while Meier et al. [3] examined how often operational algorithms could warn sites experiencing high intensity motion. These studies made clear why a single headline warning time cannot describe the performance of an entire system.

Optimization enters this chain at several timescales. Station placement is normally a planning decision. Picker calibration, alert thresholds, and message dissemination policies can be updated between events or adapted during operation. A search algorithm executed after an event starts occupies part of the warning interval itself. Consequently, offline design quality and event path execution time require different evidence. An optimizer that takes minutes to select a network can be appropriate for planning; the same computation would be inappropriate if every alert had to wait for it. Conversely, a small inference time does not establish a small end to end latency if data acquisition or delivery dominates.

Metaheuristics are relevant because network layouts and threshold combinations create discrete, nonlinear objectives. Genetic algorithms have already been used for warning network design in Istanbul [4], Almaty [5], Switzerland [6], and the Lower Rhine Embayment [7]. The scientific question is no longer whether a search method can produce a plausible layout. It is whether its objective expresses the warning task, whether its advantage survives independent scenarios, and whether the computational effort belongs outside or inside the event path.

This paper addresses those questions through a critical synthesis and a deliberately inspectable computational study. We distinguish physical arrivals, station readiness, event confirmation, alert receipt, and completed action. The resulting placement objective measures expected recipient weighted timely coverage under explicitly stated joint scenarios. We then analyze the mathematical consequences of multiple station confirmation and finite scenario sampling. Finally, we compare random search, Simulated Annealing (SA), Differential Evolution (DE), Particle Swarm Optimization (PSO), and a restarted swap heuristic against complete enumeration. These are reference implementations with published algorithmic origins, not claimed replicas of every variant used in the EEW literature.

The contribution is an evaluation framework and an auditable benchmark rather than a new earthquake detection algorithm. The literature synthesis connects the evidence needed at each stage; the formal analysis explains why independent stage scores and simple greedy guarantees can fail; the experiments quantify the relative importance of solver quality and added event path delay within a controlled planning model. All newly reported numerical results follow from executable code and saved inputs. The study does not estimate earthquake magnitude, process waveforms, predict losses, or certify an operational system. Those boundaries preserve a direct connection between the question asked and the evidence provided.

2|Literature Review

2.1|Scope and Evidence Selection

The review used a focused, purposive search completed on 22 October 2026. Searches combined EEW with network optimization, station placement, warning time, phase picking, edge processing, alert delivery, and operational evaluation. Backward and thematic searches connected regional design studies to physical limits and user oriented performance measures. The included corpus comprises 26 EEW and seismic machine

learning publications and seven methodological foundations. Each reference was checked against a publisher, author or institutional record, or publisher deposited bibliographic metadata. Abstracts or relevant full text were inspected before assigning a methodological role. The companion extraction ledger records those roles and the boundary of each use.

This is a critical review with a reproducible computational illustration, rather than an exhaustive systematic review or a quantitative meta-analysis. The selection provides an argument spanning the warning pipeline; it does not support estimates of publication prevalence or pooled algorithm effects. Published values obtained with different regions, warning definitions, hardware, and data windows cannot be treated as exchangeable observations. We therefore compare modelling choices and evidence types, and reserve cross-algorithm numerical comparisons for the common benchmark introduced below.

2.2 | Physical Limits and the Meaning of a Successful Warning

Allen and Melgar [1] organized EEW methods around point source, finite fault, and ground motion approaches, while relating their development to users and protective actions. Minson et al. [2] showed that even rapid processing cannot eliminate the time needed for a rupture to develop information about its eventual size. Their later analysis examined the consequences of ground motion variability even when earthquake source parameters were known [8]. These results separate an information limit from a computing delay: Faster hardware can reduce the latter, but cannot guarantee advance knowledge of the former.

Meier et al. [3] compared the Earthquake Point Source Integrated Code (EPIC), the Finite Fault Rupture Detector (FinDer), and the Propagation of Local Undamped Motion (PLUM) on strong motion observations from Japan. Their retrospective analysis related successful warning to intensity, source geometry, and the alerting strategy. Saunders et al. [9] subsequently evaluated intensity threshold choices using historical ShakeMaps and population based as well as area based metrics. Minson et al. [10] framed alerting around an intended level of performance for a user's shaking threshold. Together, these studies showed why threshold calibration belongs in the warning problem itself. Lower thresholds can exchange more precautionary alerts for fewer missed or late warnings; the appropriate exchange depends on the action and its costs.

The present benchmark isolates the timing component by treating every source–recipient pair as eligible for an alert. It consequently cannot estimate false alert rates, intensity forecast accuracy, or an optimal magnitude threshold. This restriction is useful for studying placement and latency, but it differs from the decision problems addressed by those threshold studies. Its timely coverage must be read as a conditional planning measure, not as the fraction of real damaging earthquakes successfully warned.

2.3 | Network Design and Metaheuristic Search

Oth et al. [4] combined simulated earthquake scenarios with a network and algorithm optimization framework for Istanbul. Their results linked the usefulness of additional stations to the warning algorithm and the available station locations. Stankiewicz et al. [5] extended the regional design perspective to Almaty, considering trigger thresholds, potential station failure, accessibility, and existing instrumentation. These applications already went beyond uniform geometrical coverage. They treated warning performance as a consequence of a network operating under a specified detection rule.

Böse et al. [6] developed a loss based assessment for Switzerland, using stochastic earthquake catalogs and a genetic algorithm to identify additional stations. Warning time was evaluated alongside the severity of the affected exposure, bringing the optimization criterion closer to the risk reduction purpose of EEW. Najdahmadi et al. [7] used a hazard compatible stochastic catalog and microgenetic search in the Lower Rhine Embayment. Their network evaluation required threshold exceedance at three recording stations, and the reported system level warning times explicitly excluded inherent system latencies at that modelling stage. That stated scope is a reason to add a latency analysis when transferring a design objective to an operational interpretation, rather than a reason to dismiss the regional optimization result.

The methodological foundations also constrain what an algorithm comparison can mean. SA introduced probabilistic acceptance of unfavorable moves under a temperature schedule [11]. DE used population differences to construct continuous proposals [12], and PSO used personal and collective search information [13]. Random keys provided a way to represent discrete decisions through continuous vectors [14]. None of these origins supplies an automatic performance ranking for a warning network. The decoder, objective, restart rule, parameter choices, and computational budget determine the implementation actually being tested.

Greedy approximation results for monotone submodular objectives are equally specific. Nemhauser et al. [15] established bounds under a diminishing returns structure. Multiple station confirmation can create complementary sites: A station may be useful only after enough other stations are present. Section 3 gives a direct counterexample to submodularity for the timing objective. This observation motivates a lookahead baseline and direct search, while avoiding an unsupported approximation guarantee.

2.4 | Detection Quality and Execution Latency

The warning pipeline contains several distinct inference tasks. FinDer v.2 estimated finite source properties and ground motion through evolving spatial observations [16]. PLUM estimated imminent shaking through nearby observed motion without first requiring a complete earthquake source description [17]. These approaches answered different questions from a single station onset picker, and their output times cannot be inferred from picker benchmarks alone.

PhaseNet learned P and S arrival probabilities from labelled three component waveforms [18]. The convolutional and recurrent earthquake detector CRED learned signal characteristics for event detection [19]. EQTransformer combined detection and phase picking using attention [20], while the Transformer Earthquake Alerting Model (TEAM) mapped observations from variable station configurations to shaking forecasts [21]. Their published evaluations concerned different outputs, data sets, and availability of waveform context. A low onset error and an early decision are related but distinct quantities: A model can locate an onset precisely after seeing data that were unavailable at the time an alert would have needed to be issued.

Woollam et al. [22] introduced SeisBench to standardize access to models, data, and processing operations. Münchmeyer et al. [23] used a common framework to compare pickers and examined transfer between data domains. Those studies supported common data protocols and explicit task definitions. For a latency constrained deployment, the protocol must additionally state the causal input window, streaming stride, buffering, preprocessing, batch size, and hardware. An inference measurement on a completed window is evidence about computation on that window, not the time from earthquake origin to alert receipt.

Recent edge studies sharpened this distinction. Ravishan et al. [24] developed a lightweight convolutional model for New Zealand waveforms and reported Raspberry Pi inference measurements using two seconds of waveform input. Xiao et al. [25] compressed a phase picking model for edge processing and concise metadata transmission, using 30 second windows in their reported inference setting. Gou et al. [26] integrated distributed acoustic sensing into the dEPIC framework, combining phase picking, location, magnitude estimation, and quality controls. These contributions addressed different bottlenecks. Their window lengths, hardware times, telemetry requirements, and full alert decision times must remain separate in a synthesis.

2.5 | Operational Delivery and Decision Evidence

Ahn et al. [27] reported trial operation and a nearby station confirmation approach intended to reduce false alarms in a machine learning assisted warning system. Lux et al. [28] evaluated ShakeAlert performance during 2019–2023, including the effects of events near the network boundary. Operational evaluations of this kind connect algorithm outputs to actual network behavior, but they still require a distinction between an internal event solution and an alert received by a person or an automated client.

McBride et al. [29] directly tested Wireless Emergency Alert (WEA) latency and geofencing using controlled devices and community responses. The reported median delivery times varied across their tests; they were not universal distribution parameters for all channels or locations. Allen et al. [30] described the Android

earthquake detection and delivery system together with user feedback. Marcou et al. [31] combined earthquake waveform replay with an empirical delivery latency model in an end to end analysis of the 2023 Kahramanmaraş sequence. This latter work directly demonstrates that including delivery latency in EEW assessment already has a substantive precedent. Our contribution concerns the interaction of that requirement with placement search and finite benchmark verification.

Scenario based optimization provides a natural way to preserve a joint model of those delays. Kleywegt et al. [32] analyzed sample average approximation for stochastic discrete problems. Hoeffding [33] supplied bounds for sums of independent bounded observations. In the present application, the independent observation is a complete earthquake and network scenario, including correlated station and recipient delays. Treating every recipient in that scenario as an independent replicate would exaggerate the effective sample size. Section 3 applies the bounded observation argument to the entire recipient weighted scenario score.

Table 1 compares representative studies by their object, uncertainty treatment, evidence, and the scope boundary relevant to this paper. The final column identifies what must be added before transferring a reported result to our full timing objective. It does not imply that each source attempted every stage of the warning task.

Table 1. Representative literature and the evidence needed for a latency based placement interpretation.

Study	Object and Method	Uncertainty or Evidence	Evaluation Focus	Boundary for This Study
Oth et al. [4]	Network and threshold search	Istanbul scenario earthquakes	Warning and alarm cost	Specified regional decision rule
Stankiewicz et al. [5]	Genetic network design	Almaty scenarios; site constraints and failure	Useful warning and triggering	Regional assumptions and trigger calibration
Böse et al. [6]	Genetic station addition	Swiss stochastic catalogs and exposure	Warning by loss severity	Regional loss model; latency remains material.
Najdahmadi et al. [7]	Microgenetic design	Hazard compatible catalog	Three station warning classes	Inherent system latency excluded from core evaluation.
Meier et al. [3]	EPIC FinDer and PLUM	Japanese strong motion replay	Timely alerts by intensity	Retrospective algorithm comparison
Münchmeyer et al. [23]	Common picker benchmark	Multiple data domains	Detection identification and timing error	Causal windows and hardware need separate audit
McBride et al. [29]	WEA delivery testing	Devices and community observations	Receipt delay and geofence	Specific platforms and test conditions
Marcou et al. [31]	Replay with delivery model	Türkiye events; MyShake delivery data	End to end warning time	Event and channel specific analysis
Ravishan et al. [24]	Lightweight edge inference	New Zealand waveform segments	Classification and device inference	Two second input distinct from inference duration.
Xiao et al. [25]	Distilled edge picker	Waveforms and metadata communication	Picking and edge computation	30 second benchmark window requires causal interpretation.

Table 2 specifies the gaps addressed here. In particular, this work does not fill the separate gaps of real waveform validation or field delivery calibration. *Table 3* groups the literature into overlapping conceptual phases. Its period boundaries organize changes of emphasis; they are not claims that one approach replaced another.

Table 2. Questions addressed by the present analysis within the reviewed literature.

Question	Relevant Evidence	Contribution Here
Does the objective end at receipt and action.	Network design [4], [7]; end to end replay [31]	Section 3 gives a joint confirmation and recipient deadline model.
Does confirmation permit a greedy guarantee.	Three station design [7]; submodular theory [15]	Section 3 provides a complementarity counterexample.
How much error belongs to search versus scenarios.	Sample average optimization [32] and concentration [33]	Uniform finite family bound and enumeration separate the errors.
How large are solver gains relative to added delay.	Delivery tests [29] and regional optimization [6]	Sections 5 and 6 compare both on shared saved scenarios.

Table 3. Development of the concepts used in this paper.

Period	Conceptual Emphasis	Representative Evidence
2010–2013	Regional optimization of sites and triggering	Istanbul and Almaty [4], [5]
2018–2020	Physical limits and alternative inference objects	Timeliness; FinDer PLUM and learned picking [2], [16], [17], [20]
2021–2023	Common benchmarks and hazard or user objectives	TEAM SeisBench loss based design and delivery tests [6], [21], [22], [29]
2024–2026	Operational scale and full chain deployment evidence	ShakeAlert Android edge inference and replay with delivery [24], [25], [28], [30], [31]

3 | Timing Model and Mathematical Properties

3.1 | Arrivals and Concurrent Station Readiness

Consider a finite set of candidate stations, a set of alert recipients, and a scenario containing an earthquake and a complete realization of network delays and availability. Origin time is zero. Station selection is a subset S of the candidate set V, with a prescribed cardinality b. The confirmation requirement is an integer k between one and b. All distances are in kilometres and all times in seconds. For the reference experiment, a point source, a homogeneous medium, and positive velocities with $v_P > v_S$ give the arrival times,

$$t_{ej}^{P} = \frac{d_{ej}}{v_P}, \qquad t_{er}^{S} = \frac{d_{er}}{v_S}. \tag{1}$$

Here e, j, and r identify the earthquake, station, and recipient. The two distances are generally different. The use of S arrival as the target deadline is a modelling proxy. Strong shaking can develop differently because of rupture extension, directivity, surface waves, and local site effects. An application can replace the proxy with a measured or modelled threshold crossing time without changing the timing logic below.

Let a_{ej} be the usable data window and any local processing delay after the P arrival, and let h_{ej} be station telemetry delay. If a station is unavailable, its report time is infinite. Otherwise its report is ready centrally at:

$$R_{ej} = t_{ej}^{P} + a_{ej} + h_{ej}. \tag{2}$$

The confirmation time is the kth order statistic of the selected reports:

$$T_e(S) = \mathrm{kth}\{R_{ej} : j \in S\}. \tag{3}$$

If fewer than k selected stations report, this statistic is infinite. Reports arrive concurrently. Adding their elapsed times would describe a serial process and would not represent this confirmation rule. The model

presumes that a usable report is a correct qualifying report. It does not infer location quality, check false station triggers, or require an azimuthal geometry test. Those criteria could instead enter the confirmation function in an extended model.

Central decision time c_e, recipient specific routing delay q_{er}, and final delivery delay d_{er} follow confirmation. An optional additional event path computation has duration o_e. Receipt occurs at:

$$A_{er}(S) = T_e(S) + c_e + q_{er} + d_{er} + o_e. \quad (4)$$

These terms are elapsed increments after the preceding step. If a deployed architecture overlaps stages, its critical path should replace this serial suffix; summing overlapping durations would double count time. The model's purpose is to expose the dependency structure, not to impose a universal architecture on every EEW system.

For action duration u_r, the remaining time after completing the action is:

$$W_{er}(S) = t^S_{er} - A_{er}(S) - u_r. \quad (5)$$

A failed confirmation gives negative infinite margin. A negative finite margin records lateness and should not be suppressed in a timing audit. One may report the nonnegative usable duration:

$$W^+_{er}(S) = \max\{0, W_{er}(S)\}. \quad (6)$$

But this truncated value merges late alerts and no alerts at zero. Their frequencies should therefore be reported separately. *Fig. 1* shows the dependencies that determine the margin.

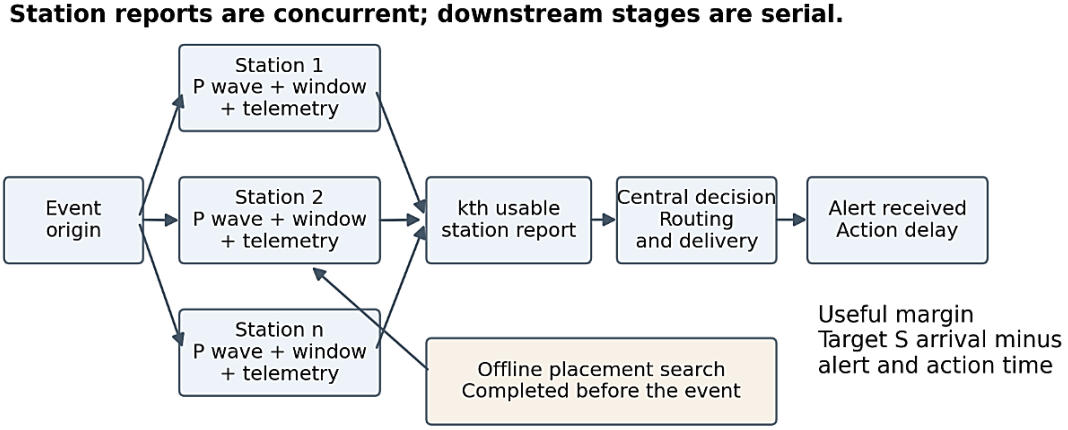


Fig. 1. Timing structure for the reference model.

Station reports arrive concurrently; the confirmation time is an order statistic. Planning search finishes before the event and contributes no event path delay. The serial suffix is an explicit architectural assumption (*Fig. 1*).

At a common source distance of 100 km, the experiment's velocities of 6 and 3.5 km/s give P and S arrivals of 16.67 and 28.57 s. The ideal separation is therefore positive 11.90 s. This elementary example does not describe the full network: A remote sensing station, three station confirmation, data windows, and delivery may change the actual margin. It fixes the physical sign and the units before any optimization is attempted.

3.2 | Timely Coverage and Group Reporting

Let the required residual lead time after action be a strictly positive value ℓ. Recipient weights w_r are nonnegative and sum to one. In each scenario, define a recipient deadline for station confirmation and the timely indicator by

$$D_{er} = t_{er}^{S} - u_r - \ell - c_e - q_{er} - d_{er} - o_e, \qquad U_{er}(S) = \mathbf{1}\{T_e(S) \le D_{er}\}. \tag{7}$$

All deadlines are finite in the experiment. Thus an unavailable confirmation can never be counted as timely. The indicator includes equality at the deadline. The recipient weighted planning objective is:

$$F(S) = \mathbb{E}\left[\sum_r w_r U_{er}(S)\right]. \tag{8}$$

The expectation includes the source scenario and the entire joint latency and outage vector. It is conditional on the model's alert eligibility assumption. It does not imply a loss reduction or a sensitivity to damaging shaking. An application would need an eligibility process, forecast errors, and action costs to support those interpretations.

A high weighted average can conceal weak service for some recipients. For any prespecified nonempty group G, define the equally weighted group measure:

$$F_G(S) = \mathbb{E}\left[\frac{1}{|G|}\sum_{r\in G} U_{er}(S)\right]. \tag{9}$$

In the reference experiment, groups are geographical coordinate bands rather than demographic categories. We report their performance and the minimum group score, but do not optimize an equity constraint. Using a minimum group constraint would define a different decision problem and would need an explicit policy target. Neither the weights nor the groups here are estimates of actual population vulnerability.

The finite design problem is:

$$S^* \in \operatorname{argmax}_{S\subseteq V,\,|S|=b} F(S), \tag{10}$$

Equal station costs make a fixed cardinality appropriate for the reference study. Unequal costs, mandatory sites, communication topology, power resilience, and restricted locations would change the feasible family. The experimental family contains every subset of the specified size; the general probability argument below also applies to any fixed finite feasible family.

Table 4. Principal notation and modelling conventions.

Symbol	Meaning	Convention
V, S, n, b	Candidates selected subset candidate count selected count	Fixed size family; no unequal costs
e, j, r	Source scenario station and recipient indices	Each scenario includes joint delays.
t_{ej}^{P}, t_{er}^{S}	P arrival at station and S arrival at recipient	Point source homogeneous proxy
$R_{ej}, T_e(S)$	Station report time and kth confirmation time	Infinite for unavailable confirmation.
c_e, q_{er}, d_{er}, o_e	Central routing delivery and added computation	Serial suffix after confirmation
u_r, ℓ, W_{er}	Action duration required lead and residual margin	Negative margin records lateness.
$w_r, U_{er}, F(S)$	Recipient weight timely indicator expected score	Weights sum to one.
N, M, η	Training scenarios feasible placements solver gap	Independent complete scenario vectors
F_G, G_N	Group coverage and training gap	Equal within group; gap in percentage points.

3.3 | Monotonicity and Complementary Stations

Proposition 1. Under fixed scenario report times and fixed downstream delays, adding stations cannot decrease timely coverage when the confirmation requirement is unchanged. Increasing any event path delay cannot increase timely coverage.

Proof: For a fixed scenario, enlarging a set of available report times cannot increase its kth smallest element. The result also holds when either order statistic is infinite. Each recipient indicator is consequently nondecreasing when stations are added. Increasing a serial delay decreases its deadline, so the same indicator is nonincreasing. Nonnegative weighting and expectation preserve both inequalities. The statement concerns the set function on all subsets, with zero coverage for sets smaller than the confirmation requirement. It does not compare objectives with a changing confirmation rule.

The exogeneity assumption is material. Adding sensors could increase traffic, queueing, or confirmation work in an actual system. Those effects are absent from the proposition and the experiment. The result therefore cannot justify unlimited sensor densification under shared congestion.

Proposition 2. For a confirmation requirement of at least two stations, the timely coverage set function need not be submodular.

Proof: Take one deterministic source–recipient scenario and three stations, all able to report before a common deadline. Require all three reports. Let the smaller set contain Station 1 and the larger set contain Stations 1 and 2. Adding Station 3 to the smaller set leaves coverage at zero; adding it to the larger set raises coverage from zero to one. The marginal gain is larger for the larger set, violating diminishing returns. The same construction works for two station confirmation with an empty smaller set and a singleton larger set, and extends to larger requirements by adding the necessary common stations.

This is a counterexample to a general property, not a claim that every instance is difficult or that greedy search always performs poorly. In particular, a triple lookahead can overcome the zero gain encountered before the first feasible confirmation in our three station model. It still has no general guarantee inherited from monotone submodular maximization.

3.4 | Sampling Error and Solver Error

Let a complete scenario vector be sampled independently from the specified distribution. The recipient weighted score for each vector lies between zero and one, even though the recipients and stations within it can be dependent. For N such vectors, the sample objective is:

$$\hat{F}_N(S) = \frac{1}{N}\sum_{i=1}^{N}\sum_{r} w_r U_{ir}(S). \tag{11}$$

Sample average approximation replaces the expectation with this common scenario average [32]. Every candidate is scored on the same fixed training set. New scenarios are reserved for assessment rather than generated afresh at each objective call.

Proposition 3. For a fixed feasible family containing M placements and independent scenario vectors, the following bound holds for every positive ε:

$$\Pr(\max_{S}|\hat{F}_N(S) - F(S)| > \varepsilon) \le 2M\exp(-2N\varepsilon^2). \tag{12}$$

Proof: For any fixed placement, Hoeffding's two sided inequality bounds the deviation of the average of independent observations in the unit interval by twice the exponential term [33]. Applying a union bound over the finite feasible family gives the result. No independence between candidates, stations, or recipients is required.

For a chosen error probability δ between zero and one, a sufficient uniform deviation radius is:

$$\varepsilon_{N,\delta} = \sqrt{\frac{\log(2M/\delta)}{2N}}. \tag{13}$$

Corollary 1. Let a solver return a placement whose training score is within η of the largest training score. With probability at least one minus δ, its population objective obeys:

$$F(S^*) - F(\tilde{S}) \le 2\varepsilon_{N,\delta} + \eta. \tag{14}$$

Proof: Insert the training scores of the population optimum and the returned placement between the two population scores. Their two sampling discrepancies are each at most the uniform radius, and their training difference is at most the solver gap.

This specialization of established concentration reasoning explains why exact training optimization is insufficient evidence of generalization. At 256 training scenarios and a 0.05 error probability, the radii are approximately 0.1390 for 495 placements and 0.1740 for 134,596 placements. These worst case bounds are broad; they are not tight confidence intervals for the observed differences. Independent assessment and a realistic scenario model remain necessary even when enumeration makes the training solver gap exactly known.

3.5 | Added Delay and Tail Accounting

For a fixed placement, add a nonnegative constant delay d after confirmation. The reduction in coverage has an exact interpretation:

$$F_0(S) - F_d(S) = \mathbb{E}\left[\sum_r w_r \, \mathbf{1}\{\ell \le W_{er,0}(S) < \ell + d\}\right]. \tag{15}$$

The affected recipients are those whose original residual margin lay inside a particular interval. This identity explains why the value of reducing latency depends on the margin distribution, rather than on the mean delay alone. It also makes a useful diagnostic: a flat sensitivity curve means that few scenario–recipient pairs lie near the decision boundary over that interval, not that latency is unimportant in general.

Marginal latency quantiles cannot simply be added and relabelled as an end to end quantile. For serial increments L_j, choose finite thresholds b_j whose individual exceedance probabilities are bounded by α_j. Then,

$$\Pr\left(\sum_j L_j \le \sum_j b_j\right) \ge 1 - \sum_j \alpha_j. \tag{16}$$

This follows because exceeding the sum of thresholds requires at least one component to exceed its threshold. The inequality does not require independent components, but it can be conservative. Adding several 95th percentile thresholds does not generally provide a 95th percentile guarantee for the sum. For example, two increments may each equal ten seconds on separate events of probability 0.04 and otherwise equal zero. Each individual 95th percentile is zero, while the sum is positive with probability 0.08. Shared congestion or outages create additional dependence that must be represented in joint scenarios.

Warning time quantiles require equal care. With a fixed deadline and continuous finite alert latency, a lower warning time quantile corresponds to an upper alert latency quantile through subtraction. Ordinary warning time quantiles themselves remain nondecreasing. With variable source distances, correlated deadlines, or no alert mass, the joint margin distribution must be computed directly. Our results therefore report timely coverage and explicit confirmation failure frequencies rather than attaching decreasing upper quantile labels to warning times.

4 | Reference Optimization Methods

4.1 | Representation and A Common Evaluation Contract

The benchmark fixes a candidate geometry and selects exactly the required number of stations. Continuous methods use one key per candidate, with the selected subset given by

$$S(x) = \{\text{indices of the b smallest entries of x}\}, \qquad x \in [0,1]^n. \tag{17}$$

Tied keys select the lower index first. This decoder always produces a feasible subset. Many key vectors represent the same placement, so convergence of a vector population does not necessarily indicate progress through different designs. Candidate clipping at zero and one can also create ties. The implementation records the number of distinct evaluated subsets as well as the total charged calls.

Every invocation of the objective consumes one unit of budget, including population initialization, repeats, and restarts. A cache stores deterministic training scores for already evaluated subsets, but repeated proposals remain charged. This contract avoids giving an algorithm extra search simply because its representation produces duplicates. It does not make objective calls equivalent to elapsed execution time: Proposal generation and cache hit rates differ. Runtime comparisons therefore require separate measurements on a specified deployment platform, which this study does not claim to provide.

An external best archive retains the best score encountered. Equal scores are resolved lexicographically by the sorted station indices in that archive; the search's internal acceptance rules retain their stated conventions. The archive ensures that a stochastic acceptance step cannot erase the returned best design. All methods maximize the same weighted timely coverage.

4.2 | Search Rules

Random search draws a uniformly distributed subset of the prescribed cardinality at each call. It is a useful baseline because it explores the decoded design space directly rather than drawing a continuous population whose geometry may matter. Repeated subsets are allowed and charged. In the smaller family, repeated random draws can cover much of the feasible space; the experiment should not be interpreted as a demanding large scale search merely because it uses many calls.

SA begins from a random subset. A proposal replaces one uniformly selected current station with one uniformly selected excluded station. For current score f and proposal score f', acceptance uses:

$$p_{accept} = \min\{1, \exp((f' - f)/\tau_t)\}. \tag{18}$$

The temperature decreases geometrically from 0.03 toward 0.0001 over the budget, using the charged call index. The first proposal uses its post-evaluation call index. This is a finite reference schedule, not an asymptotic cooling theorem. Neutral proposals are accepted. The external archive records the best placement throughout the trajectory.

DE maintains 20 key vectors. Three distinct vectors, excluding the current target, generate a proposal through,

$$v = x_{r_1} + 0.6(x_{r_2} - x_{r_3}). \tag{19}$$

The mutation is clipped to the unit cube before binomial crossover, whose probability is 0.8 with one forced coordinate. Trials replace targets only when their decoded training score strictly improves. Population updates occur sequentially, so later trials in the same sweep can use updated vectors. Strict replacement can preserve diversity on flat scores but can also leave a population trapped among equivalent decoded subsets.

PSO also uses 20 key vectors. With independent uniform coordinate vectors r_1 and r_2, its reference update is:

$$v_i \leftarrow 0.7v_i + 1.4r_1 \odot (p_i - x_i) + 1.4r_2 \odot (g - x_i),$$
$$x_i \leftarrow \mathrm{clip}_{[0,1]^n}(x_i + v_i). \quad (20)$$

The velocity is clipped coordinatewise to the interval from -0.25 to 0.25 before updating position. Initial velocities are zero. Personal bests change only on strict improvement; the global guide is the first highest scoring personal best in population order. Updates occur sequentially. These choices define the tested variant and can materially affect a random key placement problem.

The restarted swap method uses the same one station replacement proposal as SA, but accepts only strict improvements. After every 100 proposals, it draws and evaluates a new random subset, resets the proposal counter, and preserves the external archive. The restart is periodic, not triggered by an estimated convergence statistic. Its simple design tests whether local search and renewed starts explain performance that might otherwise be attributed to a more elaborate population mechanism.

4.3 | Deterministic and Exhaustive References

A deterministic triple lookahead baseline first evaluates every three station subset, chooses the first highest scoring triple in lexicographic enumeration, and then adds the best remaining station at each step. This avoids the zero utility of a one or two station seed under three station confirmation. It uses 229 calls for the smaller geometry and 2,084 for the larger geometry, below the respective stochastic budgets. It is not given an uncharged optimization phase or a submodular performance guarantee.

Complete enumeration separately evaluates every feasible full placement. The training gap in percentage points is:

$$G_N(S) = 100[\max_{|Q|=b} \hat{F}_N(Q) - \hat{F}_N(S)]. \quad (21)$$

This is an absolute difference on the unit interval, not a relative percentage of the optimum. Floating point equality within a tolerance of 10^{-12} defines a recorded optimum hit. Enumeration verifies optimality for the fixed finite training objective. It does not establish that the selected design maximizes the unknown population objective or performance in a real geographical region.

5 | Experimental Design and Reproducibility

5.1 | Geometry and Source Sampling

The six constructed planning cases cross two station geometries with three delay and outage regimes. They are transparent reference cases, not reconstructions of Istanbul, Switzerland, California, or any other operational network. The smaller geometry places 12 candidate stations on a four by three rectangular grid spanning 160 by 100 km and selects four stations. The larger geometry uses a six by four grid on the same rectangle and selects six. Indices increase first from west to east and then from south to north. The confirmation requirement is three stations in every case.

Fifteen potential source locations combine five east coordinates, 0, 40, 80, 120, and 160 km, with three north coordinates, 10, 50, and 90 km. All sources have depth 10 km. Before normalization, a source receives weight one plus two if its east coordinate is at most 40 km, plus one if its north coordinate equals 50 km. Each scenario independently samples a source from this discrete distribution. The rule creates a western concentration without representing a fitted hazard model, a recurrence rate, or a rupture catalog.

Nine recipients combine east coordinates 40, 100, and 160 km with north coordinates 15, 50, and 85 km. Their weights within each north row are 1, 2, and 4, normalized over all nine recipients. Thus the largest aggregate weight lies in the east, where source distances also tend to allow more time. The group analysis uses the three east coordinate bands and weights each of the three recipients within a band equally. This construction intentionally permits disagreement between the weighted objective and minimum group performance. *Fig. 2* displays the geometry and selected designs for the outage regime.

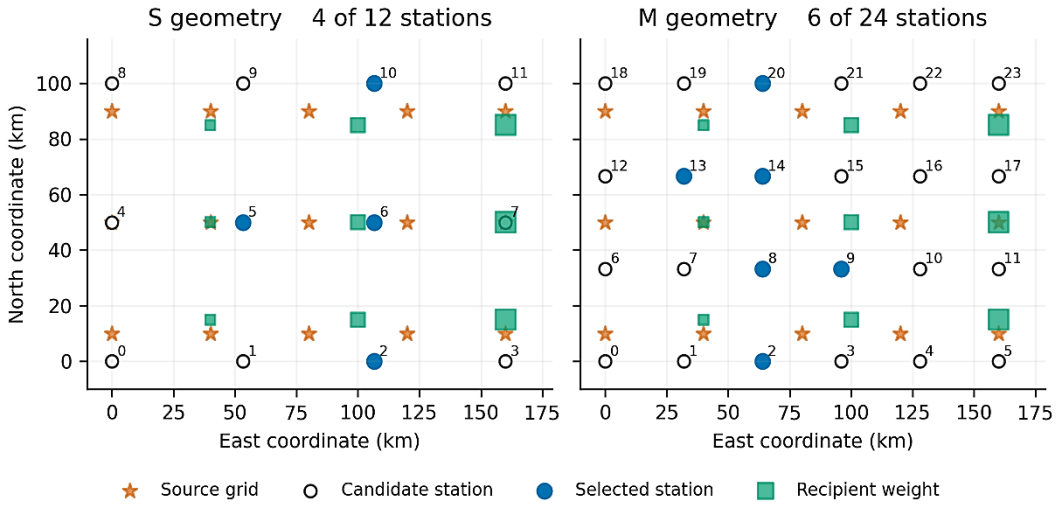


Fig. 2. Constructed planning geometry and training optimal placements for SD and MD.

Stars denote potential sources, open circles denote candidate stations, blue circles denote selected stations, and green square area represents recipient weight. The diagrams are coordinate models rather than maps of observed seismic infrastructure *(Fig. 2)*.

5.2 | Joint Delay and Outage Distributions

All cases use P and S velocities of 6 and 3.5 km/s, one second of station window plus local processing, one second of action time, and five seconds of required residual lead. Station telemetry has a lognormal component with log standard deviation 0.35. Its median is 0.35 s plus 1.2 s times one minus the station's east coordinate divided by 160 km. A western station can therefore be closer to the more heavily sampled sources while having a slower telemetry component. The values are explicit design assumptions chosen to expose this tradeoff.

A shared nonnegative shock is zero unless a Bernoulli event occurs; conditional on occurrence it is uniform between three and seven seconds. It is added to every station telemetry time and half of it is added to each recipient's downstream delay. The fast regime F uses occurrence probability 0.02, whereas the congested regime C uses 0.20. Regime D uses the same shock probability as C and introduces both a common western sector outage and independent station outages. The common outage occurs with probability 0.20 and disables stations whose east coordinate is at most 60 km. Each station independently also fails with probability 0.04. Failure means no usable report for that scenario, rather than a delayed recovery.

Central processing is uniform between 0.15 and 0.25 s. The recipient specific suffix combines a lognormal component with log standard deviation 0.45, medians 0.7, 1.5, and 3.0 s in the west, centre, and east, and a uniform component between 0.15 and 0.30 s. The latter represents routing overhead; the former represents delivery. Their sum also includes half the shared shock. Other random draws are independent except for the shared shock and sector outage. The experiment stores central delay and the combined recipient suffix; it does not infer a separate routing topology.

These distributions are not fitted to the WEA tests or to the published MyShake latency model. Those studies motivated measuring delivery as part of the chain, but they do not validate our chosen parameters. Changing the shock frequency, candidate geometry, confirmation requirement, or channel medians can change both the preferred placement and algorithm behavior. F and C differ in shock frequency; C and D add outages. These regime comparisons do not isolate correlation while keeping every marginal distribution unchanged.

Table 5. Fixed parameters and computation budgets for the reference experiment.

Component	Specification
Candidate selection	S 12 choose 4; M 24 choose 6; three station confirmation
Sources and recipients	15 sources at depth 10 km; nine recipients; all pairs alert eligible.
Wave speeds and margins	6 and 3.5 km/s; station window 1 s; action 1 s; required residual lead 5 s
Station telemetry	Median 0.35–1.55 s by east coordinate; log standard deviation 0.35.
Shared shock	Probability 0.02 in F and 0.20 in C or D; positive duration uniform 3–7 s.
Outages in D	Common western sector probability 0.20; independent station probability 0.04.
Recipient suffix	Lognormal medians 0.7 1.5 3.0 s; log standard deviation 0.45; routing uniform 0.15–0.30 s; half shared shock.
Central processing	Uniform 0.15–0.25 s; zero extra event path optimization in base cases.
Scenario and run counts	256 training and 10,000 assessment vectors per case; 30 seeds per method
Budgets and populations	1,200 calls in S and 2,400 in M; DE and PSO population 20.
SA and swap	SA temperature 0.03 to 0.0001; swap restart after 100 proposals.
DE and PSO	DE scale 0.6 crossover 0.8; PSO inertia 0.7 and attraction 1.4 each; velocity limit 0.25.

5.3 | Search and Assessment Protocol

Each case has 256 training scenario vectors and 10,000 independently generated assessment vectors. Training seeds are 8100, 8101, and 8102 for SF, SC, and SD, and 8200, 8201, and 8202 for MF, MC, and MD. Assessment seeds are respectively 9100–9102 and 9200–9202. The same stored training and assessment arrays are used by all methods within a case. This common scenario design reduces comparison noise without implying that recipient observations are independent.

Each stochastic method runs with seeds 1000 through 1029. The smaller and larger cases allow 1,200 and 2,400 charged calls per run, including the initial population. The study therefore comprises 900 optimizer runs and 1,620,000 charged calls. The distinct feasible families contain 495 and 134,596 placements; enumerating each of three regimes gives 405,273 full designs. Exact enumeration is additional reference computation and is not part of a heuristic's budget.

The training scenario set is fixed within a case, so the 30 optimizer seeds measure search variability conditional on that set. They are not 30 independent earthquake populations. Assessment uses common independent scenario arrays, and variation across optimizer seeds is summarized descriptively. We report medians, interquartile ranges, and exact optimum hit counts without treating the six designed cases as a random sample of the world's warning networks. No omnibus significance test or universal algorithm ranking is inferred.

For the added delay and group analyses, the placement returned by SA seed 1000 is fixed in advance in every case. The same assessment vectors are rescored after adding 0, 0.25, 0.5, 1, 2, 4, 8, and 12 s to the event path. No placement is retuned at each delay value. This isolates the response of a chosen design's margins, rather than attributing changes to a simultaneous search. Monte Carlo standard errors use the standard deviation of the complete recipient weighted scenario scores divided by the square root of 10,000. They describe simulation error conditional on the selected placement and the assumed distributions, not uncertainty about the physical model.

5.4 | Verification and Available Materials

An independent evaluator sorts all selected report times instead of using the optimizer's partial order statistic operation. It checks the saved score of every enumerated design and rescores every returned placement on training and assessment arrays. All 405,273 enumeration records agree with the independent evaluator, and all 900 returned placements satisfy cardinality and budget checks. Thirty representative runs, one seed for every method and case, are replayed to verify the saved placement and objective. Delay sensitivity curves are checked for the monotonicity required by *Proposition 1*.

A separate availability experiment checks three of four confirmation with independent station availability 0.9 and no timing restriction. The exact failure probability is 0.0523. In 200,000 draws with seed 62026, the observed frequency is 0.05221, a discrepancy of approximately −0.18 standard errors. This verifies a simple outage boundary case. It does not validate the sector outage distribution or the joint physical timing model.

The companion package contains the complete generator, search implementations, enumeration code, verification code, all scenario arrays, every run and trace, all enumeration scores, the source extraction and bibliographic verification ledgers, and figure sources. A captured environment file records package versions. The core benchmark uses Python and NumPy and requires no private waveform data or external service. Numerical reproduction means rerunning the stated reference experiment; it does not reproduce the field studies reviewed above.

6 | Results

6.1 | Training Optima and Search Variability

Table 6 reports the enumerated training optimum and the number of runs that reach it. SA and restarted swap attain all three small case optima in every run, giving 90 hits from 90 runs for each method. DE attains 86 of those 90 optima, random search attains 79, and PSO attains 50. The small design family is easy enough for several methods that median final scores alone would conceal differences in failure frequency. A zero median gap is compatible with poor runs away from the optimum.

Table 6. Enumerated training optima and optimum hits out of 30 runs per method.

Case	Optimum	RS	SA	DE	PSO	SW
SF	0.619048	24	30	30	18	30
SC	0.559896	28	30	29	15	30
SD	0.522693	27	30	27	17	30
MF	0.680060	0	9	1	0	16
MC	0.627790	1	12	5	1	10
MD	0.632626	0	30	19	5	30

Note: RS is random search; SW is restarted swap.

In the larger cases, SA attains 51 of 90 optima, DE 25, PSO six, restarted swap 56, and random search one. These hit rates describe the specified budgets and parameter settings. They do not establish that one algorithm family dominates after tuning or under a different encoding. In MF, restarted swap has the most hits; in MC, SA has the most; in MD, SA and restarted swap both attain the optimum in every run. *Fig. 3* shows the distributions of absolute training gaps rather than relying only on those counts.

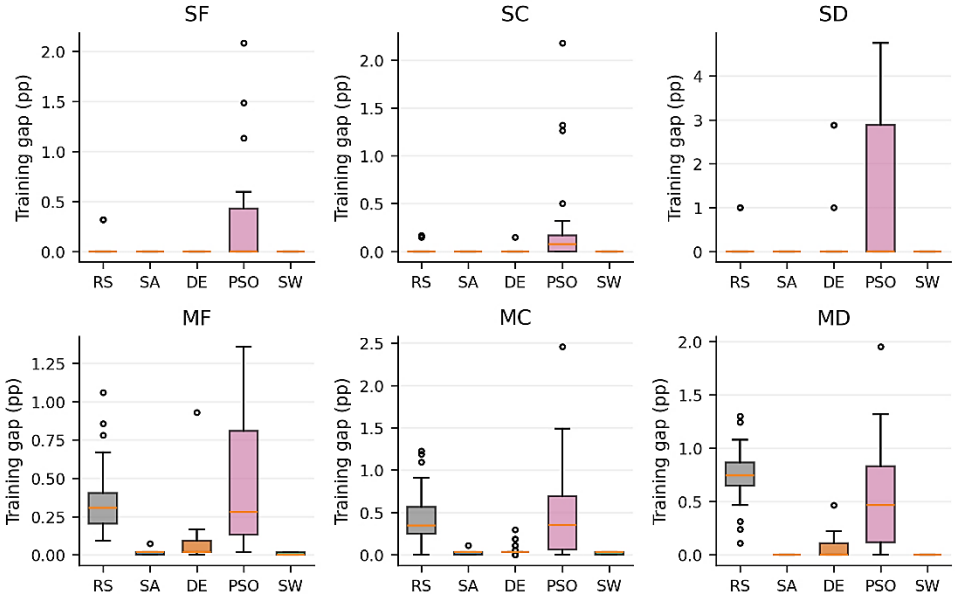


Fig. 3. Training gap distributions over 30 seeds per case and method, measured in percentage points below complete enumeration.

RS denotes random search and SW denotes restarted swap. Boxes show the interquartile range, with medians, conventional 1.5 interquartile range whiskers, and outliers. Collapsed boxes indicate repeated equal scores, not missing results (*Fig. 3*).

The largest median training gap among all method–case combinations is 0.7440 percentage points, for random search in MD. The corresponding PSO median gap is 0.4650 points. In MF and MC, the largest median gaps are 0.3069 and 0.3534 points. The annealing and swap trajectories often reduce gaps quickly, but their final distributions remain case dependent. *Fig. 4* provides budget based traces, including initialization. It should not be read as an elapsed time comparison or a guarantee of monotone progress in every population member; the curves refer to the protected best archive.

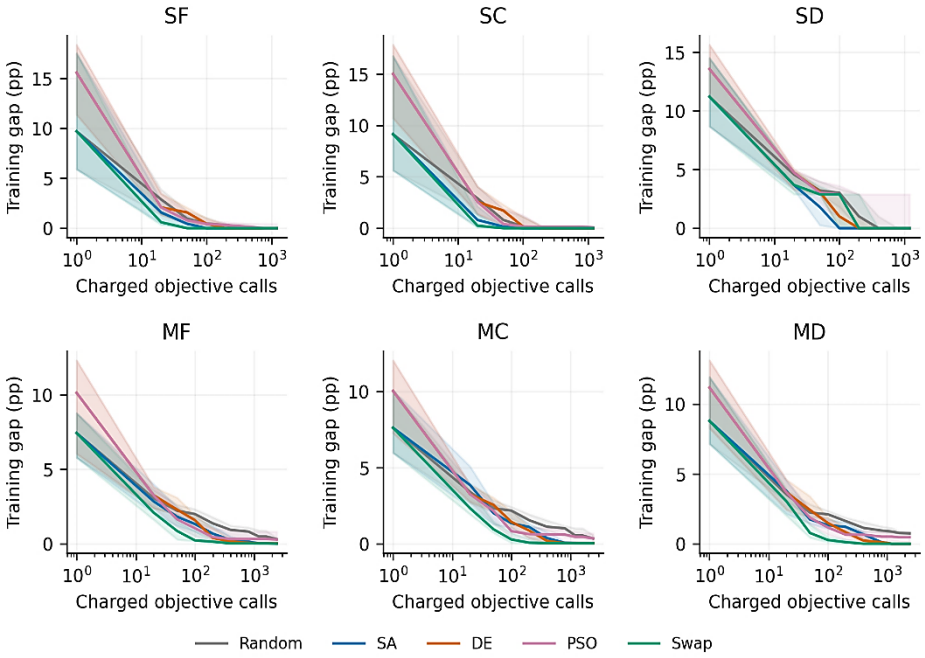


Fig. 4. Median archived training gap against charged objective calls on a logarithmic horizontal axis.

Shading denotes the interquartile range across 30 seeds. All proposal evaluations, including repeated decoded designs and restarts, are charged. The curves characterize these reference implementations under a common objective (*Fig. 4*).

6.2 | Assessment on Independent Scenarios

Table 7 reports median assessment scores across optimizer runs and the deterministic lookahead baseline. A training optimum does not guarantee the best assessment score. In SC, for example, the lookahead placement has a training score of 0.558408, slightly below the optimum of 0.559896, but its assessment score of 0.563419 is slightly above the score of 0.563176 obtained by the lexicographically preferred training optimum. The difference is small and is not evidence that lookahead is generally superior. It illustrates the distinction between training optimization and performance under a fresh scenario sample.

Table 7. Median assessment timely coverage in percent across 30 optimizer runs, with the deterministic triple lookahead score.

Case	RS	SA	DE	PSO	SW	Lookahead
SF	59.6048	59.6048	59.6048	59.6048	59.6048	59.6048
SC	56.3176	56.3176	56.3176	56.3176	56.3176	56.3419
SD	49.1810	49.1810	49.1810	49.1810	49.1810	49.1810
MF	67.8136	68.0657	68.0657	67.8229	68.1014	68.0657
MC	64.9236	65.3057	65.2533	64.9757	65.3057	65.2010
MD	63.1500	63.8757	63.8757	63.7590	63.8757	63.8757

Note: Every case uses 10,000 common assessment scenarios.

On the larger geometries, the median assessment differences between the stronger solvers are also small. Restarted swap has a median of 68.1014% in MF, while SA and DE both have medians of 68.0657%. In MC, SA and restarted swap both have 65.3057%, compared with 65.2533% for DE. Such differences must be assessed against both search variability and scenario sampling uncertainty. A shared assessment set makes paired comparison possible, but it does not remove uncertainty about the assumed source and communication model.

The minimum group measure sometimes distinguishes placements with nearly equal aggregate scores. In MF, the median minimum group score is approximately 43.89% for SA but 38.56% for restarted swap, despite the latter's slightly larger median aggregate score. The weighted objective gives more importance to eastern recipients, and it never constrained minimum group service. The result is therefore consistent with its design. It motivates explicit group reporting rather than a claim that maximizing aggregate coverage automatically distributes warning opportunities evenly.

6.3 | Added Event Path Delay

For the prespecified SA placements, four additional seconds reduce assessment coverage from 59.6048% to 48.8924% in SF, from 56.3176% to 46.5186% in SC, and from 49.1810% to 39.0886% in SD. The corresponding larger case reductions are from 68.0657% to 58.6767%, from 65.3057% to 55.9752%, and from 63.8757% to 54.2671%. The absolute losses range from 9.3305 to 10.7124 percentage points. *Fig. 5* shows the full response to added delay.

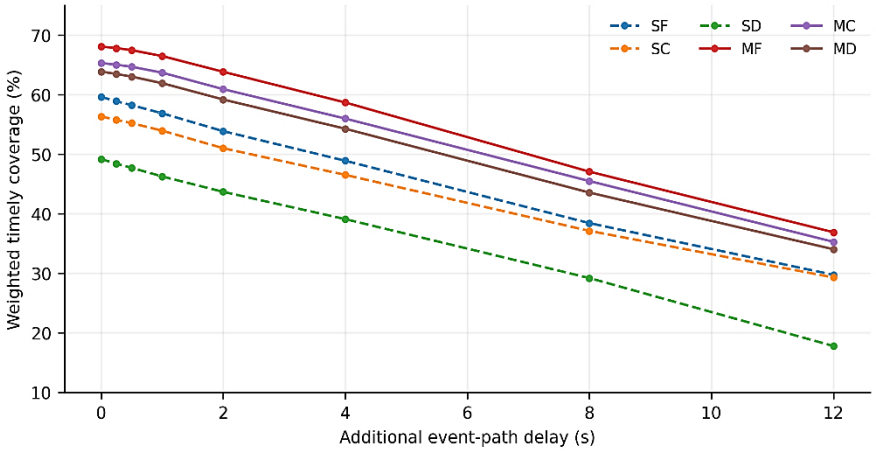


Fig. 5. Assessment coverage after adding a constant delay to the event path of the SA seed 1000 placement in each case.

The design and assessment scenarios remain fixed. Dashed curves denote small geometries and solid curves denote larger geometries. These curves are sensitivity calculations under assumed distributions, not measured delivery performance (*Fig. 5*).

The baseline Monte Carlo standard errors for these selected placements range from 0.263 to 0.309 percentage points. A four second loss is therefore large relative to their numerical integration error. It is also much larger than the observed median training gaps between the tested solvers. The comparison is a scale comparison across the reported results, not a controlled estimate of replacing one solver with a communication upgrade. A field intervention could change several distributions and would require its own evidence.

The added delay can represent computation inserted after confirmation, but it could equally represent a serial delivery or action delay. The sensitivity calculation supplies no estimate of the actual runtime of SA, DE, or PSO on an edge device. Its implication is architectural: completing expensive placement search offline avoids consuming this event path margin. If a search is proposed during a real event, its measured runtime distribution must enter the same calculation together with its effect on the alert decision.

6.4 | Group Differences and Failed Confirmation

Fig. 6 presents coordinate group coverage for the same prespecified placements. Eastern groups receive more timely warnings in every case, reflecting their source distances, selected layouts, and channel assumptions. In SF the three scores are 26.35%, 45.00%, and 75.22%; in MF they are 43.89%, 63.20%, and 76.54%. These differences arise in a constructed geometry. They are not socioeconomic estimates or evidence about access to a particular national warning service.

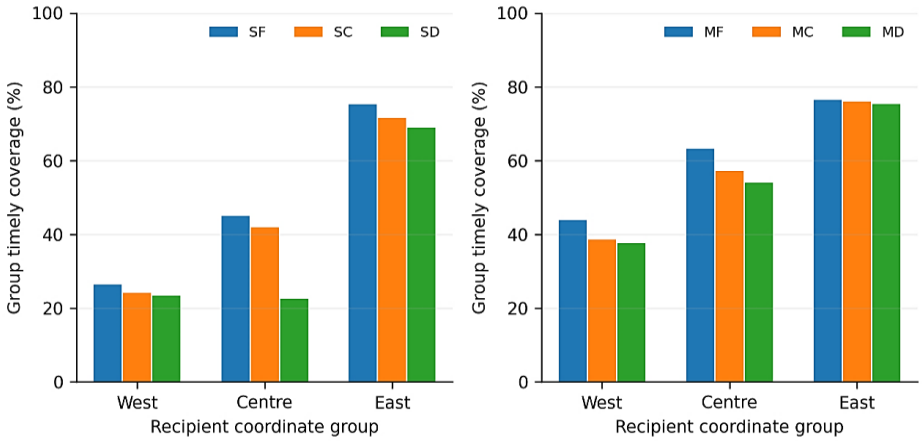


Fig. 6. Group timely coverage for the SA seed 1000 placements.

Each coordinate band averages its three recipients equally. The optimization objective uses unequal recipient weights and differs from these group averages. The figure reports a distributional diagnostic without imposing or verifying an equity policy (*Fig. 6*).

In SD, the selected placement fails to obtain three station reports in 3.00% of assessment scenarios; in MD, the frequency is 0.01%. Confirmation never fails in F or C because those regimes contain delays but no outages. Larger sensor redundancy can therefore change the outage response within this model, even though both sizes require three reports. Failed confirmation and late finite alerts remain distinct. Replacing infinite report times with a large but finite penalty without checking the deadline could blur that distinction and incorrectly reward an unavailable network.

7 | Discussion

7.1 | What a Latency Based Comparison Establishes

The literature and the benchmark support a common evaluation principle: an optimization result must be tied to the event time at which its output becomes usable. For placement, this means testing the chosen geometry under source locations, confirmation rules, communication behavior, and recipient deadlines. For phase picking, it means reporting the observation window and causal availability of the input as well as numerical onset error. For threshold calibration, it means describing the tradeoff between precautionary alerts and missed or late warnings. For dissemination, it means measuring receipt at the client rather than stopping at message issuance.

These requirements do not make every study responsible for the full warning chain. A phase picker study can validly address picking accuracy; a hazard informed placement study can validly analyze a specified system level objective. The problem arises when a result is transferred beyond that scope. In particular, the speed of a neural network forward pass cannot be substituted for warning latency, and a planning optimizer's objective value cannot establish a real time execution guarantee. *Table 8* summarizes the evidence needed for such transfers.

Table 8. Evidence required to connect component results to a deployable warning claim.

Component Claim	Evidence Needed	Error Prevented
Better station placement	Hazard or replay scenarios; complete decision rule; site feasibility; independent assessment	Coverage score mistaken for field benefit
Faster detection	Causal windows stride preprocessing and inference; timestamped decisions	Inference duration substituted for origin to decision time
Improved alert threshold	Intensity target forecast uncertainty false and missed alert costs	Timing success substituted for classification quality
Low event path latency	Joint stage timestamps queue load client receipt and missingness	Unjustified sum of marginal tail quantiles
Efficient online optimization	Wall time and energy on intended hardware and load; deadline and fallback	Offline call budget substituted for real time guarantee
Equitable warning	Defined groups eligibility access and coverage measures; appropriate linked data	Weighted average substituted for distributional service

The formal model helps identify precisely where an extension belongs. A new picker changes report readiness and potentially trigger correctness. A finite fault or ground motion model changes the confirmation and recipient eligibility processes. A dissemination policy changes delivery and perhaps its dependence on demand. A larger station set may alter congestion, violating the fixed readiness assumption behind *Proposition*

1. Modelling those effects explicitly is more informative than assigning a generic robustness label to an optimizer.

7.2 | Interpreting the Algorithm Results

The benchmark does not support choosing a solver from its name alone. The repeated swap heuristic performs well without a complex population representation, and random search often reaches the small case optimum. PSO is more variable under the chosen random key decoder and update rules. These findings identify useful baselines and implementation questions; they do not establish a theorem about the algorithms' relative merit. The study did not tune hyperparameters against a separate case development set, compare all established variants, or sample a broad class of network design problems.

Complete enumeration is particularly valuable here because it exposes how small the remaining training gaps are. Without it, the best observed heuristic score could be mistaken for a strong optimization frontier. With it, the reader can see both successful searches and the limited scale of this benchmark. Larger regional candidate sets would make exhaustive search impractical and would require exact bounds, mixed integer formulations, decomposition, or stronger heuristic validation. The present experiment supplies a reference point for checking such implementations rather than claiming to solve large scale regional network planning.

The concentration result also discourages oversolving a poorly specified sample. Reducing the training gap by a tiny amount may have little value if source probabilities, telemetry behavior, and ground motion deadlines are uncertain. The training sample contains only 256 vectors, and all 30 search seeds in a case share it. Replicating optimizer seeds measures one kind of variability; replicating scenario generation, changing hazard assumptions, and testing a new region address different kinds. Reporting them separately prevents a large count of runs from obscuring a narrow evidence base.

7.3 | Operational Interpretation and Distributional Priorities

The four second sensitivity calculation gives a practical reason to keep placement optimization outside the event path. It does not imply that four seconds is an acceptable budget or a universal failure threshold. The useful budget depends on how many recipients lie near the margin threshold, the action they intend to take, and the relevant shaking definition. A recipient with a late alert cannot be made timely by improving an unrelated coverage metric. Conversely, a small latency reduction can matter substantially when a large share of scenarios lies just below the action deadline.

Tail performance must be assessed using complete joint observations. Common infrastructure can couple sensing and delivery delays, and regional outages can remove several stations simultaneously. Our shared shock is a transparent way to illustrate that coupling, not an estimate of its occurrence in any service. Operational analysis should use synchronized station, processing, gateway, and client timestamps, with uncertainty about clock alignment and missing telemetry made explicit. Rare tail statements require enough independent events or a justified model; a high quantile extracted from a small or selectively observed sample can be unstable.

The group analysis reinforces the need for an explicit objective. Population weighted coverage, equal geographical service, and protection of critical facilities can prefer different networks. Choosing weights is a policy decision informed by hazard and action consequences. It is not resolved by metaheuristic search. The coordinate bands used here show how averages can hide differences, while remaining separate from claims about income, disability, language, handset ownership, or connectivity. Such claims would need appropriate demographic and delivery data and a justified linkage to the alert outcome.

7.4 | Limitations and Next Validation Steps

The principal limitation is model validity. The experiment uses point sources at a fixed depth, constant velocities, always eligible recipients, and correct qualifying station reports. It excludes source growth, intensity forecast errors, false triggers, magnitude saturation, wave propagation heterogeneity, communication queues,

recovery dynamics, routing topology, and adaptive warning updates. Its lognormal and shock parameters are chosen assumptions. The results verify arithmetic, search behavior, and timing consequences within that model; they do not validate those assumptions against an operating network.

The review is selective and cannot supply a pooled quantitative ranking of metaheuristics in EEW. Some studies emphasize waveform inference, some optimize station placement, and others examine alert delivery. Their metrics and evidence answer different questions. The inclusion ledger makes our selection inspectable, while a future systematic review would need a preregistered or fully specified search and screening process, a broader extraction protocol, and explicit treatment of noncomparable outcomes. We do not reconstruct absent screening counts or statistical summaries from the literature.

A useful next empirical study would combine a regional hazard or waveform replay data set with measured station availability and synchronized delivery records. Calibration and assessment should be separated by event and, where feasible, by region or time period. Candidate sites should reflect access, power, cost, and communication constraints. A proposed event path optimizer would additionally require measured runtime and energy distributions at the intended load, including acquisition, buffering, decision updates, and fallback behavior. These requirements follow from the model's omitted mechanisms and the scope of the reviewed evidence.

8 | Conclusion

Latency aware evaluation connects station selection to the recipient deadline that gives an earthquake warning its value. The reference model makes concurrent report arrivals, multiple station confirmation, serial delivery, action duration, and outages explicit. Its mathematical properties show why additional stations help only under stated exogenous delay assumptions, why multi-station confirmation can defeat submodularity, and why eliminating training optimization error does not eliminate sampling error.

The reproducible study provides exact training references for six finite cases and compares five search methods across 900 runs. Several simple methods reach the small case optima reliably, while the larger cases expose differences in repeatability. Adding four seconds to prespecified placements produces losses of approximately nine to eleven percentage points in timely coverage, larger than the median solver gaps observed here. That finding supports evaluating algorithms through their full timing consequences and completing planning search before an event. Its magnitude remains conditional on the reference geometry and distributions. Field validation requires observed waveforms, hazard, telemetry, and delivery records matched to the intended warning decision.

Data and Code Availability

The accompanying research package provides all generated scenario arrays, search and enumeration outputs, executable Python code, verification results, bibliographic ledgers, and editable figure sources. No private event records were used. The package is the reproducibility artifact for the new reference experiment; no public archive identifier is asserted. A permanent archive may be assigned when the authors deposit the final materials.

Ethics and Scope of Use

The computational study used constructed coordinates and generated scenario vectors, with no recruited participants, personal records, or operational alert transmission. It is intended for methodological analysis. Deployment decisions require independent seismological, engineering, and operational validation of the intended network and protective action.

References

[1] Allen, R. M., & Melgar, D. (2019). Earthquake early warning: Advances, scientific challenges, and societal needs. *Annual Review of Earth and Planetary Sciences*, *47*(1), 361–388. https://doi.org/10.1146/annurev-earth-053018-060457

[2] Minson, S. E., Meier, M. A., Baltay, A. S., Hanks, T. C., & Cochran, E. S. (2018). The limits of earthquake early warning: Timeliness of ground motion estimates. *Science Advances*, *4*(3), eaaq0504. https://doi.org/10.1126/sciadv.aaq0504

[3] Meier, M. A., Kodera, Y., Böse, M., Chung, A., Hoshiba, M., Cochran, E., …, & Heaton, T. (2020). How often can earthquake early warning systems alert sites with high-intensity ground motion? *Journal of Geophysical Research: Solid Earth*, *125*(2), e2019JB017718. https://doi.org/10.1029/2019JB017718

[4] Oth, A., Böse, M., Wenzel, F., Köhler, N., & Erdik, M. (2010). Evaluation and optimization of seismic networks and algorithms for earthquake early warning the case of Istanbul (Turkey). *Journal of Geophysical Research: Solid Earth*, *115*(B10), 1–22. https://doi.org/10.1029/2010JB007447

[5] Stankiewicz, J., Bindi, D., Oth, A., & Parolai, S. (2013). Designing efficient earthquake early warning systems: case study of Almaty, Kazakhstan. *Journal of Seismology*, *17*(4), 1125–1137. https://doi.org/10.1007/s10950-013-9381-4

[6] Böse, M., Papadopoulos, A. N., Danciu, L., Clinton, J. F., & Wiemer, S. (2022). Loss-based performance assessment and seismic network optimization for earthquake early warning. *Bulletin of the Seismological Society of America*, *112*(3), 1662–1677. https://doi.org/10.1785/0120210298

[7] Najdahmadi, B., Pilz, M., Bindi, D., NT Razafindrakoto, H., Oth, A., & Cotton, F. (2023). Hazard-informed optimization of seismic networks for earthquake early warning—the case of the Lower Rhine Embayment (Western Germany). *Journal of Seismology*, *27*(2), 261–277. https://doi.org/10.1007/s10950-023-10133-z

[8] Minson, S. E., Baltay, A. S., Cochran, E. S., Hanks, T. C., Page, M. T., McBride, S. K., ..., & Meier, M. A. (2019). The limits of earthquake early warning accuracy and best alerting strategy. *Scientific reports*, *9*(1), 2478. https://doi.org/10.1038/s41598-019-39384-y

[9] Saunders, J. K., Minson, S. E., & Baltay, A. S. (2022). How low should we alert? Quantifying intensity threshold alerting strategies for earthquake early warning in the United States. *Earth's Future*, *10*(3), e2021EF002515. https://doi.org/10.1029/2021EF002515

[10] Minson, S. E., Cochran, E. S., Saunders, J. K., McBride, S. K., Wu, S., Baltay, A. S., & Milner, K. R. (2022). What to expect when you are expecting earthquake early warning. *Geophysical Journal International*, *231*(2), 1386–1403. https://doi.org/10.1093/gji/ggac246

[11] Kirkpatrick, S., Gelatt Jr, C. D., & Vecchi, M. P. (1983). Optimization by simulated annealing. *Science*, *220*(4598), 671–680. https://doi.org/10.1126/science.220.4598.671

[12] Storn, R., & Price, K. (1997). Differential evolution a simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization*, *11*(4), 341–359. https://doi.org/10.1023/A:1008202821328

[13] Kennedy, J., & Eberhart, R. (1995). Particle swarm optimization. *Proceedings of ICNN'95-international conference on neural networks* (pp. 1942-1948). IEEE. https://doi.org/10.1109/ICNN.1995.488968

[14] Bean, J. C. (1994). Genetic algorithms and random keys for sequencing and optimization. *ORSA Journal on Computing*, *6*(2), 154–160. https://doi.org/10.1287/ijoc.6.2.154

[15] Nemhauser, G. L., Wolsey, L. A., & Fisher, M. L. (1978). An analysis of approximations for maximizing submodular set functions—I. *Mathematical Programming*, *14*(1), 265–294. https://doi.org/10.1007/BF01588971

[16] Böse, M., Smith, D. E., Felizardo, C., Meier, M. A., Heaton, T. H., & Clinton, J. F. (2018). FinDer v. 2: Improved real-time ground-motion predictions for M2-M9 with seismic finite-source characterization. *Geophysical Journal International*, *212*(1), 725–742. https://doi.org/10.1093/gji/ggx430

[17] Kodera, Y., Yamada, Y., Hirano, K., Tamaribuchi, K., Adachi, S., Hayashimoto, N., ..., & Hoshiba, M. (2018). The propagation of local undamped motion (PLUM) method: A simple and robust seismic

wavefield estimation approach for earthquake early warning. *Bulletin of the Seismological Society of America*, *108*(2), 983-1003. https://doi.org/10.1785/0120170085

[18] Zhu, W., & Beroza, G. C. (2019). PhaseNet: A deep-neural-network-based seismic arrival-time picking method. *Geophysical Journal International*, *216*(1), 261–273. https://doi.org/10.1093/gji/ggy423

[19] Mousavi, S. M., Zhu, W., Sheng, Y., & Beroza, G. C. (2019). CRED: A deep residual network of convolutional and recurrent units for earthquake signal detection. *Scientific Reports*, *9*(1), 10267. https://doi.org/10.1038/s41598-019-45748-1

[20] Mousavi, S. M., Ellsworth, W. L., Zhu, W., Chuang, L. Y., & Beroza, G. C. (2020). Earthquake transformer an attentive deep-learning model for simultaneous earthquake detection and phase picking. *Nature Communications*, *11*(1), 3952. https://doi.org/10.1038/s41467-020-17591-w

[21] Münchmeyer, J., Bindi, D., Leser, U., & Tilmann, F. (2021). The transformer earthquake alerting model: A new versatile approach to earthquake early warning. *Geophysical Journal International*, *225*(1), 646–656. https://doi.org/10.1093/gji/ggaa609

[22] Woollam, J., Münchmeyer, J., Tilmann, F., Rietbrock, A., Lange, D., Bornstein, T., ..., & Soto, H. (2022). SeisBench—A toolbox for machine learning in seismology. *Seismological Society of America*, *93*(3), 1695-1709. https://doi.org/10.1785/0220210324

[23] Münchmeyer, J., Woollam, J., Rietbrock, A., Tilmann, F., Lange, D., Bornstein, T., ..., & Soto, H. (2022). Which picker fits my data? A quantitative evaluation of deep learning based seismic pickers. *Journal of Geophysical Research: Solid Earth*, *127*(1), e2021JB023499. https://doi.org/10.1029/2021JB023499

[24] Ravishan, D., Prasanna, R., Herath, P., & Doyle, E. E. H. (2026). Lightweight convolutional neural network for real-time earthquake P-wave detection on edge devices in New Zealand. *Scientific Reports*, *16*(1), 11536. https://doi.org/10.1038/s41598-026-42568-y

[25] Xiao, H., Walter, J. I., & Ogwari, P. (2026). Edge seismic phase picking: Model distillation for multimodal early-warning systems. *PNAS Nexus*, *5*(10), 322. https://doi.org/10.1093/pnasnexus/pgag322

[26] Gou, Y., Nof, R. N., Pardini, B., & Allen, R. M. (2026). Integrating fiber-optic seismic arrays into earthquake early warning systems with the dEPIC framework. *Scientific Reports*, *16*(1), 960. https://doi.org/10.1038/s41598-025-30568-3

[27] Ahn, J. K., Park, E., Kim, B., Hwang, E. H., & Hong, S. (2023). Stable operation process of earthquake early warning system based on machine learning: Trial test and management perspective. *Frontiers in Earth Science*, *11*, 1157742. https://doi.org/10.3389/feart.2023.1157742

[28] Lux, A. I., Smith, D., Böse, M., McGuire, J. J., Saunders, J. K., Huynh, M., ..., & Toomey, D. (2024). Status and performance of the ShakeAlert earthquake early warning system: 2019–2023. *Bulletin of the Seismological Society of America*, *114*(6), 3041-3062. https://doi.org/10.1785/0120230259

[29] McBride, S. K., Sumy, D. F., Llenos, A. L., Parker, G. A., McGuire, J., Saunders, J. K., ... & de Groot, R. (2023). Latency and geofence testing of wireless emergency alerts intended for the ShakeAlert® earthquake early warning system for the West Coast of the United States of America. *Safety Science*, *157*, 105898. https://doi.org/10.1016/j.ssci.2022.105898

[30] Allen, R. M., Barski, A., Berman, M., Bosch, R., Cho, Y., Jiang, X. S., ..., & Wimpey, G. (2025). Global earthquake detection and warning using Android phones. *Science*, *389*(6757), 254-259. https://doi.org/10.1126/science.ads4779

[31] Marcou, S., Lux, A. I., Akimov, A., Williamson, A. L., & Allen, R. M. (2025). Exploring end-to-end earthquake early warning performance in large earthquakes using the February 2023 Kahramanmaraş, Türkiye sequence. *Scientific Reports*, *16*(1), 327. https://doi.org/10.1038/s41598-025-29755-z%0A%0A

[32] Kleywegt, A. J., Shapiro, A., & Homem-de-Mello, T. (2002). The sample average approximation method for stochastic discrete optimization. *SIAM Journal on Optimization*, *12*(2), 479–502. https://doi.org/10.1137/S1052623499363220

[33] Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, *58*(301), 13–30. https://doi.org/10.1080/01621459.1963.10500830