

Candidate-network breadth alters the value of response-based donor selection for coastal aquaculture temperature forecasting

Thi Phuong Hang Nguyen

Chonnam National University

Hieyong Jeong

h.jeong@jnu.ac.kr

Chonnam National University

Article

Keywords: aquaculture forecasting, donor selection, monitoring networks, transfer learning, water temperature, response similarity

Posted Date: September 25th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10887658/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Candidate-network breadth alters the value of response-based donor selection for coastal aquaculture temperature forecasting

Hang Thi Phuong Nguyen¹ and Hieyong Jeong^{1,*}

¹Department of Artificial Intelligence Convergence, Chonnam National University,
Gwangju, Republic of Korea

*Corresponding author: h.jeong@jnu.ac.kr

Abstract

Forecasting water temperature at aquaculture sites with short, fragmented records often requires pooling information from surrounding monitoring stations. We tested whether the benefit of formal donor selection changes as the eligible candidate domain broadens and whether candidate-response dispersion can serve as an exploratory screening diagnostic. Candidate-domain radius was varied from 60 to 200 km around five data-sparse aquaculture sites while candidate count, donor-window budget, forecast architecture, chronology, validation data and retraining seeds were fixed. Across 12 independently sampled candidate-set blocks, selection advantage increased by 0.104 °C per tenfold increase in block span (95% confidence interval 0.021–0.187; $p = 0.019$), equivalent to 0.031 °C per doubling. Selected-pool error showed no systematic trend, whereas random-pool error increased. Maximum response similarity remained nearly constant (0.869–0.882), while the

minimum fell from 0.520 to 0.355. Wider domains therefore increased exposure to poorly matched donors rather than improving the best available donor. Selection showed no clear advantage in the most compact domain. Response heterogeneity co-varied with breadth in this network and remains an exploratory diagnostic rather than an independently established cause.

Keywords: aquaculture forecasting; donor selection; monitoring networks; transfer learning; water temperature; response similarity

1 Introduction

Aquatic foods support food security and livelihoods, and aquaculture is expected to remain an important component of future food systems [1, 2]. Korea has a mature seaweed-cultivation sector in which laver is an economically important crop grown in shallow coastal waters [3]. Water temperature and dissolved oxygen (DO) are operationally important because thermal extremes and oxygen depletion can disrupt coastal ecosystems and aquaculture production [4]. Forecasts issued far enough in advance can support intensified monitoring, harvesting, seeding and other risk-reduction decisions, although the present study does not measure mortality or economic loss directly.

The sites where forecasts are most useful are not always the sites with the best observations. Coastal stations begin operation at different times, measure different variables and contain maintenance or communication gaps. In the network studied here, the five aquaculture-adjacent target sites have short, fragmented temperature records and no target-site DO record, whereas surrounding stations have longer histories. Existing coastal forecasting studies have improved model architecture and accuracy for available observations [5–7], but an operational programme must also decide how information should be borrowed across an uneven monitoring network. Here, a *donor* is an existing monitoring station whose historical windows are pooled during model training; it is not a proposed location for a new sensor. The endpoint evaluated here is target-site water temperature. DO motivates the broader aquaculture-monitoring problem but cannot be

externally validated at these five targets. The study concerns data-sparse targets with enough historical observations to estimate a response profile, not truly ungauged sites with no target history. The study network and nested eligibility domains are shown in Fig. 1.

A practical research sequence follows from this observation. A forecasting pipeline must first retain locally valid windows from globally fragmented records instead of discarding an entire station because its history is incomplete. Once multi-site pooling is feasible, donor-pool construction becomes a separate decision. An earlier stage of this research programme, reported as a preprint, evaluated donor selection within a fixed coastal monitoring network and identified forcing-response similarity as a useful basis for constructing budget-matched training pools [8]. That analysis held the eligible candidate network fixed. In practice, expanding the area from which donor records are considered changes the breadth of the candidate set and may also change its response heterogeneity. It therefore remains unclear whether formal donor selection is beneficial across candidate networks of different extents.

Transfer-learning theory predicts that sources help when their input–output relations align with the target and can cause negative transfer when they do not [9, 10]. Hydrological regionalization studies likewise show that donor criterion, donor count and network structure affect skill [11–13]. Those comparisons generally treat the eligible network as given. Sensor-placement research typically optimizes locations within a prespecified eligible domain [14, 15], leaving the effect of changing that domain less directly examined.

A wider eligible domain can act through two competing mechanisms. It may reveal donors whose response to atmospheric and tidal forcing is better matched to the target, in which case the carefully selected pool should improve as the search widens. Alternatively, the best-matched donors may already be present nearby, while wider domains mainly add poorly matched candidates. The selected pool would then remain stable, but indiscriminate pooling would become riskier. The two mechanisms make different observable predictions: better-donor discovery requires improvement of the upper similarity tail and selected-pool error, whereas poor-donor avoidance requires a lengthening lower tail and degradation of random pools.

We distinguish these mechanisms through a controlled candidate-domain sweep. Candidate-domain radius is varied while the number of candidates per block, donor-window budget, forecast model, chronology, validation data and retraining seeds are held fixed. Our primary objective is to test whether donor-selection advantage changes as the eligible domain broadens. Our second objective is to determine whether any increase arises from better-donor discovery or from avoidance of poorly matched donors. Finally, we explore, without treating it as confirmatory, whether candidate-response dispersion can screen networks in which formal selection is worth evaluating. The candidate-set block, rather than the retraining seed or target site, is treated as the primary inferential unit. The fixed forecasting information flow and controlled pool comparison are summarized in Fig. 2.

2 Results

Wider domains add lower-similarity candidates while the best match remains stable

The monitoring network and the four nested candidate domains are shown in Fig. 1. Eligible donors are concentrated along the coast, and the most distant one lies 196 km from the target-region centroid, so the candidate set saturates by 200 km. Candidate colour in Fig. 1 is response similarity to the targets, computed from training-period records alone; the best-matched donors sit immediately around the target cluster while the eastern and south-western extremities are the least similar. Distance alone therefore does not determine similarity, which is why the analysis treats the two separately.

The candidate distributions show the asymmetry predicted by the poor-donor avoidance mechanism. Candidate count rose with radius and saturated at 74 donors (Fig. 3a). As radius increased from 60 to 200 km, maximum response similarity changed only from 0.869 to 0.882, whereas the minimum fell from 0.520 to 0.355 (Fig. 3b). The median fell from approximately 0.80 to 0.71 and the standard deviation rose from 0.079 to 0.136. Widening the domain did not

reveal a materially better upper match; it admitted a longer lower-similarity tail. These quantities were computed before any recurrent-model retraining within the final sweep, so they provided an ex-ante prediction for that sweep rather than a post hoc explanation of its model outcomes.

Selection advantage increases with candidate-domain extent

Mean selection advantage increased from -0.0149°C at 60 km to $+0.0365^{\circ}\text{C}$ at 200 km (Table 1, Fig. 4). Across the 12 independently sampled candidate-set blocks, the fitted slope was $+0.104^{\circ}\text{C}$ per tenfold increase in block span (standard error 0.0373°C , 95% confidence interval 0.0209 – 0.1872 , $t = 2.79$, $p = 0.019$). This corresponds to an estimated increase of $0.104 \log_{10}(2) = 0.031^{\circ}\text{C}$ for a doubling of span. The 12 block means are listed in Supplementary Table S1. The distribution-free check agreed (Spearman $\rho = 0.588$, $p = 0.044$). A sensitivity analysis retaining all 36 seed-level effects and clustering standard errors by the 12 blocks gave the same point estimate ($p = 0.001$; Supplementary Table S2); because only 12 clusters are available, this result is treated as a sensitivity check rather than as stronger inferential evidence. The number of blocks favouring selection rose monotonically from zero of three at 60 km to three of three at 200 km, so the trend is not carried by one or two extreme runs.

Per-domain confidence intervals in Table 1 are wide because each rests on only three independently sampled candidate sets. They are reported descriptively; the prespecified primary test is the cross-domain trend.

The trend comes from random-pool degradation, not selected-pool improvement

Decomposing the advantage into its two arms changes the interpretation of the aggregate trend (Fig. 5). Selected-pool error showed no systematic trend with block span (-0.0115°C per decade, $p = 0.749$), whereas random-pool error increased by $+0.0925^{\circ}\text{C}$ per decade ($p = 0.0013$). The random arm accounts for approximately 89% of the positive trend in advantage; the remainder is

a small and statistically unresolved change in the selected arm. Wider domains did not materially improve the selected pool. They increased the cost of allowing lower-similarity donors into pools assembled without regard to response.

Neither absolute error series is monotonic: both arms fell at the 80-km domain before rising again. The paired advantage nonetheless increased across all four domains, because that common level shift largely cancelled within each candidate-set comparison. Reporting the arms separately is what makes the mechanism visible, and a reader who saw only the fitted trends might infer a smoothness that the underlying levels do not have.

The identity of the top-ranked donors was correspondingly stable (Supplementary Fig. S1). Of the eight most similar candidates in the 60-km domain, seven remained in the top eight at 80 km and six remained at both 120 and 200 km. At 200 km the retained fraction was 0.75 against an expected random overlap of $8/74 = 0.11$. Expanding the domain by a factor of three in candidate count therefore replaced two of the eight donors a selection procedure would choose, while changing much more about which donors an indiscriminate procedure might choose instead.

Compact domains and an exploratory candidate-dispersion diagnostic

At 60 km the mean advantage was -0.0149°C and all three candidate-set block means were negative, although the block-level confidence interval included zero. The direction is consistent with the mismatch–variance condition derived below, but the evidence does not establish a universal harmful-distance threshold. At 80 km the mean remained negative with much greater between-block uncertainty; at 120 and 200 km most or all block means were positive. Descriptively, all five target-site mean effects were negative at 60 km and positive at 200 km (Supplementary Fig. S2), although target sites are not independent replicates because they share one fitted model.

Because candidate response heterogeneity may be more portable than kilometres, we examined dispersion directly. Across the four domain means, selection advantage tracked the standard deviation of candidate response similarity ($R^2 = 0.93$) with a fitted crossing near 0.100 (Supplementary Fig. S3). The statistic can be computed before recurrent-model retraining, but two cautions are

essential. Extent and dispersion are almost collinear in this network ($r = 0.998$), so these data cannot identify which is the operative variable. Only four domain means are available, so the crossing is an exploratory scale indicator rather than a calibrated threshold.

Taken together, these results provide the empirical basis for the scoped donor-selection hierarchy summarized in Table 2. Selection advantage increased as the assigned candidate domain broadened because random-pool error rose while selected-pool error remained approximately stable. The best-ranked donors were largely retained as the domain widened, whereas the lower-similarity tail expanded. These findings support a network-level distinction between candidate sets in which formal selection is worth evaluating and compact sets in which its benefit was not evident. They do not establish a universal distance threshold or independently identify response heterogeneity, distance, data volume or redundancy as causal factors. The operational interpretation of this hierarchy is developed in the Discussion.

3 Discussion

The main result separates two explanations that imply different monitoring strategies. An increasing advantage of selection could mean that wider searches reveal increasingly useful donors. Instead, selected-pool error and the upper similarity tail were nearly stable, whereas random-pool error and lower-tail risk increased. In this network, the value of selection in a wide domain was therefore protection against poor donor choices rather than discovery of a better distant donor.

Table 2 translates the empirical findings into a two-level decision hierarchy with explicit scope conditions, rather than a universal distance rule. The first decision is made at the network level and asks whether formal selection is worth evaluating at all. Across the assigned candidate radii, realized candidate-block span increased, and the measured penalty of indiscriminate pooling rose with that span. Formal selection was therefore best supported in the broader candidate networks examined here. Candidate-response heterogeneity may provide a more portable screening diagnostic, but it was observed rather than manipulated, was nearly collinear with network breadth,

and therefore remains an exploratory hypothesis.

The second decision is made at the station level and asks how to construct the donor pool once the network-level evidence supports formal selection. Forcing-response similarity provides a provisional primary ordering among quality-controlled candidates when the target has sufficient history to estimate a response profile. This conditionality is empirically important: at 60 km, all three candidate-set blocks favoured random ordering. Moreover, the comparison was against random ordering only, and the selected and random policies were matched on donor-window count rather than exact station count. Distance should therefore be treated only as a secondary eligibility, cost or tie-breaking screen. Physical connectivity is a plausible refinement but was not evaluated here, while redundancy remains an exploratory safeguard.

These findings extend fixed-network donor-selection studies by showing that the eligible candidate distribution is itself part of the selection problem. Informed donor choice and multiple-donor averaging can improve predictions relative to random donors [11–13], but a ranking rule that is useful in one monitoring network need not provide the same benefit in another network with a different breadth or candidate distribution. Reporting selection performance without characterizing that network can obscure whether an apparent gain reflects better-donor discovery or avoidance of poorly matched donors. The selected–random arm decomposition is therefore essential because these mechanisms imply different monitoring decisions.

The findings should not be interpreted as an optimal distance for installing a new sensor. The experiment evaluates which *existing records* should enter a pooled training set; it does not prospectively optimize sensor locations. A nearby station may duplicate information already present, and a distant station may be poorly matched, but distance is only a proxy for response similarity, physical connectivity and information complementarity. A distant station may still be useful if its forcing response resembles the target or it contributes underrepresented conditions. The descriptive crossing in the fitted distance relation is therefore not a universal siting rule.

Several limitations bound the conclusions. The analysis concerns one coastal network, one response score and one recurrent architecture; independent replication in a geometry where

candidate extent and response dispersion vary independently is the most important extension. Only three candidate-set blocks were sampled per radius, so the cross-domain trend is resolved but per-domain contrasts remain weak. A fixed count of overlapping windows is a computational budget rather than a measure of effective information. Selected and random policies were matched on donor-window count but not station count, and selected pools used up to 0.7 more stations on average; a station-count-matched sensitivity analysis would be needed to isolate the ordering effect. The compact-domain sign pattern does not identify redundancy as its unique cause. The response score requires enough target history to fit a target coefficient vector and therefore does not directly support a truly ungauged target. Target-site DO was unavailable, so the conclusions apply to temperature rather than to externally verified aquaculture-site DO forecasts. The conservative DO range screen may also remove genuine severe-hypoxia observations, although DO is only an auxiliary input for the endpoint analysed here. Finally, future ERA5 weather is a retrospective ceiling condition, and straight-line radius should be compared with water-path or hydrographic connectivity before the geometry is generalized.

In summary, broadening the eligible donor domain increased the measured value of response-based selection because randomly assembled pools deteriorated, not because selected pools found progressively better donors. The resulting rule is two-stage: first determine whether the eligible candidate network is broad enough to justify formal selection; then, among quality-controlled candidates with adequate target-comparable records, use forcing-response similarity as a provisional ordering while treating distance, station count and redundancy as secondary considerations. These conclusions concern the pooling of historical records in one coastal network and do not define a universal sensor-siting distance, a validated response-dispersion threshold or a rule for truly ungauged targets.

4 Methods

Study network and target region

The National Institute of Fisheries Science (NIFS) operates the Real-time Information System for Aquaculture Environment, which reports coastal water temperature and related variables along the coast of the Republic of Korea [16]. This study analysed the southern coast. Five target sites were retained because they define one aquaculture decision region and have short records beginning between October 2020 and April 2023. Pairwise great-circle separations calculated from the station coordinates range from 4.7 to 39.2 km. They were retained irrespective of long-term completeness because their data scarcity defines the forecasting problem.

Donor eligibility was evaluated separately among non-target stations. Stations were retained when at least 80% of their reported hourly temperature records passed the quality screen described below in at least five years during 2019–2025, which yielded 74 eligible donors. Because the criterion is evaluated over reported records, it certifies measurement quality rather than temporal completeness; gaps in the reporting stream are handled by the masking and window-exclusion rules below. The most distant eligible donor lies 196 km from the target-region centroid and the maximum pairwise span of the complete donor domain is approximately 323 km. Straight-line radius is an experimental definition of eligibility and does not represent along-water connectivity around islands or peninsulas.

Observations and quality control

The NIFS archive is distributed as annual files without registration. Records are reported at approximately 30-min resolution and were aggregated to hourly means. Sensors are deployed at up to three nominal layers, but only the surface layer is available network-wide: among the stations analysed here, mid-depth sensors exist at nine and a bottom sensor at one, so all analyses use surface records only. Surface sensors sit at nominal depths of 1–5 m (median 2 m) in the operator’s station metadata. Installation dates span 2004 to 2024 (median 2018), so the network

has expanded throughout the study period and the five target sites are among its most recent additions. Their short records are consequently a property of the operational network rather than an artefact of the experimental design.

The operator distinguishes this archive from its real-time service, noting that the most recent quality control is applied to the latter. An explicit range screen was therefore applied rather than treating the archive as analysis-ready. Temperatures outside 0–35 °C and dissolved oxygen outside 0.5–15 mg L⁻¹ were masked. The DO screen was chosen conservatively to limit obvious sensor drift and may also remove genuine severe-hypoxia values; this is acknowledged as a limitation. Masked values were set to missing rather than deleted and a flag was retained for each, so that the screening is auditable. Gaps of at most 3 h were linearly interpolated and longer gaps left missing. A valid sample required 168 h of input history and at least one observed target within the 72-h forecast interval. After anomaly standardization, missing numerical inputs were replaced by zero and accompanied by binary presence channels, while unobserved targets were excluded from the loss. Each station therefore contributes locally valid windows without long gaps being filled or missing targets being imputed.

Forcing and chronology

Meteorological inputs were taken from ERA5 reanalysis [17, 18]. Deterministic astronomical tidal terms, calendar terms and solar-geometry terms were also included. No future ocean observation enters the decoder. Tides, calendar and solar terms are known exactly in advance, but future ERA5 weather is a retrospective oracle condition rather than an archived operational forecast. Because every experimental arm receives identical forcing, donor-pool comparisons remain internally controlled, and absolute accuracy should be interpreted as an upper bound. Training used 2019–2023, validation 2024 and testing 2025. Climatology, scaling parameters, ridge-surrogate coefficients and donor scores were estimated from the training period only, and windows crossing a split boundary were excluded.

Forecast model and information flow

The forecaster is a gated recurrent encoder–decoder [19, 20]. Let $\mathcal{H} = \{6, 12, \dots, 72\}$ denote the forecast leads. For station s and forecast origin t , the encoder maps a locally valid 168-h observation window, past forcing, a binary input-presence mask and an eight-dimensional station representation to a hidden state,

$$\mathbf{h}_{s,t} = E_{\theta_E}(\mathbf{X}_{s,t-167:t}, \mathbf{F}_{s,t-167:t}^{\text{past}}, \mathbf{M}_{s,t-167:t}^{\text{in}}, \mathbf{e}_s), \quad (1)$$

where $\mathbf{X}$ contains temperature anomaly and any available auxiliary observations. The recurrent decoder generates the forecast sequence from the hidden state and lead-time-aligned future covariates,

$$\{\hat{a}_{s,t+h} : h \in \mathcal{H}\} = D_{\theta_D}(\mathbf{h}_{s,t}, \{\mathbf{u}_{s,t+h} : h \in \mathcal{H}\}), \quad (2)$$

where $\mathbf{u}_{s,t+h}$ contains deterministic tidal terms, calendar and solar-geometry terms, and retrospective ERA5 weather for lead h . The complete parameter set is $\Theta = \{\theta_E, \theta_D, \{\mathbf{e}_s\}\}$. The model predicts temperature anomaly, which is transformed back to degrees Celsius using the station- and day-of-year climatology estimated on the training split and smoothed with a 15-day centred window. Hidden width is 96 and the model contains approximately 89,000 trainable parameters.

Figure 2a separates training from target inference. During training, the chosen donor pool determines which station windows contribute to Θ ; target-site windows are also included whenever locally valid and are identical across paired arms. During inference, a target forecast uses only that target’s own recent history, presence information, station representation and future forcing. Donor observations are not supplied as real-time spatial inputs, so the experiment evaluates transfer through pooled parameter learning rather than spatial data assimilation. Training used AdamW with learning rate 2×10^{-3} , batch size 256, a Huber loss and early stopping on the common 2024 target-site validation set. The primary endpoint is 72-h RMSE averaged across the five targets during 2025. Target errors are not treated as independent replicates because they share one fitted model and regional forcing.

Response-similarity score and candidate dispersion

For each station, a ridge surrogate relates the 72-h temperature anomaly to the current anomaly and to forcing summaries. Define

$$\mathbf{z}_{s,t,H} = \left[a_s(t), \mathbf{f}_s(t+H)^\top, \bar{\mathbf{f}}_{s,(t,t+H]}^\top \right]^\top, \quad H = 72 \text{ h}, \quad (3)$$

and fit

$$a_s(t+H) = \alpha_s + \mathbf{b}_s^\top \mathbf{z}_{s,t,H} + e_s(t). \quad (4)$$

Ridge regression stabilizes coefficients under correlated predictors [21]. Predictors were standardized jointly across stations, and a common penalty was selected by blocked cross-validation on contiguous training-period segments because random splits are optimistic under serial dependence [22]. The intercept was excluded from the response vector. With target set T , define

$$\bar{\mathbf{b}}_T = \frac{1}{|T|} \sum_{q \in T} \mathbf{b}_q, \quad r_s = \frac{\mathbf{b}_s^\top \bar{\mathbf{b}}_T}{\|\mathbf{b}_s\|_2 \|\bar{\mathbf{b}}_T\|_2}. \quad (5)$$

All fitted coefficient vectors used in the ranking had non-zero norm. The score is forecast-model-free but not data-free: it requires inexpensive surrogate fits to target and candidate training records, but no retraining of the recurrent forecaster. It is treated as fixed in the domain sweep.

For descriptive screening, response dispersion in the full eligible domain was computed as

$$D(R) = \left[\frac{1}{|C(R)| - 1} \sum_{s \in C(R)} (r_s - \bar{r}_R)^2 \right]^{1/2}, \quad \bar{r}_R = \frac{1}{|C(R)|} \sum_{s \in C(R)} r_s. \quad (6)$$

Because $D(R)$ was not manipulated and almost perfectly co-varies with extent, its association with selection advantage is treated as exploratory.

Candidate blocks, training budget and experimental design

Let $C(R)$ be the set of eligible donors within assigned radius R of the target centroid. Experimental block b samples without replacement

$$S_{Rb} \subset C(R), \quad |S_{Rb}| = 20. \quad (7)$$

Let $\mathcal{W}_s$ be the eligible donor windows from station s and $q_s = |\mathcal{W}_s|$. For an ordering π of S_{Rb} and a fixed donor-window budget $W = 19,066$, define the shortest prefix reaching the budget as

$$m(\pi, W) = \min \left\{ m : \sum_{j=1}^m q_{\pi_j} \geq W \right\}. \quad (8)$$

The donor training set is then a common-seed sample

$$\mathcal{D}_{\pi, W} \subseteq \bigcup_{j=1}^{m(\pi, W)} \mathcal{W}_{\pi_j}, \quad |\mathcal{D}_{\pi, W}| = W, \quad (9)$$

with the same target-site windows added to every paired arm. The selected ordering sorts candidates by decreasing r_s ; two reference orderings are independent random permutations.

Selected pools required slightly more stations than random pools to reach the same donor-window budget: mean counts were 8.0 versus 7.8 at 60 km, 8.3 versus 8.0 at 80 km, 8.0 versus 7.3 at 120 km and 7.7 versus 7.0 at 200 km. Thus the experiment controls donor-window count, not station count or effective information; realized pools contained 7–9 donors. Figure 2b shows that the donor-pool construction policy is the only factor changed within a paired block, while model, chronology, validation set, optimizer and seeds are held fixed.

Assigned radius was varied over 60, 80, 120 and 200 km, whose full domains contained 25, 35, 60 and 74 eligible donors. At every radius, three independently sampled candidate sets of 20 donors were used (Table 3). Fixing the draw size prevents a wider domain from gaining merely because more candidates are considered, while the fixed window budget controls nominal training volume. Within each block, selected and random runs shared seeds 9300, 9301 and 9302;

pairing common seeds and splits reduces optimization variance that can otherwise obscure small differences [23, 24]. The correction of an earlier structurally unbalanced pilot sweep is documented in Supplementary Note.

Three pilot-informed predictions were fixed before the final 108-run sweep, each with an explicit falsification condition. Domain expansion should lengthen the lower-similarity tail while leaving the upper edge nearly unchanged; this is falsified if the upper edge improves materially with extent. Random-pool error should rise with extent while selected-pool error stays approximately flat; this is falsified if the selected pool improves systematically or the random pool does not degrade. Where candidate similarities are tightly distributed, mismatch reduction may be too small to offset estimation variance; this is falsified if selection is consistently beneficial in every domain.

The last prediction follows from a standard squared-error decomposition. For pool-construction policy $g \in \{\text{random}, \text{selected}\}$, let

$$\text{MSE}_g = \mathbb{E}[(Y - \hat{f}_g(X))^2] = \sigma_\epsilon^2 + B_g^2 + V_g, \quad (10)$$

where B_g^2 is integrated squared bias and V_g is integrated estimator variance under the common target test distribution. The RMSE advantage is

$$\Delta_{\text{RMSE}} = \sqrt{\text{MSE}_{\text{random}}} - \sqrt{\text{MSE}_{\text{selected}}} = \frac{\text{MSE}_{\text{random}} - \text{MSE}_{\text{selected}}}{\sqrt{\text{MSE}_{\text{random}}} + \sqrt{\text{MSE}_{\text{selected}}}}. \quad (11)$$

The denominator is positive, and the common irreducible noise cancels. Therefore

$$\Delta_{\text{RMSE}} > 0 \iff B_{\text{random}}^2 - B_{\text{selected}}^2 > V_{\text{selected}} - V_{\text{random}}. \quad (12)$$

Similarity-based selection can reduce mismatch-related bias, but it need not reduce estimator variance. Hence its sign is not guaranteed, particularly in a compact domain where the reducible bias difference may already be small. The experiment tests this sign condition; it does not identify redundancy or any other variance component as the unique cause.

Statistical analysis

For assigned radius R , block b and seed s , the seed-level advantage is

$$\Delta_{Rbs} = \frac{1}{2} \sum_{j=1}^2 \text{RMSE}_{\text{random},j,Rbs} - \text{RMSE}_{\text{selected},Rbs}, \quad (13)$$

with positive values favouring response-based selection. Because seeds within a candidate set are repeated optimizations rather than new donor networks, the primary block effect averages them, $\bar{\Delta}_{Rb} = \frac{1}{3} \sum_s \Delta_{Rbs}$, and the candidate-set block is the inferential unit. Assigned radius is the manipulated factor. The primary continuous exposure metric is the realized maximum pairwise block span L_{Rb} , and the 12 seed-averaged block effects are regressed on its base-10 logarithm,

$$\bar{\Delta}_{Rb} = \beta_0 + \beta_1 \log_{10} L_{Rb} + \varepsilon_{Rb}. \quad (14)$$

A positive β_1 indicates increasing selection value with extent; the fitted change for a doubling of span is $\beta_1 \log_{10} 2$. Spearman correlation over the 12 block means is a distribution-free check. A sensitivity analysis retains all 36 seed-level effects with standard errors clustered by block (Supplementary Table S2), but it is not treated as stronger evidence because only 12 clusters are available. Arm-specific trends are estimated from the same 12 block means. Per-radius summaries are descriptive analyses over three blocks and are not treated as four separately powered confirmatory tests. Target-specific results and the four-point dispersion relation are likewise descriptive.

Use of generative artificial-intelligence tools

OpenAI ChatGPT was used to assist with English-language editing, LaTeX organization and consistency checks. All experimental analyses, numerical results, interpretations, citations and final wording were checked and approved by the authors.

Data availability

All data generated or analysed during this study are included in this published article and its supplementary information files.

Code availability

The source code and numerical result files required to reproduce the findings of this manuscript are publicly available in Zenodo at [10.5281/zenodo.22206710](https://zenodo.org/doi/10.5281/zenodo.22206710) [25].

Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2026-25469002), the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (IITP-2026-RS-2026-25619158, AI Star Fellowship Support Program), and the Culture, Sports and Tourism R&D Program through the Korea Creative Content Agency grant funded by the Ministry of Culture, Sports and Tourism in 2026 (Project Name: Content Illegal Streaming Infringement Analysis Digital Forensics Human Resource Development Center, Project Number: RS-2026-25513329).

Author contributions

H.T.P.N.: conceptualization, methodology, software, validation, formal analysis, investigation, data curation, visualization, writing — original draft. H.J.: conceptualization, methodology, validation, resources, supervision, project administration, funding acquisition, writing — review and editing. Both authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

References

- [1] Food and Agriculture Organization of the United Nations. The state of world fisheries and aquaculture 2024: Blue transformation in action. Tech. Rep., Food and Agriculture Organization of the United Nations, Rome (2024).
- [2] Naylor, R. L. *et al.* A 20-year retrospective review of global aquaculture. *Nature* **591**, 551–563 (2021).
- [3] Hwang, E. K. & Park, C. S. Seaweed cultivation and utilization of Korea. *ALGAE* **35**, 107–121 (2020).
- [4] Breitburg, D. *et al.* Declining oxygen in the global ocean and coastal waters. *Science* **359**, eaam7240 (2018).
- [5] Kim, M., Yang, H. & Kim, J. Sea surface temperature and high water temperature occurrence prediction using a long short-term memory model. *Remote Sensing* **12**, 3654 (2020).
- [6] Jung, S., Kim, Y. J., Park, S. & Im, J. Prediction of sea surface temperature and detection of ocean heat wave in the South Sea of Korea using time-series deep-learning approaches. *Korean Journal of Remote Sensing* **36**, 1077–1093 (2020).
- [7] Choi, H.-M., Kim, M.-K. & Yang, H. Deep-learning model for sea surface temperature prediction near the Korean Peninsula. *Deep Sea Research Part II: Topical Studies in Oceanography* **208**, 105262 (2023).
- [8] Nguyen, H. T. P. & Jeong, H. Conditional value of monitoring-network expansion for coastal temperature forecasting: Paired-block estimation, forcing-response selection, and detectable

deployment scale. SSRN preprint (2026). URL <https://doi.org/10.2139/ssrn.7275372>.
SSRN abstract 7275372.

[9] Ben-David, S. *et al.* A theory of learning from different domains. *Machine Learning* **79**, 151–175 (2010).

[10] Pan, S. J. & Yang, Q. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering* **22**, 1345–1359 (2010).

[11] Oudin, L., Andréassian, V., Perrin, C., Michel, C. & Le Moine, N. Spatial proximity, physical similarity, regression and ungauged catchments: A comparison of regionalization approaches based on 913 French catchments. *Water Resources Research* **44**, W03413 (2008).

[12] Zhang, Y. & Chiew, F. H. S. Relative merits of different methods for runoff predictions in ungauged catchments. *Water Resources Research* **45**, W07412 (2009).

[13] Pool, S., Vis, M. & Seibert, J. Regionalization for ungauged catchments—lessons learned from a comparative large-sample study. *Water Resources Research* **57**, e2021WR030437 (2021).

[14] Krause, A., Singh, A. & Guestrin, C. Near-optimal sensor placements in Gaussian processes: Theory, efficient algorithms and empirical studies. *Journal of Machine Learning Research* **9**, 235–284 (2008). URL <https://www.jmlr.org/papers/v9/krause08a.html>.

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

[16] National Institute of Fisheries Science. Real-time information system for aquaculture environment (RISA) observation data. National Institute of Fisheries Science, Republic of Korea (2026). URL <https://www.nifs.go.kr/risa/main.do>. Accessed 6 August 2026.

[17] Hersbach, H. *et al.* The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society* **146**, 1999–2049 (2020).

- [18] Hersbach, H. *et al.* ERA5 hourly data on single levels from 1940 to present. Copernicus Climate Change Service Climate Data Store (2023). URL <https://doi.org/10.24381/cds.adbb2d47>.
- [19] Cho, K. *et al.* Learning phrase representations using RNN encoder–decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, 1724–1734 (2014).
- [20] Sutskever, I., Vinyals, O. & Le, Q. V. Sequence to sequence learning with neural networks. In *Advances in Neural Information Processing Systems*, vol. 27, 3104–3112 (2014). URL <https://proceedings.neurips.cc/paper/2014/hash/a14ac55a4f27472c5d894ec1c3c743d2-Abstract.html>.
- [21] Hoerl, A. E. & Kennard, R. W. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* **12**, 55–67 (1970).
- [22] Roberts, D. R. *et al.* Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography* **40**, 913–929 (2017).
- [23] Bouthillier, X. *et al.* Accounting for variance in machine learning benchmarks. In *Proceedings of Machine Learning and Systems*, vol. 3, 747–769 (2021). URL https://proceedings.mlsys.org/paper_files/paper/2021/hash/0184b0cd3cfb185989f858a1d9f5c1eb-Abstract.html.
- [24] Reimers, N. & Gurevych, I. Reporting score distributions makes a difference: Performance study of LSTM-networks for sequence tagging. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 338–348 (2017).
- [25] Nguyen, H. T. P. & Jeong, H. Software and reproducibility materials for candidate-network breadth and donor selection in coastal aquaculture temperature forecasting. Zenodo (2026). URL <https://doi.org/10.5281/zenodo.22206710>. Research software and numerical reproducibility materials.

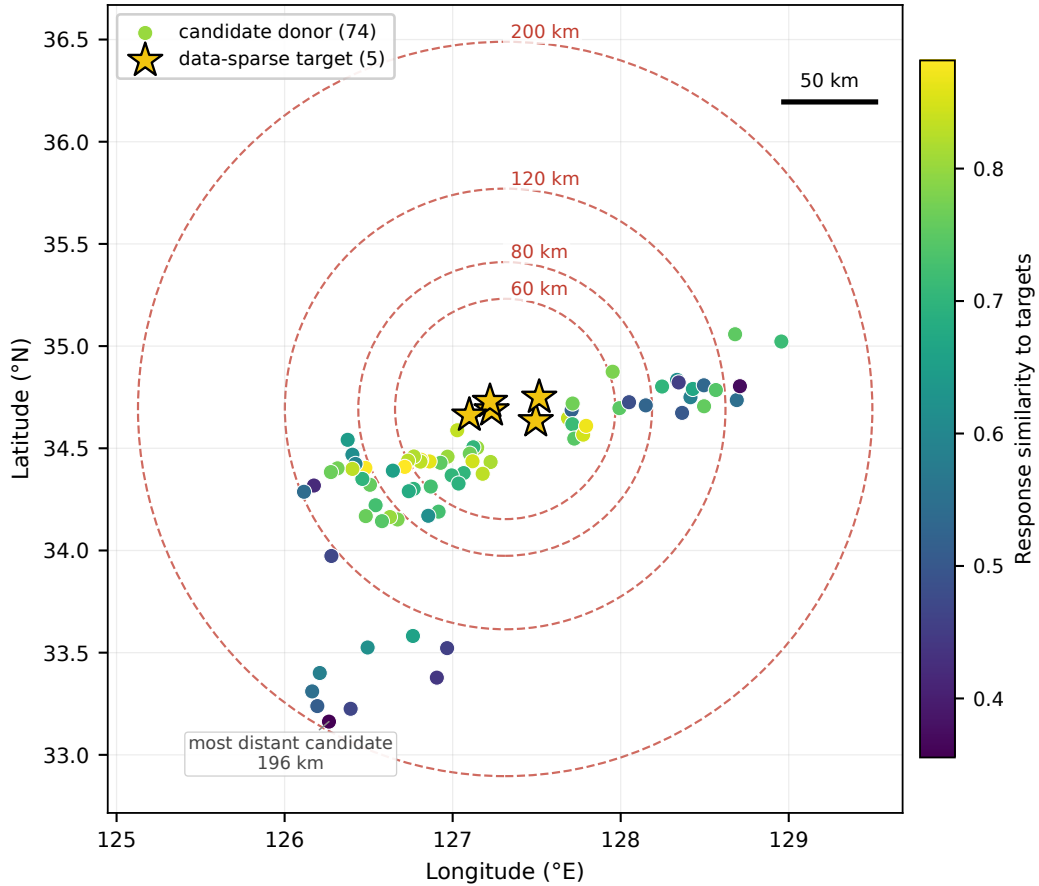
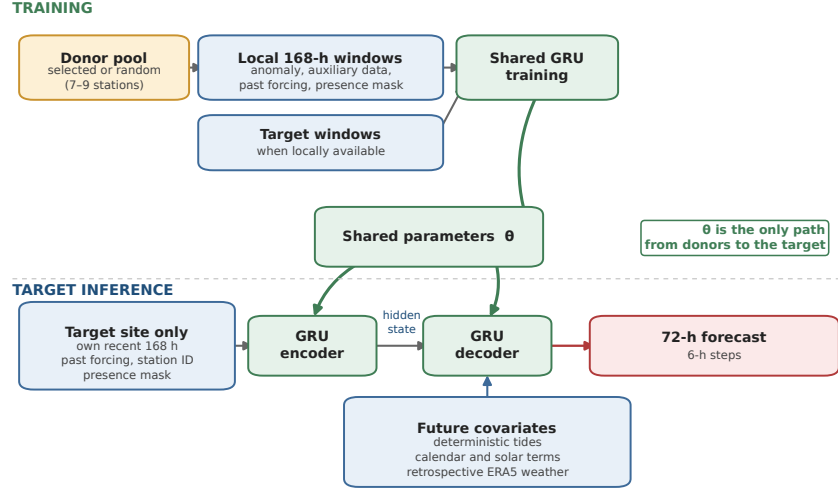
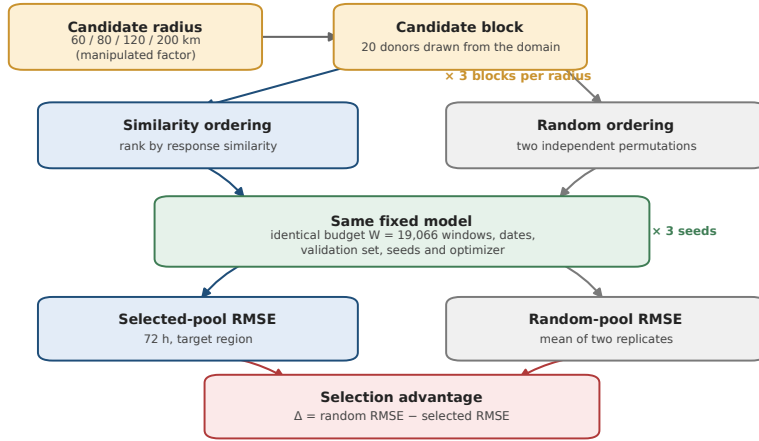


Figure 1: **Study network and nested candidate domains.** Seventy-four eligible donor stations are coloured by response similarity to the five data-sparse targets, computed from training-period records only. Dashed circles mark the 60-, 80-, 120- and 200-km domains centred on the target-region centroid. The targets form one purposeful application region rather than a spatially representative sample of the southern coast. The most distant eligible donor is approximately 196 km from the centroid, which is why the candidate set saturates by 200 km.



(a) Fixed forecasting model: training and target inference



(b) Controlled donor-selection experiment

Figure 2: **Forecast information flow and controlled donor-selection comparison.** (a) Historical windows from the selected or random donor pool and locally available target windows train one shared GRU encoder–decoder. At target inference, only the target’s own recent history, past forcing, presence mask, station identity and lead-time-specific future covariates are used; no donor observation enters the target forecast, and donor records affect it only through the shared parameters. (b) Assigned radii are sampled into 20-donor blocks. A response-similarity ordering and two random orderings are evaluated under the same 19,066-window donor budget, model, chronology, validation set, optimizer and seeds. Only the donor-pool construction policy differs between paired arms; the 12 candidate-set blocks are the inferential units, and the 108 fitted models produce the paired 72-h RMSE contrast.

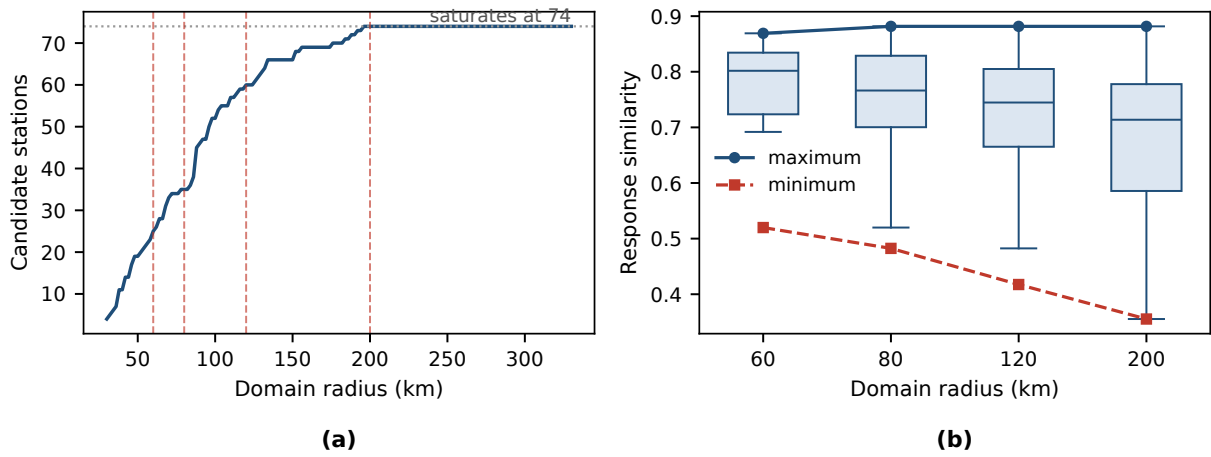


Figure 3: **Candidate structure before any forecast-model retraining.** (a) Number of eligible donors within a given radius of the target-region centroid; vertical lines mark the experimental radii and the horizontal line marks saturation at 74 donors. (b) Response-similarity distributions in the four full domains. Boxes show the interquartile range and median; overlaid lines show the maximum and minimum. The upper edge is nearly flat while the lower edge falls, so widening the domain adds lower-similarity candidates while leaving the best available match nearly unchanged.

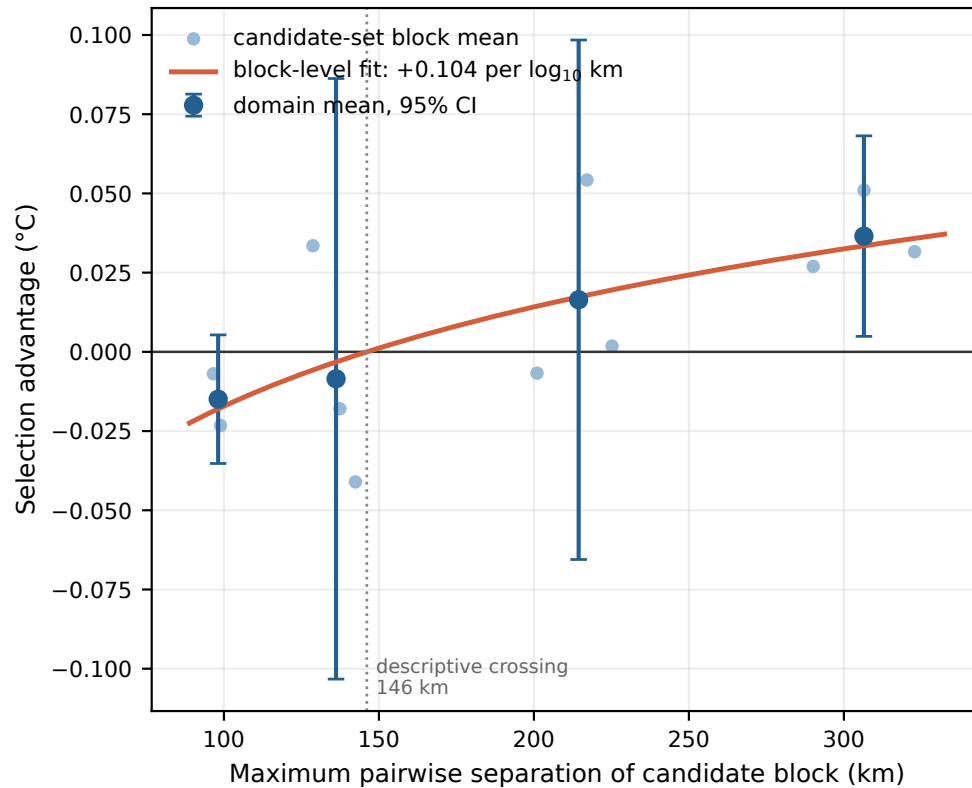


Figure 4: **Selection advantage versus candidate-block span.** Faint points are the 12 seed-averaged candidate-set block effects. Large points are domain means with 95% confidence intervals over three blocks. The line shows the primary regression on log-transformed span for the 12 block means. Its zero crossing near 146 km is descriptive and is not interpreted as a universal distance threshold.

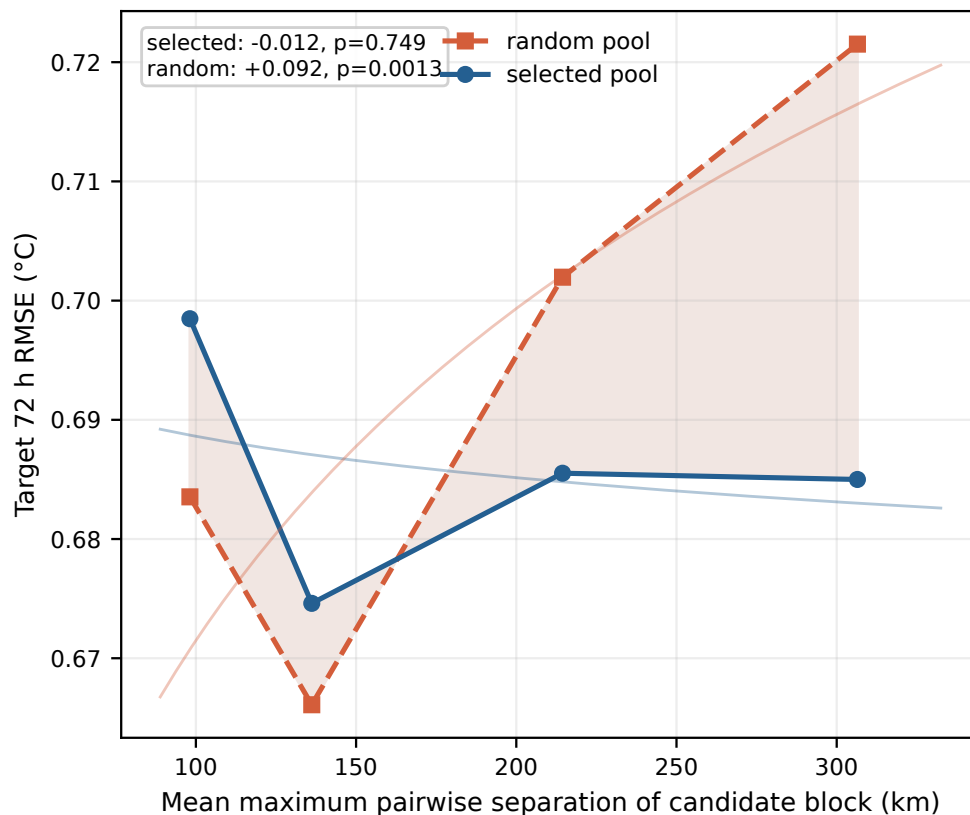


Figure 5: **The advantage decomposed into its two arms.** Selected- and random-pool target RMSE are plotted against mean candidate-block span; the shaded region between them is the selection advantage. Thin lines are arm-specific logarithmic fits to the 12 seed-averaged block means. The selected arm shows no detectable trend whereas random-pool error rises with extent.

Table 1: **Target-region 72-h RMSE and selection advantage by domain radius.** Advantage is random-pool RMSE minus selected-pool RMSE. Confidence intervals and positive-block counts use the three independently sampled candidate-set blocks at each radius; retraining seeds are averaged within blocks. All quantities are in degrees Celsius. Per-radius intervals rest on three blocks and should be read descriptively; the prespecified primary test is the cross-domain trend of equation (14).

Radius (km)	Selected	Random	Advantage	95% CI	Positive blocks
60	0.6985	0.6835	−0.0149	[−0.0352, +0.0053]	0/3
80	0.6746	0.6661	−0.0085	[−0.1033, +0.0863]	1/3
120	0.6855	0.7020	+0.0164	[−0.0655, +0.0984]	2/3
200	0.6850	0.7215	+0.0365	[+0.0049, +0.0682]	3/3

Table 2: **Empirical evidence and scoped decision hierarchy for donor-station selection.** The table translates the domain-sweep results into a two-level framework while distinguishing direct performance contrasts from constraints, indirect observations and exploratory interpretations. It concerns pooling of existing records, not prospective sensor siting. “Direct” denotes a manipulated performance contrast; “Constraint” a factor held fixed for comparability; “Indirect” an observed but unmanipulated association; and “Exploratory” an interpretation requiring a dedicated test. Ranks are assigned within each decision level.

Decision level	Rank	Criterion	Evidence from this experiment	Recommended use
Network	1	Breadth of the eligible candidate set	Direct. Eligibility radius was assigned experimentally; realized maximum pairwise block span was the continuous exposure metric. Selection advantage increased by 0.104 °C per $\log_{10}$ increase in span (95% CI 0.021–0.187).	Decide whether formal selection is worth evaluating. The descriptive 146-km crossing is specific to this network and is not a transferable threshold.
Network	2	Response heterogeneity of the candidate set	Exploratory. Heterogeneity was not manipulated and is almost collinear with breadth ($r = 0.998$). The association ($R^2 = 0.93$ over four domain means; crossing near 0.100 SD) rests on four points.	Use only as a pre-retraining screening hypothesis until validated where breadth and heterogeneity vary independently.
Station	1	Forcing-response similarity to the target	Direct policy contrast against random ordering under an equal donor-window budget. Distance-, correlation- and volume-based orderings were not tested here. Similarity-based pools used 0.2–0.7 more stations on average, so the isolated marginal effect of ordering was not identified. All three 60-km blocks favoured random ordering.	Provisional primary ordering only after the network-level decision favours formal selection and sufficient target history exists to estimate a response profile. Confirm with a station-count-matched sensitivity analysis.
Station	2	Sufficient locally valid observations	Constraint. Every arm used exactly 19,066 donor windows and required 7–9 donors. The marginal effect of data volume was not estimated.	Apply the prespecified donor-eligibility and quality-control screen before ranking; report the realized usable-window count and number of included stations.
Station	3	Distance and physical connectivity	Indirect. Maximum similarity barely changed with extent while low-similarity candidates accumulated near the domain margins; stations at comparable distances could differ in response similarity.	Use distance for eligibility, cost or tie-breaking rather than as the primary ranking rule. Where available, evaluate water-path or hydrographic connectivity as a refinement; it was not tested here.
Station set	1	Redundancy and information complementarity	Exploratory. The compact-domain sign pattern is compatible with limited incremental information from similar donors, but redundancy was not isolated as its cause.	Treat diversification as a secondary safeguard and verify donor–donor correlation or event coverage before adopting it as a rule.

Table 3: **Controlled candidate-domain sweep.** Block span is the maximum pairwise separation of each sampled 20-candidate set. Each radius contains three candidate-set blocks, three common retraining seeds, one selected ordering and two random orderings, for 27 model runs.

Radius (km)	Eligible donors	Block span (km)	Windows per run	Model runs
60	25	96.6–98.9	19,066	27
80	35	128.7–142.4	19,066	27
120	60	201.0–225.2	19,066	27
200	74	290.1–322.8	19,066	27

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplementary2.pdf](#)
- [datasetarchive.zip](#)