

# Supplementary Information

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

Hang Thi Phuong Nguyen

Hieyong Jeong

### Contents

- **Supplementary Table S1.** Seed-averaged results for the twelve candidate-set blocks. Cited in Results, “Selection advantage increases with candidate-domain extent”.
- **Supplementary Table S2.** Primary and cluster-robust analyses of the domain-extent trend. Cited in Results, “Selection advantage increases with candidate-domain extent”, and in Methods, “Statistical analysis”.
- **Supplementary Fig. S1.** Stability of the top-ranked candidate set across candidate-domain radii. Cited in Results, “The trend comes from random-pool degradation, not selected-pool improvement”.
- **Supplementary Fig. S2.** Descriptive target-specific selection advantages. Cited in Results, “Compact domains and an exploratory candidate-dispersion diagnostic”.
- **Supplementary Fig. S3.** Exploratory relation between selection advantage and candidate-response dispersion. Cited in the compact-domain Results subsection.
- **Supplementary Note.** Correction of the preliminary domain sweep. Cited in Methods, “Candidate blocks, training budget and experimental design”.

## Supplementary Table S1

Table S1: **Seed-averaged results for the twelve candidate-set blocks.** Each row is one independently sampled set of 20 candidate donors at one radius, averaged over the three common retraining seeds. Span is the maximum pairwise separation within the sampled candidate set and is the predictor used in the primary trend test. Advantage is random-pool RMSE minus selected-pool RMSE, so positive values favour response-based selection. All errors and advantages are in degrees Celsius. These twelve values are the independent units of the primary analysis reported in the main text.

| Radius (km) | Block | Span (km) | Selected RMSE | Random RMSE | Advantage |
|-------------|-------|-----------|---------------|-------------|-----------|
| 60          | 1     | 96.6      | 0.6906        | 0.6837      | −0.0069   |
| 60          | 2     | 98.9      | 0.7024        | 0.6792      | −0.0233   |
| 60          | 3     | 98.9      | 0.7024        | 0.6878      | −0.0147   |
| 80          | 1     | 137.4     | 0.6775        | 0.6596      | −0.0180   |
| 80          | 2     | 128.7     | 0.6398        | 0.6733      | +0.0335   |
| 80          | 3     | 142.4     | 0.7065        | 0.6655      | −0.0410   |
| 120         | 1     | 201.0     | 0.6988        | 0.6921      | −0.0067   |
| 120         | 2     | 217.1     | 0.6547        | 0.7089      | +0.0542   |
| 120         | 3     | 225.2     | 0.7031        | 0.7049      | +0.0018   |
| 200         | 1     | 306.5     | 0.6642        | 0.7152      | +0.0510   |
| 200         | 2     | 322.8     | 0.6999        | 0.7315      | +0.0316   |
| 200         | 3     | 290.1     | 0.6909        | 0.7179      | +0.0270   |

The block means show why per-radius intervals in the main text are wide. At 80 and 120 km the three blocks disagree in sign, so the domain mean is poorly resolved even though the cross-domain trend is not. At 60 km all three blocks are negative and at 200 km all three are positive; the count of positive blocks rises monotonically across the four domains.

## Supplementary Table S2

Table S2: **Primary and cluster-robust analyses of the domain-extent trend.** Both analyses regress the selection advantage on the base-10 logarithm of candidate-block span. The primary analysis averages the three retraining seeds within each block and uses the twelve independently sampled blocks as the inferential units. The sensitivity analysis retains all 36 seed-level effects and clusters standard errors by the twelve blocks, which relaxes the assumption that seeds carry no information while still refusing to treat them as independent networks. The two give the same point estimate; the seed-level version is more precise because it uses the within-block replication that averaging discards.

| Analysis                   | Effects | Clusters | Slope ( $^{\circ}\text{C}$ per decade) | 95% CI             | <i>p</i> |
|----------------------------|---------|----------|----------------------------------------|--------------------|----------|
| Seed-averaged block means  | 12      | 12       | +0.1040                                | [+0.0209, +0.1872] | 0.0192   |
| Seed-level, cluster-robust | 36      | 12       | +0.1040                                | [+0.0525, +0.1556] | 0.0010   |

The seed-averaged analysis is reported as primary because retraining seeds are repeated optimizations of the same donor network rather than new draws from the population of candidate sets. Treating them as independent would overstate the effective sample size. The cluster-robust analysis preserves the same point estimate and is reported only as a sensitivity check; with 12 clusters, its smaller standard error should not be interpreted as stronger evidence than the primary block-level analysis.

## Supplementary Fig. S1

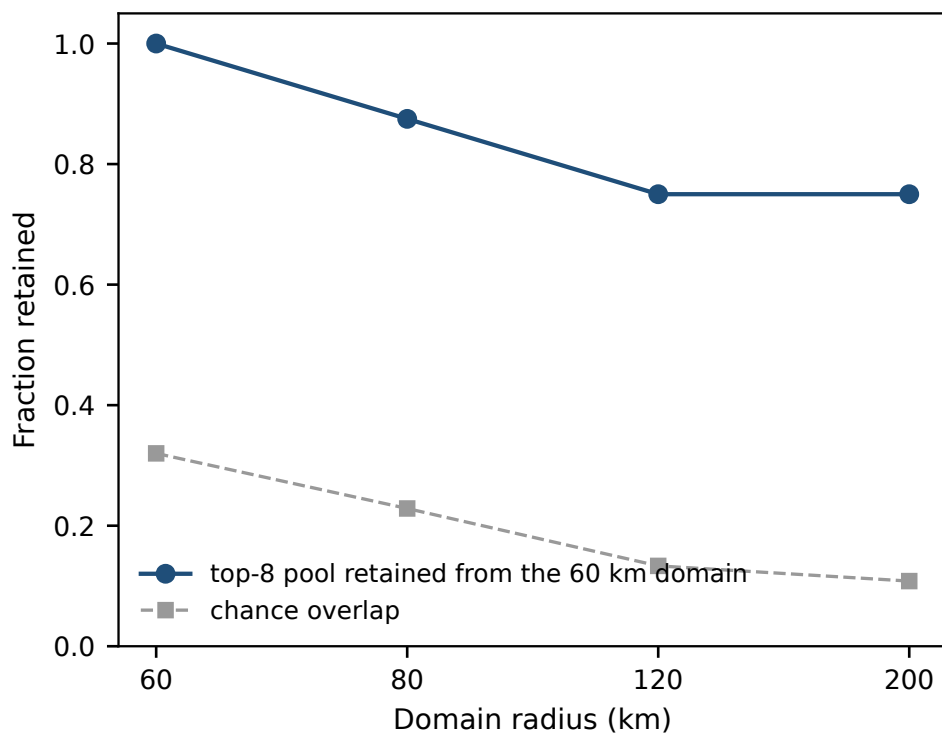


Figure S1: **Stability of the top-ranked candidate set across domain radii.** The solid line is the fraction of the 60-km top eight retained in each wider full domain. The dashed line is the overlap expected under random top-eight membership,  $8/|C(R)|$ . Six of the eight highest-ranked 60-km donors remain in the top eight at both 120 and 200 km, supporting the interpretation that widening the domain changes the lower-similarity tail more than the best available donor set. This ranking diagnostic is descriptive; the fixed-window training pools contained 7–9 donors.

## Supplementary Fig. S2

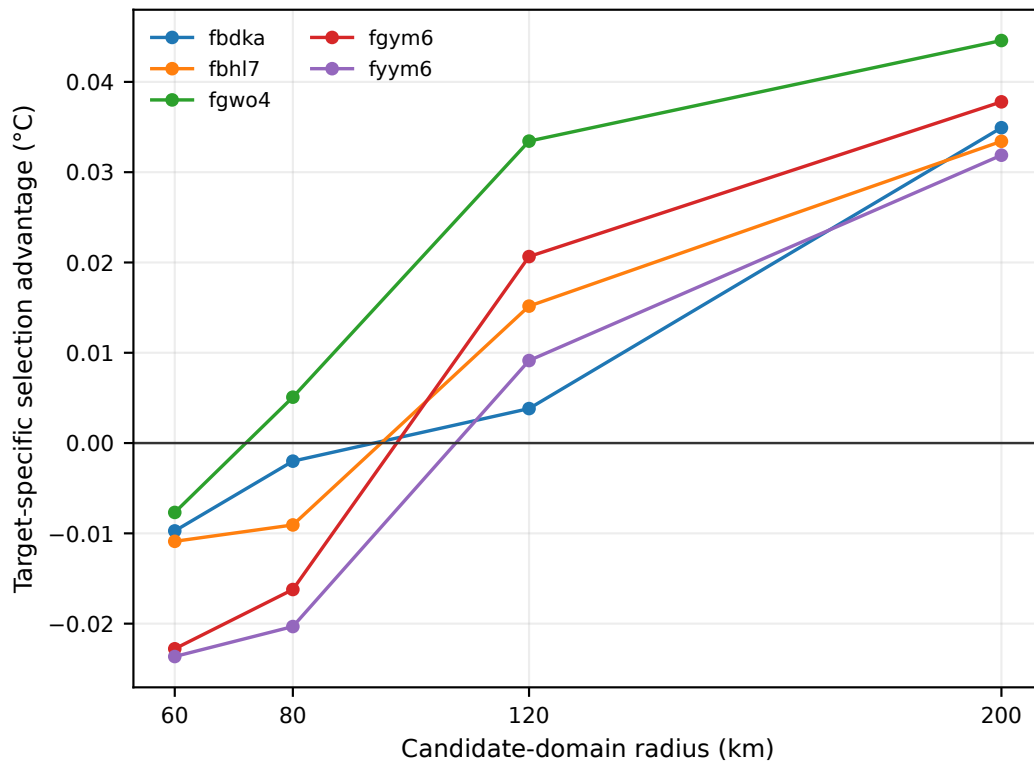


Figure S2: **Descriptive target-specific selection advantages across candidate-domain radii.** Positive values indicate lower 72-h RMSE for the response-selected pool than for the mean of the two paired random pools. Each value averages three candidate-set blocks and three retraining seeds. All five target means are negative at 60 km and positive at 200 km, so the direction of the main result is shared across the application region rather than driven by one site. The five targets nevertheless share every fitted model and one application-defined regional setting, so these lines are descriptive and are not treated as independent inferential replications.

## Supplementary Fig. S3

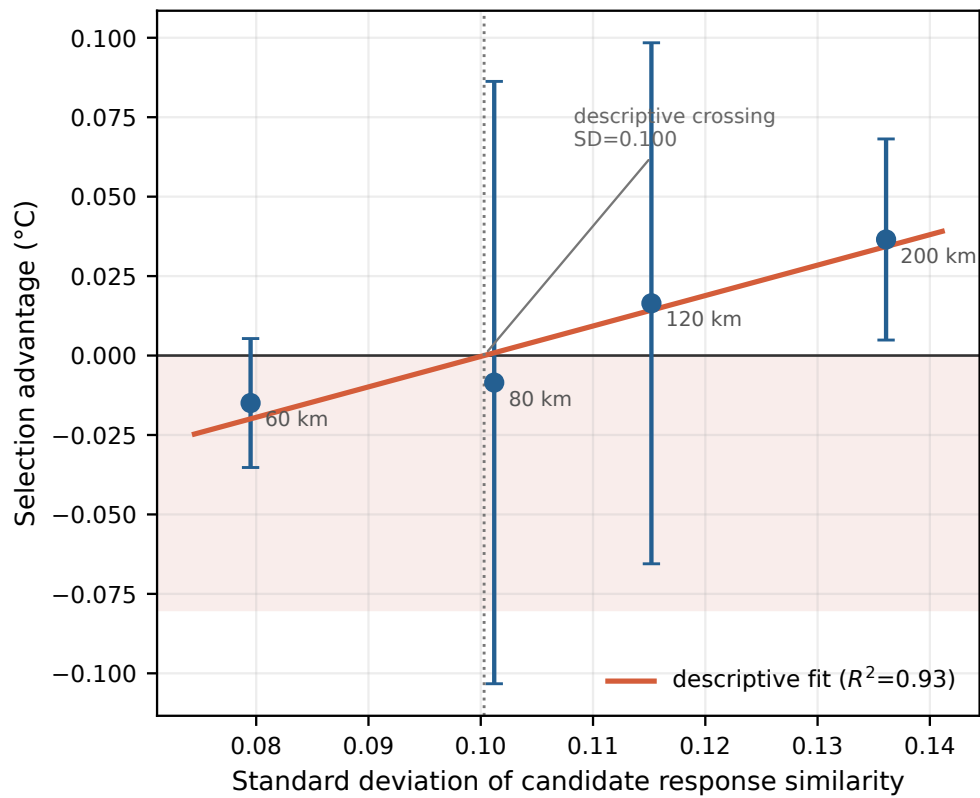


Figure S3: **Selection advantage versus candidate-response dispersion.** Points are domain means with 95% confidence intervals over three candidate-set blocks. The fit has  $R^2 = 0.93$  and crosses zero near a similarity standard deviation of 0.100. Both quantities are exploratory: radius and dispersion are nearly collinear in this network ( $r = 0.998$ ), only four domain means are available, and the crossing is not a calibrated decision threshold.

## Supplementary Note. Correction of the preliminary domain sweep

A preliminary experiment used candidate radii of 60, 120 and 400 km and sampled 25 candidate stations per domain. It appeared to produce a non-monotonic relation between domain extent and selection advantage. Two structural defects made that result uninterpretable, and both are reported here because the corrections they forced explain two design choices in the main text that would otherwise look arbitrary.

The first defect concerned candidate-set replication. The full 60-km domain contained exactly 25 eligible donors, so sampling 25 returned the same complete set in every nominal candidate draw. The assumed replication at that radius did not exist, and the apparent variation between draws was retraining noise alone. The final design therefore reduced the draw to 20, which provides genuine candidate-set variation at every radius including the most compact one.

The second defect concerned spatial extent. The eligible network saturates at 74 donors within 200 km of the target centroid, because the most distant qualifying station lies 196 km away. The 400-km label therefore described the same donor domain as 200 km and added no extent. The final design replaced it with 80 km, which places an experimental radius where the sign of the advantage actually changes rather than beyond the point at which the network stops growing.

Both defects were corrected before the final 108-run sweep analysed in the manuscript. Candidate draw size and the exact 19,066-window training budget were held constant throughout that sweep. The preliminary results are not used in any final-sweep hypothesis test or effect estimate; they are reported so that the corrected design can be evaluated against the reasoning that produced it. The run-level output of both sweeps is included in the archived code repository, so the candidate counts and radii described here can be verified directly.