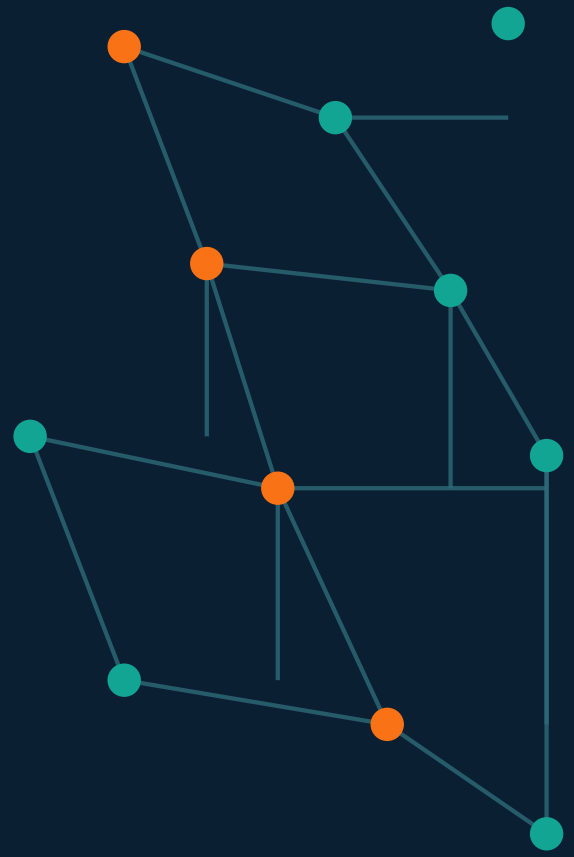
FROM DATA TO RESILIENT DECISIONS

Temporally Validated Information-Theoretic Optimization of Surface-Water Monitoring Networks

Vitalii Bezsonnyi • Oksana Davydova • Natalia Balatska
Nataliia Cherevychna • Dmytro Kramarenko

Simon Kuznets Kharkiv National University of Economics



CED-2026 • Area 4

Digital Worlds and Smart Technologies: From Data to Solutions

29 October 2026

Presenter: Vitalii Bezsonnyi

DATA → INFORMATION → DECISIONS

Why optimize a monitoring network?

The challenge is not merely to reduce stations — it is to preserve decision-relevant information.

The management tension

RESOURCE LOAD Laboratory, field and logistics costs

REDUNDANCY Nearby stations may carry similar information

RELIABILITY Removing the wrong station can erase unique signals

Research question

How much of an existing network can be removed before information needed for decisions materially degrades?

We test four questions

- How quickly does information deteriorate?
- Does informed selection beat random reduction?
- Does a stronger optimizer transfer better to future data?
- Is resilience basin-specific?

Data: from a national archive to a stable experimental cohort

Open harmonized monitoring data for Ukraine, 2003–July 2026.

Selection funnel

59,670 curated observations

1,005 unique Post_ID values

917 Post_ID values in 5 focal basins

362 present every year, 2012–2017

343 stable Core-10 cohort

Five river basins

Dnipro	168
Danube	51
Dniester	47
Don	36
Southern Bug	41

Core-10 parameters

BOD_5 • suspended solids • dissolved oxygen
sulfate • chloride • ammonium • nitrate
nitrite • phosphate • COD

Temporal design

SELECT 2012–2014 → **VALIDATE** 2015–2017

The primary findings apply to this 343-station stable cohort — not automatically to all 1,005 Post_ID values.

How information retention is measured

A transparent metric: compare full-network and reduced-network marginal distributions.

1 Jensen–Shannon divergence

$$JSD(P,Q) = \tfrac{1}{2}\, D_{kl}(P \,\|\, M) + \tfrac{1}{2}\, D_{kl}(Q \,\|\, M)$$
$$M = (P + Q) / 2$$

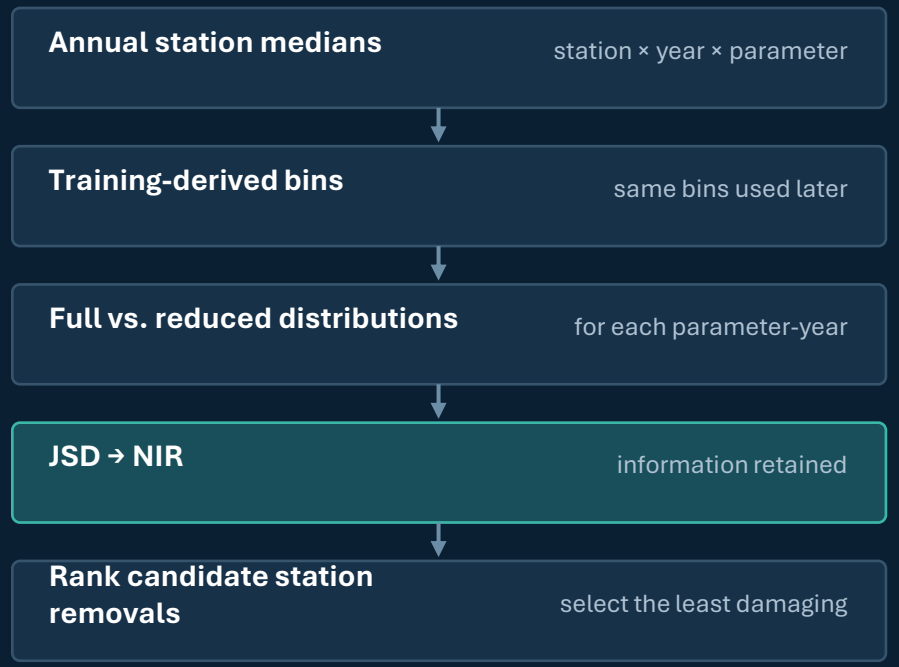
2 Network Information Retention

$$NIR(S) = 1 - \text{mean JSD across years and parameters}$$

NIR = 1 → identical discretized marginal distributions under the metric

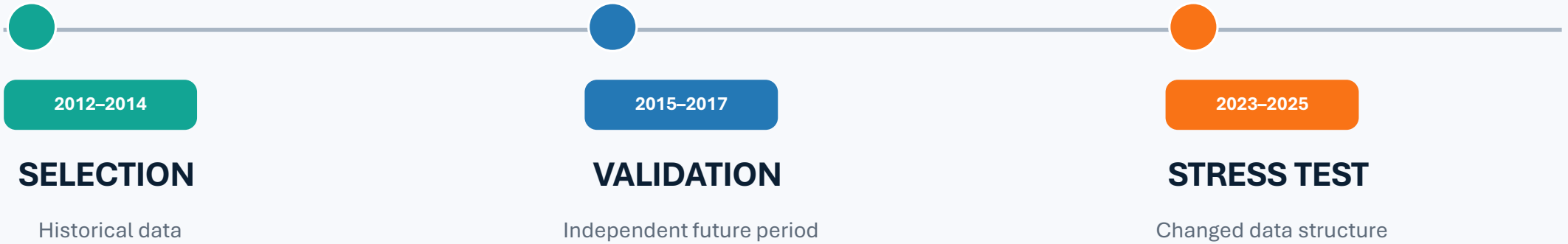
Important: NIR does not preserve the full joint multivariate dependence structure.

From measurements to a decision metric



Experimental design: selection and evaluation are separated

No 2015–2017 validation observations are used to choose stations.



Four comparison strategies

Greedy

remove the station with the smallest training-information loss

Random × 2,000

exact-size subsets + 95% Monte-Carlo interval

Adversarial

remove the most information-critical station first

Genetic algorithm

stronger fixed-cardinality search at 50% and 30%

Why this matters: better training performance is not assumed to mean better future-period performance.

Main result: informed station choice matters more as the network shrinks

Mean independent-validation NIR across five basins.

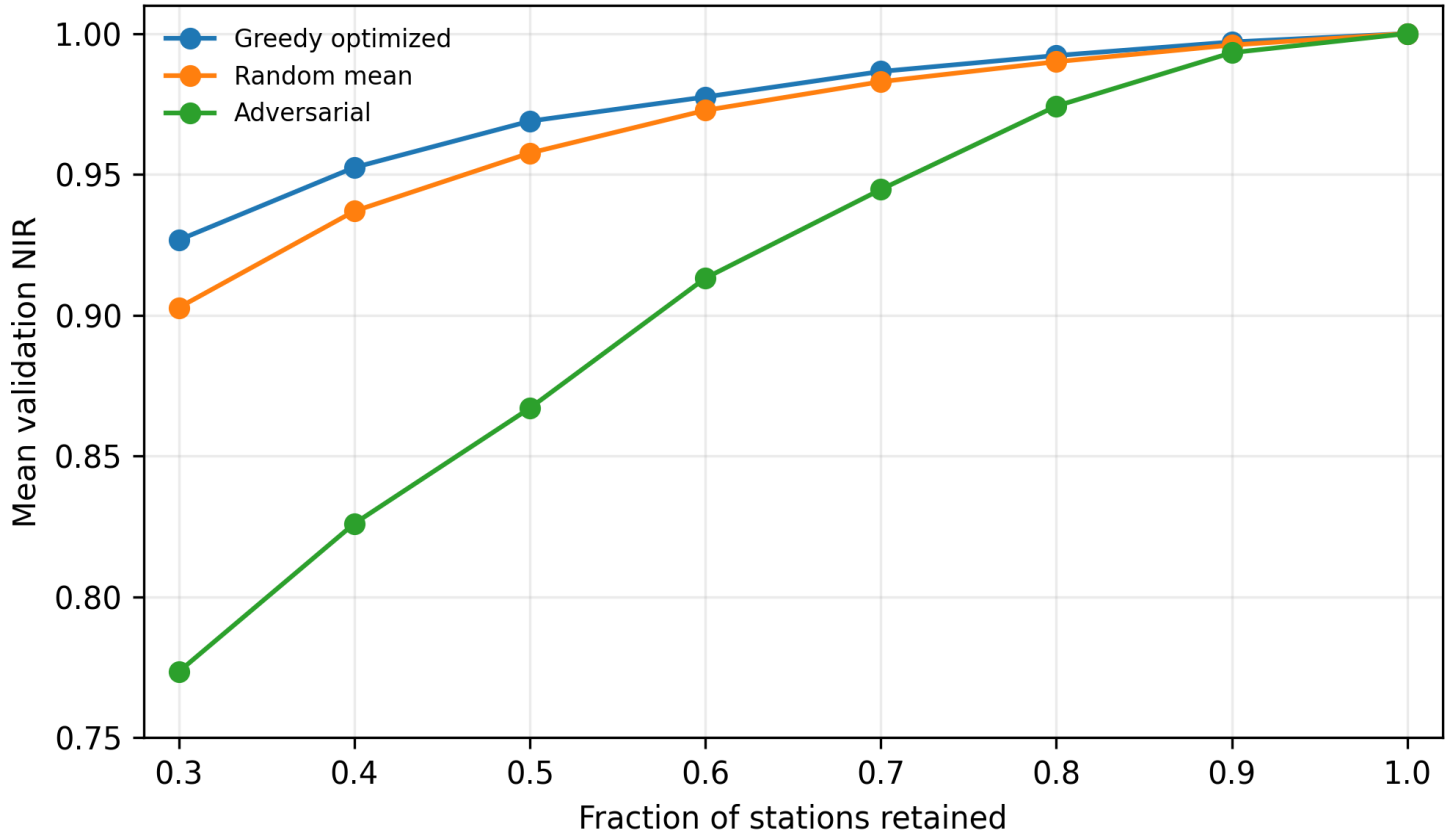


Fig. 1. Cross-basin mean validation performance

At 50% retained

0.9689 greedy

0.9576 random

0.8670 adversarial

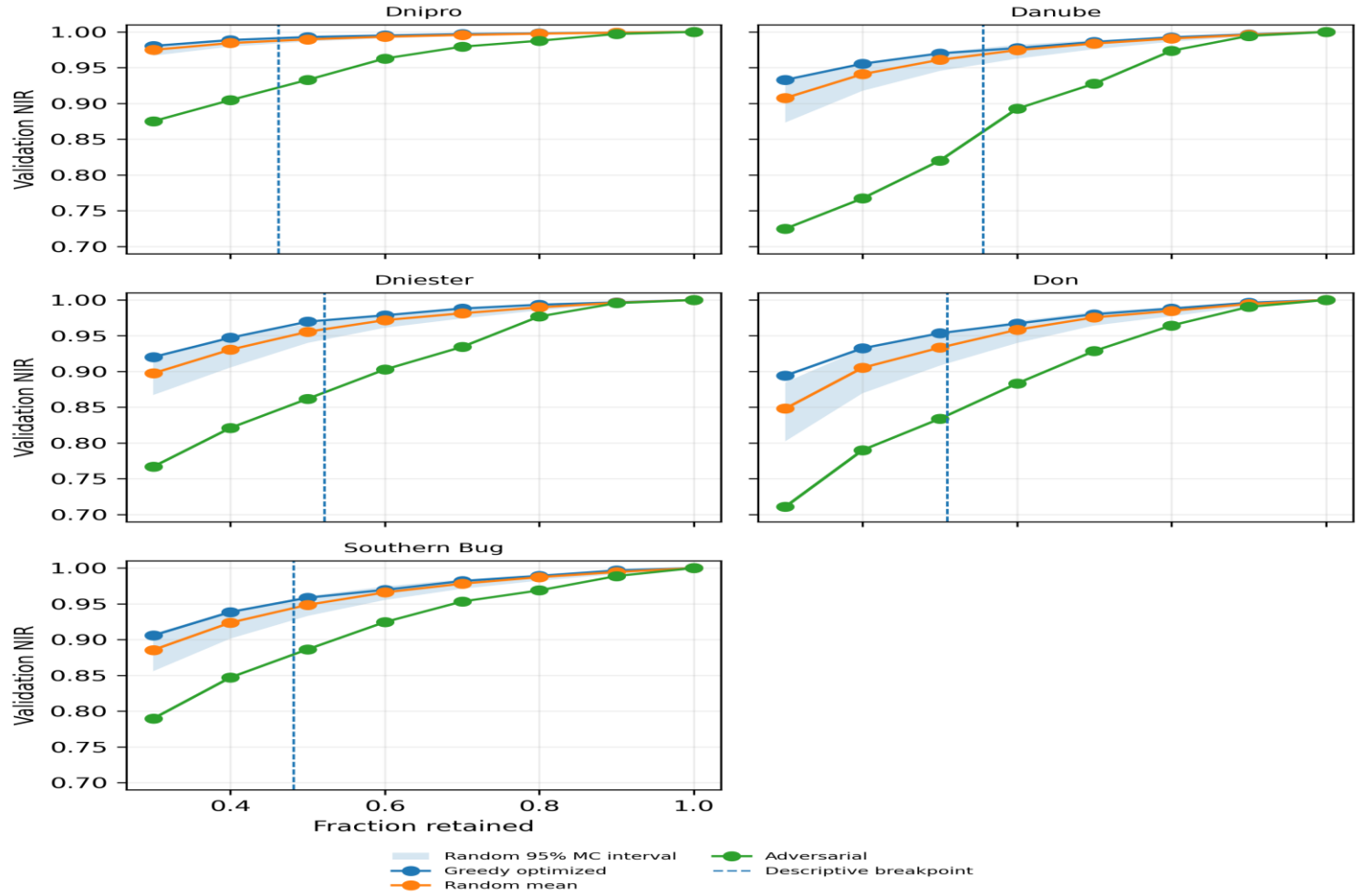
At 30% retained

0.9267 greedy
0.9027 random
0.7734 adversarial

The smaller the network, the more station identity matters.

Resilience is basin-specific — and the breakpoints are descriptive

Five basins show similar nonlinearity but different levels of robustness.



Descriptive slope changes

Dnipro	46%
Danube	56%
Dniester	52%
Don	51%
Southern Bug	48%

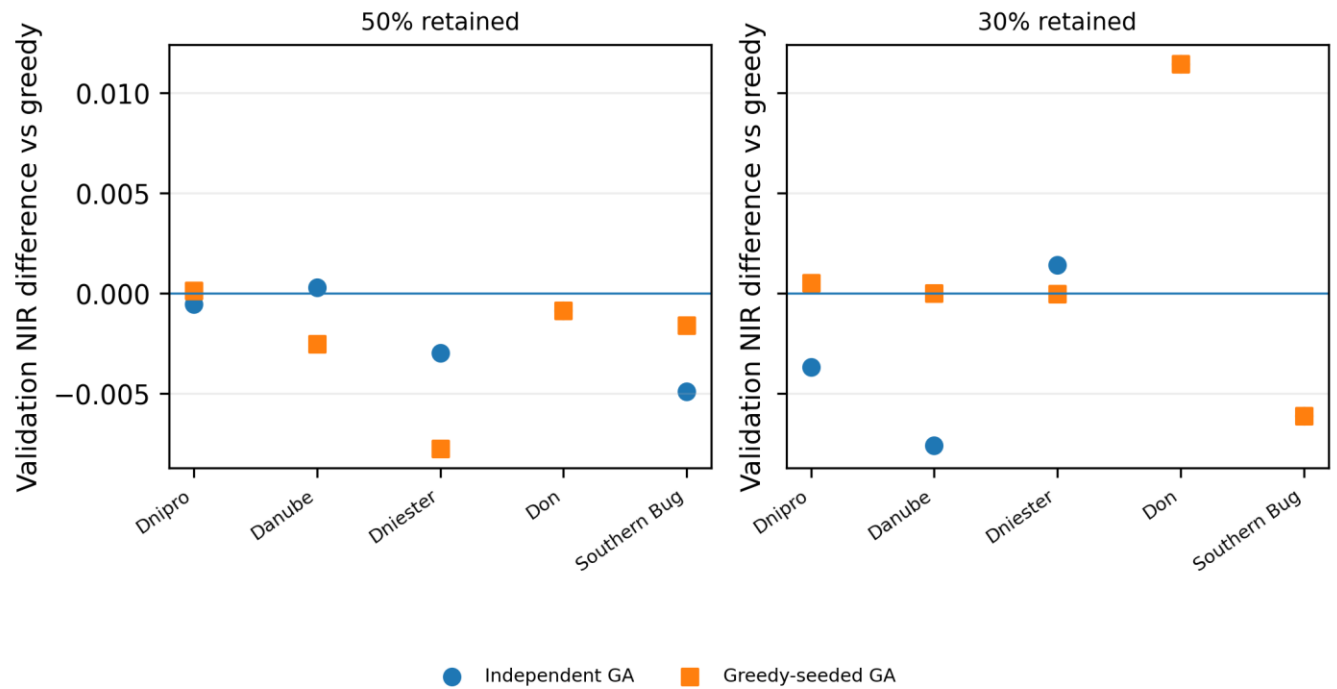
NOT operational thresholds

They do not define a safe or recommended minimum network size.

Regulatory, hotspot, transboundary and seasonal objectives remain outside this breakpoint metric.

Does a stronger optimizer produce a better future network?

Genetic-algorithm search improves the training objective — but not consistently the validation period.



GA validation NIR minus greedy validation NIR

Cross-basin validation means

50% retained

Greedy	**0.9689**
Independent GA	**0.9671**
Greedy-seeded GA	**0.9664**

30% retained

Greedy 0.9267
Independent GA 0.9258
Greedy-seeded GA 0.9278

Better calibration ≠ better temporal transfer

Stress test: optimality can change when the data structure changes

2023–2024 selection → 2025 validation, using a stricter contemporary cohort.

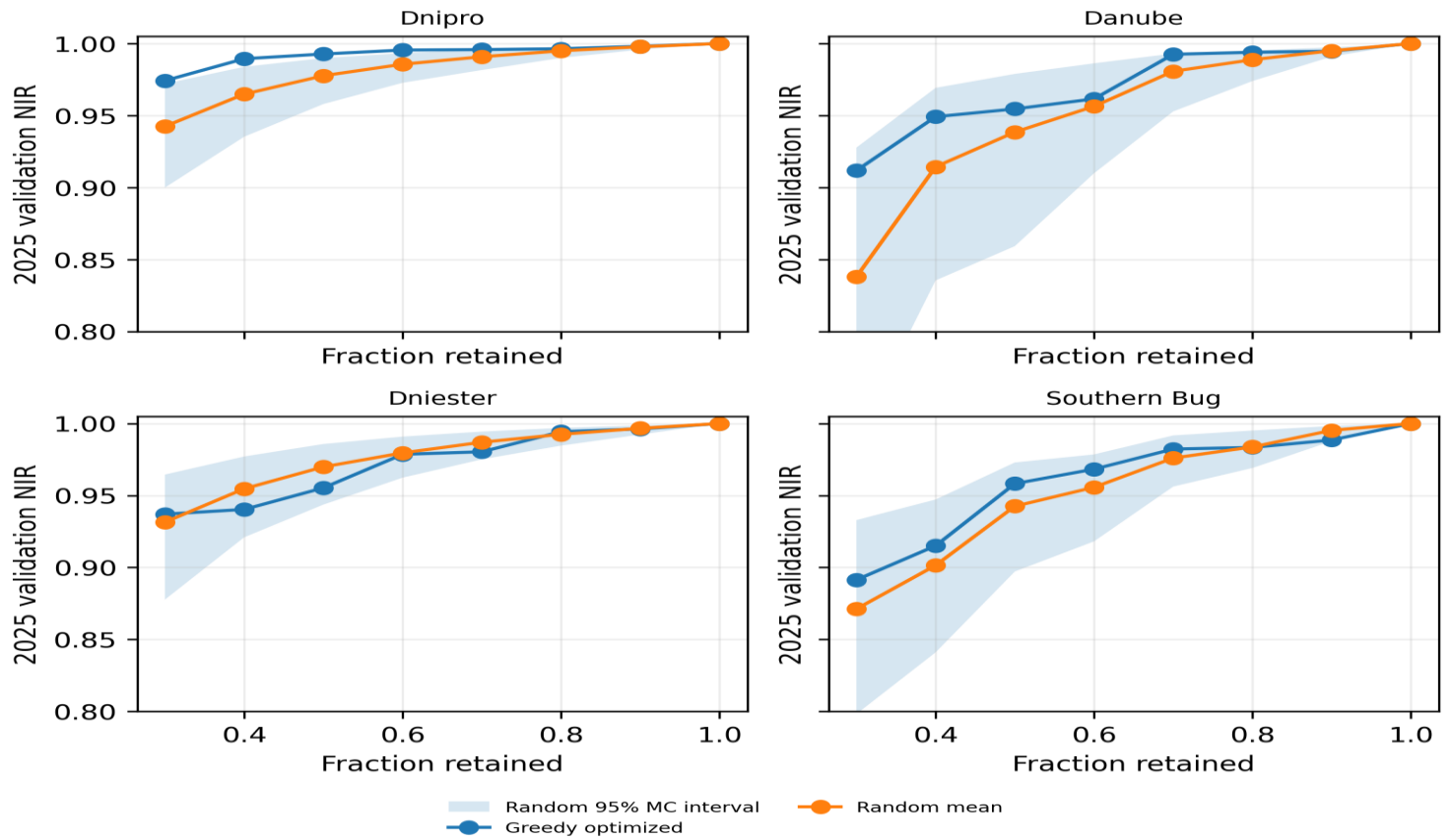


Fig. 4. Structural and temporal stress test

2025 at 50% retention

Dnipro	**0.9927**	> random
Danube	**0.9546**	> random
Southern Bug	**0.9582**	> random
Dniester	**0.9553**	< random

Adaptive management

Re-evaluate the network periodically.

Primary-study limits

343-station stable cohort • annual medians • marginal distributions • spatial coverage is diagnostic, not optimized

Three take-home messages

1

Redundancy exists

Moderate reduction can preserve high marginal information retention.

2

Transferability matters

A network optimized on historical data may not be best in a later period.

3

Optimization should be adaptive

Use NIR for screening, then combine it with spatial, legal, seasonal and hotspot constraints.

Reproducible research

DATASET	10.5281/zenodo.22722596
CURATION	10.5281/zenodo.22729481
ANALYSIS	10.5281/zenodo.23151239

The goal is not fewer stations. The goal is resilient decisions with defensible information.

Thank you

vitalii.bezsonnyi@hneu.net