BAYESIAN MATRIX-WEIGHTED GRAPHS FOR CONTEXT-DEPENDENT MULTIVARIATE RELATIONSHIPS

Anonymous authors
Paper under double-blind review

ABSTRACT

Many graph-based models involve nodes with multivariate feature vectors, so scalar edge weights are insufficient for direction-dependent interactions. We develop the Bayesian matrix-weighted graph (BMWG), which assigns each edge a symmetric positive-definite matrix W_e and infers its context-dependent posterior geometry. The geodesic distance induced by the affine-invariant Riemannian metric (AIRM) quantifies the magnitude of change, and generalized eigenvalues resolve its signed directions of strengthening and weakening. We show that BMWG achieves competitive recovery of the full graph precision matrix against fused graphical lasso, Bayesian multiple-GGM, and common principal components, while retaining identifiable matrix-weighted edges and more accurately recovering edge-level deformation directions. In Bay Area Meteostat weather data, 12-hour evolution reconfigures spatial coupling by an amount comparable to seasonal variation. In TCGA-BRCA, estrogen-receptor-associated reconfiguration concentrates on specific gene-module pairs and persists under scaffold sparsification and removal of subgroup mean differences. BMWG therefore provides a unified framework for quantifying and interpreting context-dependent multivariate reconfiguration.

1 INTRODUCTION

Classical weighted graphs assign a nonnegative scalar weight w_e to each edge $e = (i, j)$, where $x_i, x_j \in \mathbb{R}$ are scalar node values, and use the penalty $w_e(x_i - x_j)^2$. Matrix-weighted graphs generalize this construction by assigning a symmetric positive-semidefinite matrix to each edge (Tuna, 2016; Trinh et al., 2018; Hansen, 2021). When node i carries a multivariate feature vector $y_i \in \mathbb{R}^d$, the corresponding edge penalty becomes $(y_i - y_j)^\top W_e(y_i - y_j) \geq 0$. Such a matrix weight can distinguish interaction strengths along different multivariate directions, which no single scalar weight can express. A weather station reports temperature, relative humidity, wind, and pressure; a biological module comprises several genes. The relationship between two such nodes can be strong along one combination of variables and weak along another, and it can rotate when the scientific context changes, a reconfiguration that no single scalar weight can express.

We therefore represent each edge e of a fixed graph $G = (V, E)$ by a symmetric positive-definite (SPD) matrix $W_e \in \mathcal{S}^d_{++}$, where $\mathbb{S}^d_{++}$ denotes the cone of real $d \times d$ SPD matrices. We infer the posterior distribution of $W = \{W_e\}_{e \in E}$, thereby characterizing matrix-weighted graph geometries rather than scalar edge strengths.

Our aim is not to fit a graph in each context, but to characterize how the multivariate relationships it encodes change across contexts: different weather regimes, atmospheric state versus short-horizon innovation, or distinct tumor subgroups. We distinguish three complementary aspects of change: (1) *magnitude*, how far the corresponding SPD edge matrices move; (2) *evidence beyond finite-sample variation*, whether that separation exceeds the variation expected from independently fitted posteriors under no structural change; and (3) *direction*, which multivariate combinations strengthen or weaken between contexts.

The ingredients are individually classical: affine-invariant geometry on SPD matrices (Bhatia, 2007; Pennec et al., 2006), and Gaussian graphical models with cross-context comparison of covariance or precision structure (Lauritzen, 1996; Flury, 1984; Danaher et al., 2014; Peterson et al., 2015).

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

We claim no new SPD metric or generic MCMC method. Our contribution is their graph-specific synthesis into a posterior over matrix-weighted edges, and the demonstration that this representation directly characterizes edge-level directional reconfiguration. ConeMALA (Dey, 2026) provides the geometry-aware MCMC inference engine used to sample these posteriors.

Our contributions are as follows. First, we use AIRM-induced geodesic distance as an intrinsic edge effect size and a generalized eigenproblem to resolve that distance into signed strengthening and weakening directions. Second, we introduce a same-truth reference that distinguishes data-generating structural change from the finite-sample separation between independently fitted no-change posteriors. Third, we establish identifiability of the matrix-weighted edge parameterization under an explicit injectivity condition. For the incidence model used in the controlled and Meteostat studies, this condition holds for every simple graph and any node dimension, without an additional rank condition. Fourth, on a 120-condition controlled grid we compare BMWG with fused graphical lasso (Danaher et al., 2014), Bayesian multiple-GGM (Peterson et al., 2015), and common principal components (Flury, 1984), each in its native inference, and isolate the role of matrix-weighted edges through an Isotropic BMWG ablation $W_e = w_e I_d$. Finally, we evaluate the same definitions in controlled, meteorological, and biological settings, where BMWG resolves orientation changes invisible to ordinary eigenvalues, finds that 12-hour weather innovation reconfigures spatial geometry about as much as whole seasons differ, and localizes ER-associated change in TCGA-BRCA to module pairs that persist under scaffold and mean-shift perturbation.

Figure 1 summarizes the BMWG pipeline: multivariate feature vectors are represented through matrix-weighted graph edges, posterior inference produces distributions over edge geometries, and context comparisons resolve both the magnitude and direction of structural reconfiguration.

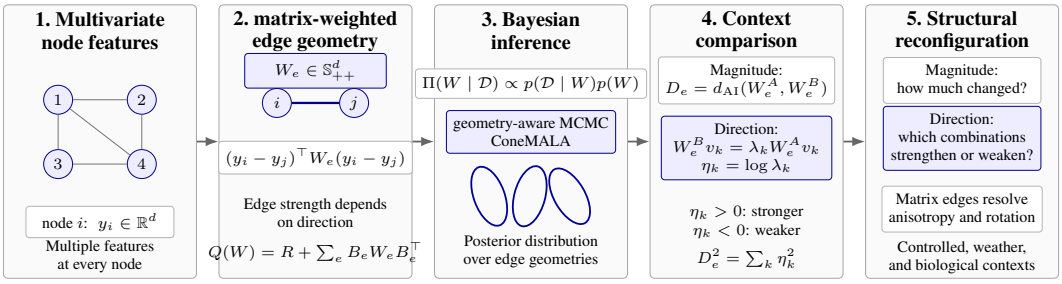


Figure 1: **Overview of Bayesian matrix-weighted graphs (BMWG).** Each node carries a multivariate feature vector and each graph edge is parameterized by an SPD matrix W_e, which assigns direction-dependent penalties to endpoint discrepancies. Bayesian inference produces a posterior over matrix-weighted graph geometries. Across contexts, affine-invariant distance quantifies the magnitude of edge deformation, while generalized eigenvalues identify multivariate directions that strengthen or weaken.

2 Bayesian Matrix-Weighted Graph Model

Let $G = (V, E)$ be a fixed undirected graph with $m = |V|$ nodes, each carrying a d-vector. Choose an arbitrary orientation for each edge, let $b_e \in \mathbb{R}^m$ be its incidence vector, and define $B_e = b_e \otimes I_d \in \mathbb{R}^{md \times d}$. For edge matrices $W = (W_e)_{e \in E}$, define

$$L(W) = \sum_{e \in E} B_e W_e B_e^\top, \qquad Q(W) = L(W) + R, \qquad R \succ 0. \tag{1}$$

For $y = (y_1^\top, \ldots, y_m^\top)^\top$,

$$y^\top L(W) y = \sum_{e=(i,j) \in E} (y_i - y_j)^\top W_e (y_i - y_j), \tag{2}$$

so W_e describes how disagreement between endpoint vectors is penalized in different directions.

Remark 1. For $W_e \in \mathbb{S}^d_{++}$, every nonzero edge discrepancy $\delta_e = y_i - y_j$ satisfies $\delta_e^\top W_e \delta_e > 0$, so W_e assigns strictly positive, direction-dependent penalties. Hence $B_e W_e B_e^\top \succeq 0$ and therefore

$L(W) \succeq 0$. The stabilizer $R \succ 0$ gives

$$Q(W) = L(W) + R \succ 0,$$

ensuring a nonsingular Gaussian precision matrix. The strict SPD constraint also places the edge weights on the open cone where the affine-invariant geometry is defined.

The inverse $Q(W)^{-1}$ is the covariance under the Gaussian working model. For nodes with different variables, Section A.6 replaces $y_i - y_j$ by $A_{e,i}y_i - A_{e,j}y_j$ in a fixed edge-specific shared space; TCGA-BRCA uses its one-dimensional specialization.

Theorem 1. *Let $R \succ 0$ be fixed and define the edge-assembly map*

$$\mathcal{L}(H) = \sum_{e \in E} B_e H_e B_e^\top.$$

Then W is globally identifiable from the law $Y \sim N(0, Q(W)^{-1})$ if and only if $\mathcal{L}$ is injective. In the incidence model $B_e = b_e \otimes I_d$, $\mathcal{L}$ is injective for every simple graph and every d. In the scalar fixed-map model $B_e = v_e$, $W_e = w_e > 0$, identifiability is equivalent to linear independence of $\{v_e v_e^\top\}_{e \in E}$.

The TCGA fixed maps satisfy the corresponding rank condition (rank 10 for ten edges). Formal statements and proofs are given in Section A.7.

For centered observations $Y_1, \ldots, Y_n \in \mathbb{R}^{md}$, let $S = n^{-1} \sum_r Y_r Y_r^\top$ and use

$$Y_r \mid W \sim N(0, Q(W)^{-1}), \qquad \ell(W) = \frac{n}{2}\{\log\det Q(W) - \operatorname{tr}(SQ(W))\} + C. \tag{3}$$

This Gaussian likelihood is a working model for second-order dependence; the inferred edges are associational, not causal. Its edgewise gradient is

$$\nabla_{W_e} \ell(W) = \frac{n}{2} B_e^\top \{Q(W)^{-1} - S\} B_e. \tag{4}$$

All edge gradients therefore depend on the common global inverse $Q(W)^{-1}$.

With independent Wishart priors $W_e \sim \mathcal{W}_d(\nu, \Psi)$, ConeMALA works relative to product affine-invariant Riemannian volume. After the Lebesgue-to-AIRM measure conversion (Appendix A; (Dey, 2026)), the intrinsic negative log density is, up to a constant,

$$\Phi(W) = \frac{n}{2}\{\operatorname{tr}(SQ(W)) - \log\det Q(W)\} - \frac{\nu}{2} \sum_{e \in E} \log\det W_e + \frac{1}{2} \sum_{e \in E} \operatorname{tr}(\Psi^{-1} W_e). \tag{5}$$

3 Posterior Geometry and Structural Reconfiguration

This section turns context-specific posteriors over matrix-weighted edges into interpretable summaries of structural reconfiguration. We first use the affine-invariant Riemannian metric to quantify how much each edge changes and generalized eigenvalues to identify which multivariate directions strengthen or weaken. We then compare candidate-context separation with an independently fitted same-truth reference to distinguish structural change from finite-sample posterior separation. Finally, Section 3.1 aggregates the learned edge penalties into an observation-level structural energy for new data.

Context- or group-specific posterior inference is standard in Bayesian multigroup graphical modeling; see, for example, (Peterson et al., 2015). For context c, let $\Pi_c(W) = p(W \mid \mathcal{D}_c)$ denote the posterior based on observations $\mathcal{D}_c$. A context may be a weather regime, season, biological group, controlled perturbation, or a different representation of the same data.

For $A, B \in \mathbb{S}^d_{++}$, the geodesic distance induced by the AIRM is

$$d_{\mathrm{AI}}(A, B) = \left\| \log(A^{-1/2} B A^{-1/2}) \right\|_F, \tag{6}$$

where $\|\cdot\|_F$ is the Frobenius norm. It is invariant to a common invertible change of coordinates. For edge e and contexts a, b we use

$$D_e^{a,b} = d_{\mathrm{AI}}(W_e^a, W_e^b) \tag{7}$$

as the posterior edge effect size. We report distances between posterior-mean blocks separately from draw-pair distances because the latter also contain posterior uncertainty from both fits.

AIRM says how much an edge changed but not how. For $A, B \in \mathbb{S}^d_{++}$, solve

$$Bv_k = \lambda_k A v_k, \qquad \eta_k = \log \lambda_k. \tag{8}$$

The $\lambda_k > 0$ are the generalized eigenvalues of the pair (B, A).

Proposition 1. *For the generalized eigenvalues in* (8)*,*

$$d_{\mathrm{AI}}(A, B)^2 = \sum_{k=1}^{d} (\log \lambda_k)^2 = \sum_{k=1}^{d} \eta_k^2. \tag{9}$$

This representation of the affine-invariant distance in terms of the generalized eigenvalues of the matrix pair (B, A) is classical (Bhatia, 2007; Förstner & Moonen, 2003). A positive η_k means that B is stronger relative to A along the corresponding generalized direction; a negative value means weaker. Different η_k may have different signs, so one matrix edge can strengthen and weaken simultaneously in different directions.

For an edge e, let $D_{e,c} := d_{\mathrm{AI}}\left(W_e^{(0)}, W_e^{(c)}\right)$ denote the AIRM distance between an independently paired draw from the baseline posterior and a draw from the posterior under candidate context c. This is a draw-level measure of the size of the inferred edge change. Likewise, let $D_{e,\mathrm{null}} = d_{\mathrm{AI}}\left(W_e^{(0)}, W_e^{(0')}\right)$, where the two posteriors are fitted to independent datasets generated from the same underlying edge matrix.

Even when the underlying graph is unchanged, there is a positive distance between a pair of posterior draws from models fitted to two independent finite datasets. Therefore $\mathbb{P}(D_{e,c} > 0)$ is not a useful equality test, see Remark 4. In the known-ground-truth structural-resolution study, we instead define

$$R_e(c) = \mathbb{P}(D_{e,c} > D_{e,\mathrm{null}} \mid \mathcal{D})\,. \tag{10}$$

Thus $D_{e,c}$ measures the magnitude of the candidate change, whereas $R_e(c)$ measures how often that change is larger than the finite-sample separation observed under no change. The prespecified rule $R_e(c) \geq 0.95$ declares a change detectable when its posterior draw-pair distance exceeds the corresponding no-change distance with at least 95% probability. This threshold is used only as a prespecified detection rule in the controlled known-ground-truth experiment and is not proposed as a universal scientific significance rule.

3.1 From posterior graph geometry to structural energy

The edge penalty introduced above, $(y_i - y_j)^\top W_e (y_i - y_j)$, quantifies how strongly a discrepancy between the multivariate feature vectors at two nodes is penalized by the learned edge coupling. We now aggregate these edgewise penalties into a graph-level structural-energy score for a new sample and connect this score directly to the posterior reference geometry.

For a reference posterior $\Pi(W \mid \mathcal{D}_{\mathrm{ref}})$, define

$$\bar{W}_e = \mathbb{E}[W_e \mid \mathcal{D}_{\mathrm{ref}}], \qquad \bar{Q} = Q(\bar{W}).$$

For a graph-level sample $y = (y_1^\top, \ldots, y_m^\top)^\top$, where $y_i \in \mathbb{R}^d$ is the feature vector at node i, define the structural-energy score

$$\mathcal{E}_{\bar{Q}}(y) = y^\top \bar{Q} y = y^\top R y + \sum_{e=(i,j)\in E} (y_i - y_j)^\top \bar{W}_e (y_i - y_j). \tag{11}$$

Thus $\mathcal{E}_{\bar{Q}}(y)$ measures how strongly the new sample disagrees with the multivariate edge geometry learned from the reference data. For the fixed-map model of Section A.6, $y_i - y_j$ is replaced by the corresponding edge discrepancy $\delta_e(y)$.

Theorem 2. *For the reference posterior,* $\mathbb{E}[Q(W) \mid \mathcal{D}_{\mathrm{ref}}] = \bar{Q}$*, and consequently, for every fixed graph-level sample* y*,*

$$\mathbb{E}\left[y^\top Q(W) y \mid \mathcal{D}_{\mathrm{ref}}\right] = \mathcal{E}_{\bar{Q}}(y).$$

Proof. Since $Q(W) = R + \sum_{e \in E} B_e W_e B_e^\top$ is affine in the edge blocks,

$$\mathbb{E}[Q(W) \mid \mathcal{D}_{\mathrm{ref}}] = R + \sum_{e \in E} B_e \bar{W}_e B_e^\top = Q(\bar{W}) = \bar{Q}.$$

Therefore, $\mathbb{E}[y^\top Q(W) y \mid \mathcal{D}_{\mathrm{ref}}] = y^\top \bar{Q} y = \mathcal{E}_{\bar{Q}}(y)$. □

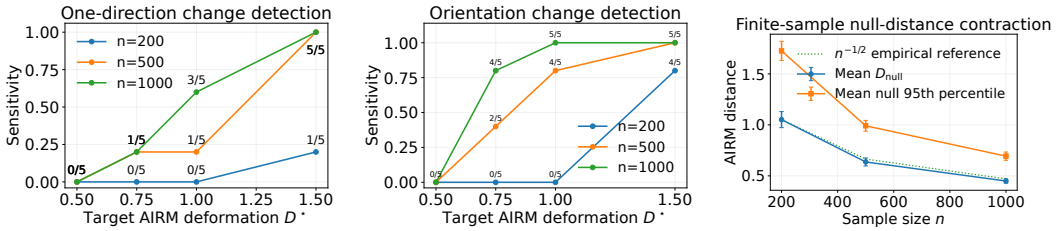
Thus (11) is exactly the posterior expected structural energy rather than an arbitrary plug-in score. Averaging $y^\top Q(W) y$ over posterior draws targets the same quantity and therefore adds no new posterior information for this quadratic functional.

4 Experiments

In this section, we perform three complementary empirical studies with distinct roles. The known-ground-truth structural-resolution study measures finite-sample resolution and quantifies the separation expected under no change. The Meteostat analysis examines how spatial matrix-weighted graph geometry varies across physical and observational contexts. The TCGA-BRCA population analysis studies the reconfiguration of the module-pairs associated with ER. All main posterior fits use multiple chains and the prespecified diagnostic gate defined in Section B.

4.1 Known-ground-truth structural resolution

We use five modules of dimension $d = 3$ connected by six fixed edges. For each of five independent generating seeds, a fresh baseline truth is sampled and then held fixed while controlled perturbations are applied. We use $n \in \{200, 500, 1000\}$ and apply changes with target AIRM distances $D^\star \in \{0.5, 0.75, 1, 1.5\}$. A one-direction perturbation changes one generalized direction. An orientation perturbation uses an orthogonal congruence $RAR^\top$ and is tuned to the same AIRM magnitude while preserving trace, determinant, and ordinary eigenvalues.



(a) One-direction change. (b) Orientation change. (c) Null contraction.

Figure 2: Panels (A-B) show the sensitivity of the prespecified $R_e \geq 0.95$ rule across five independently generated synthetic data replicates for the one-direction and orientation perturbations. Panel (C) shows the separation between two no-change posteriors fitted to independent datasets and how this separation decreases as the sample size increases.

Figure 2 summarizes the finite-sample resolution of the controlled experiment. No unchanged edge is falsely detected in any reported condition (specificity 1.00), while orientation and one-direction deformations with the same target AIRM magnitude can have markedly different finite-sample detection rates; full results are in Section C.1. Detection improves with both sample size and prespecified deformation magnitude, while the orientation perturbation is more readily resolved than the one-direction perturbation at intermediate effect

The generalized eigenvector directions recover how the edge changed. At $D^\star = 1$, mean direction alignment improves from 0.853 to 0.956 for the one-direction perturbation and from 0.928 to 0.978 for the orientation perturbation as n increases from 200 to 1000. Sign-recovery probability is essentially one at the largest sample size. Full detection tables, magnitude-recovery results, and direction/sign analyses are reported in Section C.

4.2 Comparison with native multigroup estimators

To determine whether the gains come from the matrix-weighted graph representation rather than only from the sampling method, we compare BMWG with three established approaches using their standard fitting procedures. Fused graphical lasso (FGL) (Danaher et al., 2014) estimates a separate 15×15 precision matrix for each context while encouraging sparsity and similarity between the two matrices. We compare its estimated precision matrices and precision change with the corresponding BMWG quantities. The Bayesian multiple-GGM model of (Peterson et al., 2015) also estimates context-specific precision matrices, using G-Wishart MCMC. We compare its posterior-mean precision matrices and precision changes with BMWG, and separately examine its posterior probabilities of changes in edge support. Common principal components (CPC) (Flury, 1984) operates on the three-dimensional contrast associated with the changed edge. We therefore compare CPC and BMWG on recovery of that edge's covariance, covariance change, and leading deformation direction.

All methods use the same training observations, while fresh common validation/test samples are used only for tuning and evaluation. The full comparison spans five independent generating seeds, three sample sizes $n \in \{200, 500, 1000\}$, four target deformation magnitudes $D^\star \in \{0.5, 0.75, 1, 1.5\}$, and two deformation families (one-direction and orientation), giving $5 \times 3 \times 4 \times 2 = 120$ matched conditions. Thus the study compares complete statistical procedures, rather than attributing differences solely to parameterization.

For the candidate context, $Q_1^\star$ denotes the data-generating precision matrix and $\widehat{Q}_1$ its estimate; the corresponding precision change is

$$\Delta Q^\star = Q_1^\star - Q_0^\star, \ \widehat{\Delta Q} = \widehat{Q}_1 - \widehat{Q}_0. \tag{12}$$

For the changed edge $e = (i, j)$, let $\delta_e(Y) = B_e^\top Y = Y_i - Y_j$ denote its d-dimensional contrast, with context-specific covariance

$$\Sigma_{e,c} = \mathrm{Cov}\{\delta_e(Y)\} = B_e^\top Q_c^{-1} B_e, \qquad \Delta\Sigma_e = \Sigma_{e,1} - \Sigma_{e,0}.$$

Here $\Sigma_{e,c}^\star$ and $\Delta\Sigma_e^\star$ denote the corresponding data-generating quantities. "Axis error" is the absolute principal angle, in degrees, between the estimated and data-generating leading deformation directions. NLL denotes Gaussian negative log-likelihood, so lower values indicate better predictive fit.

At $D^\star = 1$ and $n = 1000$, BMWG has the smallest candidate-precision error for both perturbation families (Table 1, Panel A), while FGL is more accurate for the global ΔQ error in the orientation case. Across the full 120-condition grid, BMWG outperforms multi-GGM on all matched precision-recovery and predictive comparisons, outperforms FGL on 109/120 candidate-precision and 120/120 predictive-NLL comparisons, and outperforms CPC on 115/120 directional-recovery comparisons. FGL remains more accurate for global ΔQ over part of the orientation-change regime. Full ΔQ, localization, multi-GGM support-probability, and MCMC-stability results are reported in Section C.3.

Matrix-Weighted Representation Ablation. To isolate the contribution of matrix-weighted edges, we additionally fit Isotropic BMWG, which retains the same graph scaffold and Gaussian precision construction but restricts $W_e = w_e I_d$, $w_e > 0$. BMWG has lower candidate-edge error in all $120/120$ matched conditions and lower deformation error in $111/120$, including all $80/80$ conditions with $n \geq 500$. The contrast is representational rather than a sampling failure: all isotropic fits satisfy the diagnostic gate. For the isospectral orientation family, the oracle isotropic representation gap is 1, so no component of the data-generating additive deformation lies in the isotropic edge subspace. Full derivations, numerical results, and recovery curves are reported in Section C.4.

4.3 Context-dependent spatial geometry in Meteostat

We use one year (2023) from eight Bay Area weather stations (Meteostat, 2026). Each station carries $d = 4$ variables such as temperature, relative humidity, wind speed, and pressure, and a fixed symmetrized 3-nearest-neighbor scaffold defines 16 undirected edges: the initial $8 \times 3 = 24$ directed neighbor links reduce to 16 distinct edges after reciprocal links are merged. Hourly observations are aggregated to 3-hour means. All preprocessing parameters are estimated from the training period only; exact transformations are reported in Section D.

Table 1: Panels A-B report mean ± sample standard deviation across five independent replicates at $D^\star = 1$ and $n = 1000$; lower is better. Panel C reports pairwise win counts over the full 120-condition grid: an entry $a/120$ means that BMWG has strictly smaller error (or NLL) than the indicated comparator in a of the 120 matched conditions.

Panel A: global precision recovery and predictive fit

Family	Method	$\|\widehat{Q}_1 - Q_1^\star\|_F / \|Q_1^\star\|_F$	$\|\widehat{\Delta Q} - \Delta Q^\star\|_F / \|\Delta Q^\star\|_F$	Test NLL
One-direction	BMWG	**0.066 ± 0.014**	**0.303 ± 0.068**	**15.752 ± 0.299**
	FGL	0.168 ± 0.037	0.743 ± 0.066	15.782 ± 0.297
	multiGGM	0.086 ± 0.013	0.416 ± 0.076	15.787 ± 0.295
Orientation	BMWG	**0.075 ± 0.010**	0.829 ± 0.212	**15.859 ± 0.278**
	FGL	0.085 ± 0.005	**0.686 ± 0.099**	15.883 ± 0.280
	multiGGM	0.101 ± 0.009	1.074 ± 0.221	15.892 ± 0.274

Panel B: changed-edge multivariate geometry

Family	Method	$\|\widehat{\Delta\Sigma}_e - \Delta\Sigma_e^\star\|_F / \|\Delta\Sigma_e^\star\|_F$	Axis error (deg.)	Edge NLL
One-direction	BMWG	**0.210 ± 0.065**	**0.70 ± 0.58**	**3.208 ± 0.167**
	CPC	0.642 ± 0.141	20.07 ± 19.14	3.214 ± 0.170
Orientation	BMWG	**0.358 ± 0.088**	**4.05 ± 4.07**	**3.643 ± 0.157**
	CPC	1.033 ± 0.019	25.29 ± 5.90	3.662 ± 0.157

Panel C: matched conditions in which BMWG performs better

Comparator	Context/candidate error	Change error	NLL / Direction
FGL	109/120	62/120	120/120 / –
multiGGM	120/120	120/120	120/120 / –
CPC	98/120	111/120	112/120 / 115/120

For FGL and multiGGM, "context/candidate error" and "change error" are the relative Frobenius errors of Q_1 and ΔQ, respectively. For CPC, Panel C reports changed-edge covariance and covariance-change errors; NLL is computed on the edge contrast, and "Direction" denotes leading-axis error. In the orientation family, FGL can recover the aggregate ΔQ more accurately, whereas BMWG is more accurate for the edge-level deformation direction.

We compare three representations of the same spatial field: standardized raw state, anomaly after removing the average time-of-day pattern estimated from the training data, and 12-hour innovation $I_t = Y_{t+12\mathrm{h}} - Y_t$. For contexts a, b, summarize global graph separation by

$$\bar{D}_{a,b} = \frac{1}{|E|} \sum_{e \in E} d_{\mathrm{AI}}\Big(\bar{W}_e^{(a)}, \bar{W}_e^{(b)}\Big). \tag{13}$$

The mean separations are

$$\bar{D}_{\mathrm{raw,anomaly}} = 0.887, \qquad \bar{D}_{\mathrm{raw,innovation}} = 3.184, \qquad \bar{D}_{\mathrm{anomaly,innovation}} = 3.576.$$

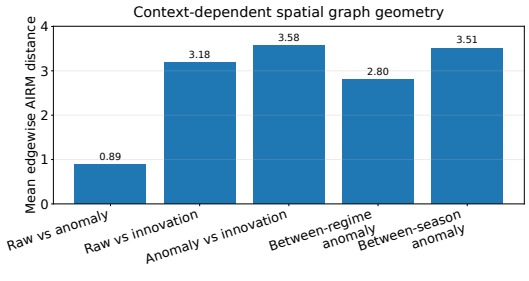
For comparison, mean anomaly-graph separation is 2.800 across training-derived regimes and 3.511 across seasons. Thus replacing atmospheric state by its 12-hour evolution changes the learned spatial geometry about as much as whole seasons differ, whereas removing the diurnal mean changes it relatively little.

The large separation is directional rather than a complete loss of common structure. For each edge, the leading-mode share is $\lambda_{\max}(\bar{W}_e)/\operatorname{tr}(\bar{W}_e)$. The leading ordinary eigenmode remains pressure-dominated on essentially every edge and accounts, on average, for 84.6%, 84.8%, and 89.2% of the total eigenvalue mass in the raw, anomaly, and innovation geometries, respectively. In contrast, the generalized directions reveal strong edge-specific strengthening and weakening. Full directional, effective-resistance, and chronological OOD analyses are reported in Section D.

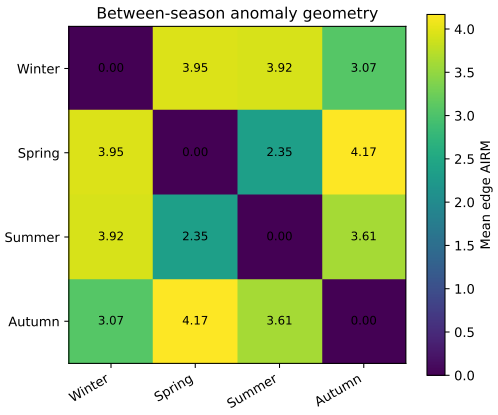
4.4 ER-ASSOCIATED GENE-MODULE RECONFIGURATION

We analyze 1097 unique primary TCGA-BRCA tumors from UCSC Xena (The Cancer Genome Atlas Network, 2012; Goldman et al., 2020; UCSC Xena, 2026). Five prespecified three-gene groups serve as compact representatives of Immune, Metabolic, Structural/EMT, Signalling, and Epigenetic programs (Ahn et al., 2017; Wong et al., 2022; Liberzon et al., 2015; Sanchez-Vega et al., 2018; Xu et al., 2010; Oh et al., 2017). These labels organize the 15-gene testbed and are not claimed to be complete pathways or learned modules.

Because different modules contain different genes, we avoid arbitrary coordinate matching. Before using ER labels, a regularized pairwise canonical correlation analysis (CCA) is fit for each of the ten



(a) Global/context separation.



(b) Seasonal anomaly geometry.

Figure 3: Mean edgewise AIRM separation across global representations, training-derived regimes, and seasons. The largest seasonal separation is autumn–spring.

module pairs and its leading shared coordinate is frozen. For edge $e = (i, j)$,

$$s_{e,i}^{(k)} = a_{e,i}^{\top} y_i^{(k)}, \qquad s_{e,j}^{(k)} = a_{e,j}^{\top} y_j^{(k)}, \qquad \delta_e(y^{(k)}) = s_{e,i}^{(k)} - s_{e,j}^{(k)}.$$

The primary scaffold is K_5, so all ten module pairs are present simultaneously. The comparison uses $n_{\mathrm{ER}-} = 179$ and $n_{\mathrm{ER}+} = 601$; details of the fixed-map construction and preprocessing are in Sections A.6 and E.

For each scalar edge, define

$$D_e = \left| \log \frac{\bar{w}_e^{+}}{\bar{w}_e^{-}} \right|, \qquad \eta_e = \log \frac{\bar{w}_e^{+}}{\bar{w}_e^{-}}. \tag{14}$$

The mean edge AIRM over K_5 is 0.537, and six of ten 90% signed-change intervals exclude zero (Figure 4; Table 12). Thus the ER-associated reconfiguration is distributed across several module pairs but heterogeneous in magnitude rather than concentrated on a single edge. Structural/EMT–Signalling shows the largest change, with $\eta_e = 1.625$ and 90% interval $[1.173, 2.314]$; since $e^{1.625} \approx 5.08$, its posterior-mean edge weight is approximately 5.08 times larger in ER-positive than in ER-negative tumors.

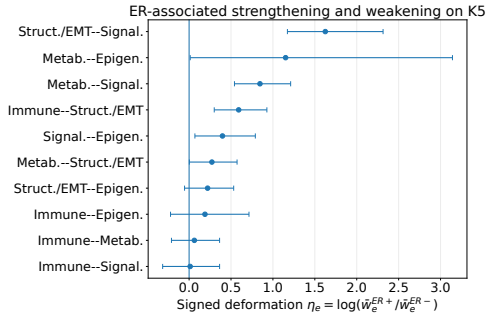


Figure 4: ER-associated signed change $\eta_e = \log(\bar{w}_e^{+}/\bar{w}_e^{-})$ for the ten module pairs; bars are 90% paired-posterior intervals. Six intervals exclude zero.

Because BMWG conditions on a fixed scaffold, we refit all 12 labeled five-cycle scaffolds C_5, each retaining five of the ten K_5 edges. For each edge, its K_5 distance agrees closely with the average over the six cycles containing it (Pearson $r = 0.990$, Spearman $\rho = 0.988$); the Structural/EMT–Signalling edge ranks first in all six cycles in which it appears. Removing ER-specific means retains 94.1% of the mean edge separation and preserves all ten edge ranks (Spearman $\rho = 1.000$). Thus the leading ER-associated reconfigurations persist under substantial scaffold sparsification and removal of subgroup mean differences.

5 Discussion and limitations

The experiments show that matrix-weighted edge inference captures context-dependent changes that scalar edge summaries cannot represent. The controlled study illustrates why both deformation magnitude and direction matter. AIRM quantifies how far an edge moves, while generalized eigenvalues resolve that change into signed strengthening and weakening directions. BMWG recovers orientation changes even when trace, determinant, and ordinary eigenvalues are unchanged, and the Isotropic BMWG ablation shows that this advantage is representational rather than merely a consequence of the sampling procedure. Native-method comparisons further show that BMWG remains competitive for global precision and predictive recovery while providing substantially more accurate edge-level deformation geometry.

The two real-data studies show complementary forms of structural reconfiguration. In Meteostat, replacing atmospheric state by its 12-hour evolution changes the learned spatial geometry by an amount comparable to seasonal variation. This change is primarily directional rather than a complete loss of common structure: the leading ordinary mode remains pressure-dominated, while generalized eigenvector directions reveal edge-specific strengthening and weakening. In TCGA-BRCA, ER-associated reconfiguration is not confined to a single module pair: six of the ten edges on the complete K_5 scaffold have 90% signed-change intervals excluding zero. The strongest changes remain stable under substantial scaffold sparsification and removal of subgroup mean differences. Together, these results show that posterior matrix-weighted edge geometry can reveal both the magnitude and the multivariate direction of context-dependent change, as well as whether that reconfiguration is localized or distributed across a network.

Graph topology is fixed rather than learned, so inference is conditional on the supplied scaffold. The Gaussian precision likelihood is a working model for second-order dependence and does not imply causal relationships. The controlled detection boundary depends on the chosen graph, prior, sample sizes, and deformation families. TCGA-BRCA analysis is restricted to the specified 15-gene system and requires external-cohort validation before broader biological conclusions are drawn. Finally, $\widehat{R}$ and ESS support numerical stability of the sampled posteriors but do not provide a finite-time MCMC convergence guarantee.

Reproducibility Statement

All experiments use fixed graph scaffolds, explicit priors, archived random seeds, and prespecified diagnostic gates. The controlled study records all generating seeds and nested sample sizes; Meteostat records chronological splits, transforms, climatology, and graph construction; TCGA records the cohort, gene list, normalization, frozen CCA maps, topology sensitivity, and mean-removal sensitivity. Exact sampler configurations and additional results are in the appendices.

AI Use Statement

Generative AI tools, primarily ChatGPT (OpenAI), were used as an interactive research and writing assistant during preparation of this work. They supported iterative manuscript revision, code development and debugging, controlled synthetic experiments, and literature discovery and preliminary synthesis for the comparative study. The central research questions and scientific direction originated with the author. The theoretical development grew from the author's mathematical questions and was developed iteratively through interaction with the AI.

The author read the cited papers directly, reran the reported experiments, verified the numerical results against generated outputs, and determined the final scientific interpretations and claims. All AI-assisted material was reviewed and revised by the author, who takes full responsibility for the final paper.

References

Ryuhjin Ahn, Valérie Sabourin, Alicia M. Bolt, et al. The Shc1 adaptor simultaneously balances Stat1 and Stat3 activity to promote breast cancer immune suppression. *Nature Communications*, 8: 14638, 2017. doi: 10.1038/ncomms14638.

Rajendra Bhatia. *Positive Definite Matrices*. Princeton University Press, 2007.

Patrick Danaher, Pei Wang, and Daniela M. Witten. The joint graphical lasso for inverse covariance estimation across multiple classes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(2):373–397, 2014. doi: 10.1111/rssb.12033.

Papri Dey. Geometry-aware Langevin sampling for matrix-valued graph learning. arXiv:2603.24913v2, 2026. URL https://arxiv.org/abs/2603.24913v2.

Bernhard N. Flury. Common principal components in k groups. *Journal of the American Statistical Association*, 79(388):892–898, 1984. doi: 10.1080/01621459.1984.10477108.

Wolfgang Förstner and Boudewijn Moonen. A metric for covariance matrices. In Erik W. Grafarend, Friedrich W. Krumm, and Volker S. Schwarze (eds.), *Geodesy—The Challenge of the 3rd Millennium*, pp. 299–309. Springer, Berlin, Heidelberg, 2003. doi: 10.1007/978-3-662-05296-9_31.

Chris Godsil and Gordon Royle. *Algebraic Graph Theory*, volume 207 of *Graduate Texts in Mathematics*. Springer, New York, 2001. doi: 10.1007/978-1-4613-0163-9.

Mary J. Goldman, Brian Craft, Mim Hastie, Kristupas Repečka, Fran McDade, Akhil Kamath, Ayan Banerjee, Yunhai Luo, Dave Rogers, Angela N. Brooks, Jingchun Zhu, and David Haussler. Visualizing and interpreting cancer genomics data via the Xena platform. *Nature Biotechnology*, 38:675–678, 2020. doi: 10.1038/s41587-020-0546-8.

Jakob Hansen. Expansion in matrix-weighted graphs. *Linear Algebra and its Applications*, 630: 252–273, 2021. doi: 10.1016/j.laa.2021.08.009.

Roger A. Horn and Charles R. Johnson. *Topics in Matrix Analysis*. Cambridge University Press, Cambridge, 1991. doi: 10.1017/CBO9780511840371.

Steffen L. Lauritzen. *Graphical Models*. Oxford University Press, 1996.

Arthur Liberzon, Chet Birger, Helga Thorvaldsdóttir, Mahmoud Ghandi, Jill P. Mesirov, and Pablo Tamayo. The molecular signatures database (MSigDB) hallmark gene set collection. *Cell Systems*, 1(6):417–425, 2015.

Meteostat. Open weather and climate data. https://meteostat.net/, 2026.

Maher Moakher. A differential geometric approach to the geometric mean of symmetric positive-definite matrices. *SIAM Journal on Matrix Analysis and Applications*, 26(3):735–747, 2005. doi: 10.1137/S0895479803436937.

Robb J. Muirhead. *Aspects of Multivariate Statistical Theory*. John Wiley & Sons, New York, 1982. doi: 10.1002/9780470316559.

Tae Gyu Oh, Shu-Ching M. Wang, and George E. O. Muscat. Therapeutic implications of epigenetic signaling in breast cancer. *Endocrinology*, 158(3):431–447, 2017. doi: 10.1210/en.2016-1716.

Xavier Pennec, Pierre Fillard, and Nicholas Ayache. A Riemannian framework for tensor computing. *International Journal of Computer Vision*, 66(1):41–66, 2006. doi: 10.1007/s11263-005-3222-z.

Christine B. Peterson, Francesco C. Stingo, and Marina Vannucci. Bayesian inference of multiple Gaussian graphical models. *Journal of the American Statistical Association*, 110(509):159–174, 2015. doi: 10.1080/01621459.2014.896806.

Håvard Rue and Leonhard Held. *Gaussian Markov Random Fields: Theory and Applications*. Chapman & Hall/CRC, 2005.

F. Sanchez-Vega et al. Oncogenic signaling pathways in The Cancer Genome Atlas. *Cell*, 173(2): 321–337.e10, 2018.

Lene Theil Skovgaard. A Riemannian geometry of the multivariate normal model. *Scandinavian Journal of Statistics*, 11(4):211–223, 1984.

The Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature*, 490(7418):61–70, 2012. doi: 10.1038/nature11412.

Minh Hoang Trinh, Chuong Van Nguyen, Young-Hun Lim, and Hyo-Sung Ahn. Matrix-weighted consensus and its applications. *Automatica*, 89:415–419, 2018. doi: 10.1016/j.automatica.2017.12.024.

S. Emre Tuna. Synchronization under matrix-weighted Laplacian. *Automatica*, 73:76–81, 2016. doi: 10.1016/j.automatica.2016.06.012.

UCSC Xena. TCGA Breast Cancer (BRCA) Gene Expression RNA-seq Dataset (HiSeqV2). UCSC Xena TCGA Hub, 2026. Data file: TCGA.BRCA.sampleMap/HiSeqV2.gz.

Grace L. Wong, Sara G. Manore, Daniel L. Doheny, and Hui-Wen Lo. STAT family of transcription factors in breast cancer: Pathogenesis and therapeutic opportunities and challenges. *Seminars in Cancer Biology*, 86(Pt 3):84–106, 2022. doi: 10.1016/j.semcancer.2022.08.003.

Jinhua Xu, Yinghua Chen, and Olufunmilayo I. Olopade. MYC and breast cancer. *Genes & Cancer*, 1(6):629–640, 2010. doi: 10.1177/1947601910378691.

A Theory and Derivations

A.1 Sufficiency of the empirical second moment

Conditional on the edge matrices $W = \{W_e\}_{e\in E}$, the centered observations are independent with

$$Y_r \mid W \sim \mathcal{N}\big(0, Q(W)^{-1}\big), \qquad r = 1, \dots, n,$$

where $Q(W) = R + \sum_{e\in E} B_e W_e B_e^\top$. Writing $Q = Q(W)$ for brevity, their joint likelihood is

$$p(Y_1, \dots, Y_n \mid W) \propto (\det Q)^{n/2} \exp\left\{-\frac{1}{2}\sum_{r=1}^{n} Y_r^\top Q Y_r\right\}.$$

Since

$$\sum_{r=1}^{n} Y_r^\top Q Y_r = \sum_{r=1}^{n} \operatorname{tr}(Q Y_r Y_r^\top) = n \operatorname{tr}(SQ), \qquad S = \frac{1}{n}\sum_{r=1}^{n} Y_r Y_r^\top,$$

the likelihood becomes

$$p(Y_1, \dots, Y_n \mid W) \propto \exp\left\{\frac{n}{2}\left[\log\det Q(W) - \operatorname{tr}\{SQ(W)\}\right]\right\}.$$

Thus the observations enter the parameter-dependent part of the likelihood only through $\sum_r Y_r Y_r^\top = nS$. By the Fisher–Neyman factorization theorem, S is therefore sufficient for W under the centered Gaussian model. Consequently, conditional on n, replacing the individual observations by S loses no likelihood information about the edge matrices.

This sufficiency also gives a useful computational reduction. The empirical second moment S is formed once before posterior sampling, and subsequent likelihood and score evaluations depend on the data only through S rather than through all n observations. Thus increasing the number of observations does not require revisiting the full dataset at every MCMC iteration.

A.2 All-edge score

Using $Q(W) = R + \sum_{f\in E} B_f W_f B_f^\top$, a perturbation H_e of the edge block W_e gives

$$D_{W_e} Q(W)[H_e] = B_e H_e B_e^\top.$$

Using the matrix differential identities

$$d \log\det Q = \operatorname{tr}(Q^{-1} dQ), \qquad d\,\operatorname{tr}(SQ) = \operatorname{tr}(S\,dQ)$$

and eq (3), we get

$$D_{W_e}\ell(W)[H_e] = \frac{n}{2}\operatorname{tr}\big[\{Q^{-1} - S\} B_e H_e B_e^\top\big].$$

Using cyclic invariance of the trace,

$$D_{W_e}\ell(W)[H_e] = \operatorname{tr}\Big[\frac{n}{2}B_e^\top(Q^{-1}-S)B_e\,H_e\Big].$$

By definition, the Euclidean matrix gradient with respect to W_e is the unique symmetric matrix $\nabla_{W_e}\ell(W)$ satisfying

$$D_{W_e}\ell(W)[H_e] = \langle\nabla_{W_e}\ell(W), H_e\rangle_F = \operatorname{tr}\big[\nabla_{W_e}\ell(W)^\top H_e\big]$$

for every symmetric perturbation H_e. Since $B_e^\top(Q^{-1}-S)B_e$ is symmetric, comparison of the two expressions yields

$$\nabla_{W_e}\ell(W) = \frac{n}{2}B_e^\top(Q^{-1}-S)B_e,$$

which is eq (4).

A perturbation H_f of the edge block W_f induces the perturbation

$$H_Q := D_{W_f}Q(W)[H_f] = B_f H_f B_f^\top$$

of the full precision matrix. Applying the chain rule to the composition $W_f \mapsto Q(W) \mapsto Q(W)^{-1}$ and using

$$D_Q(Q^{-1})[H_Q] = -Q^{-1}H_Q Q^{-1},$$

we obtain

$$D_{W_f}[Q(W)^{-1}][H_f] = -Q^{-1}B_f H_f B_f^\top Q^{-1}.$$

Note that S does not depend on W_f. Therefore, we obtain

$$D_{W_f}\big[\nabla_{W_e}\ell(W)\big][H_f] = \frac{n}{2}B_e^\top D_{W_f}[Q(W)^{-1}][H_f]B_e = -\frac{n}{2}B_e^\top Q^{-1}B_f H_f B_f^\top Q^{-1}B_e.$$

For $e \neq f$ this term is generally nonzero, showing explicitly that the likelihood couples distinct edge blocks through the common global precision inverse $Q(W)^{-1}$.

A.3 Gaussian Fisher geometry and the pullback metric

For one observation with log likelihood

$$\ell_Q(y) = \frac{1}{2}\log\det Q - \frac{1}{2}y^\top Qy + C,$$

the directional score is

$$D\ell_Q[H] = \frac{1}{2}\operatorname{tr}(Q^{-1}H) - \frac{1}{2}y^\top Hy.$$

Using Gaussian fourth-moment identities gives the Fisher information metric of the multivariate normal model (Skovgaard, 1984), i.e.,

$$g_Q^{\mathrm{F}}(H,K) = \mathbb{E}[D\ell_Q[H]D\ell_Q[K]] = \frac{1}{2}\operatorname{tr}(Q^{-1}HQ^{-1}K).$$

For n independent observations this is multiplied by n.

A tangent vector $U = (U_e)_{e\in E}$ in edge space induces $L(U) = \sum_e B_e U_e B_e^\top$. Substitution into the Fisher metric yields

$$g_W^{\mathrm{pull}}(U,V) = \frac{n}{2}\operatorname{tr}\left[Q^{-1}L(U)Q^{-1}L(V)\right],$$

which contains cross-edge terms and is therefore not the block-diagonal product metric used for the proposal.

A.4 Reference measure of the product AIRM

For one SPD matrix $W \in \mathbb{S}^d_{++}$, the affine-invariant Riemannian volume is, up to a constant, (cf. (Muirhead, 1982; Pennec et al., 2006)),

$$d\mathrm{vol}_{\mathrm{AI}}(W) = (\det W)^{-(d+1)/2}\, dW,$$

where dW is Lebesgue measure on the independent symmetric coordinates. For all edges,

$$d\mathrm{vol}_{\mathrm{prod}}(W) \propto \prod_{e\in E} (\det W_e)^{-(d+1)/2}\, dW_e.$$

The Lebesgue Wishart factor contributes $(\det W_e)^{(\nu-d-1)/2}$. Converting the posterior density from Lebesgue measure to product-AIRM volume therefore multiplies it by $(\det W_e)^{(d+1)/2}$, producing the power $\frac{\nu-d-1}{2} + \frac{d+1}{2} = \frac{\nu}{2}$. This yields the $-\frac{\nu}{2}\log\det W_e$ term in eq (5).

A.5 Proof of Proposition 1

Let $C = A^{-1/2}BA^{-1/2} \in \mathbb{S}^d_{++}$. The generalized eigenvalues of (B, A) are precisely the ordinary eigenvalues λ_k of C. Write $C = U\,\mathrm{diag}(\lambda_1, \ldots, \lambda_d)U^\top$. Then $\log C = U\,\mathrm{diag}(\log\lambda_1, \ldots, \log\lambda_d)U^\top$. Orthogonal invariance of the Frobenius norm gives

$$d_{\mathrm{AI}}(A, B)^2 = \|\log C\|_F^2 = \sum_{k=1}^{d} (\log\lambda_k)^2.$$

This proves eq (9).

A.6 Fixed node-to-edge maps and unequal node dimensions

The simple edge difference $y_i - y_j$ assumes that all nodes have the same dimension and directly aligned coordinates, as in Meteostat. For heterogeneous nodes, let $y_i \in \mathbb{R}^{d_i}$ and $p = \sum_{i\in V} d_i$. For edge $e = (i, j)$, choose a shared edge-space dimension r_e and fixed linear maps

$$A_{e,i} \in \mathbb{R}^{r_e\times d_i}, \qquad A_{e,j} \in \mathbb{R}^{r_e\times d_j},$$

and define

$$\delta_e(y) = A_{e,i}y_i - A_{e,j}y_j. \tag{15}$$

Let $D_e \in \mathbb{R}^{r_e\times p}$ be the block-row operator containing $A_{e,i}$ in the columns of node i, $-A_{e,j}$ in those of node j, and zeros elsewhere, so that $D_e y = \delta_e(y)$. Define $\widetilde{B}_e = D_e^\top$. For $W_e \in \mathbb{S}^{r_e}_{++}$,

$$\mathcal{E}_e(y) = \delta_e(y)^\top W_e \delta_e(y) = y^\top \widetilde{B}_e W_e \widetilde{B}_e^\top y. \tag{16}$$

Hence

$$L_A(W) = \sum_{e\in E} \widetilde{B}_e W_e \widetilde{B}_e^\top, \qquad Q_A(W) = L_A(W) + R. \tag{17}$$

Since each summand in $L_A(W)$ is positive semidefinite and $R \succ 0$, $Q_A(W) \succ 0$.

For the same Gaussian working likelihood $Y_r \mid W \sim N(0, Q_A(W)^{-1})$, the edge gradient is

$$\nabla_{W_e}\ell(W) = \frac{n}{2}\widetilde{B}_e^\top\{Q_A(W)^{-1} - S\}\widetilde{B}_e. \tag{18}$$

Thus the shared precision solve, global cross-edge dependence, and Metropolis-Hastings correction are unchanged; only B_e is replaced by $\widetilde{B}_e$. With independent $W_e \sim \mathcal{W}_{r_e}(\nu_e, \Psi_e)$, the intrinsic negative log posterior relative to product-AIRM volume is

$$\Phi_A(W) = \frac{n}{2}\{\mathrm{tr}(SQ_A(W)) - \log\det Q_A(W)\} - \frac{1}{2}\sum_{e\in E}\nu_e \log\det W_e + \frac{1}{2}\sum_{e\in E}\mathrm{tr}(\Psi_e^{-1}W_e). \tag{19}$$

For contexts a and b,

$$W_e^b v_k = \lambda_k W_e^a v_k, \qquad \eta_k = \log\lambda_k,$$

and

$$d_{\mathrm{AI}}(W_e^a, W_e^b)^2 = \sum_{k=1}^{r_e}\eta_k^2.$$

Example. If $y_1 = (y_{11}, y_{12})^\top$, $y_2 = y_{21}$, $A_{e,1} = [1\ 2]$, and $A_{e,2} = [3]$, then

$$\delta_e(y) = y_{11} + 2y_{12} - 3y_{21}, \qquad \widetilde{B}_e = [1 \quad 2 \quad -3]^\top.$$

For a scalar edge weight $w_e > 0$,

$$\mathcal{E}_e(y) = w_e(y_{11} + 2y_{12} - 3y_{21})^2 = y^\top \widetilde{B}_e w_e \widetilde{B}_e^\top y.$$

This illustrates that the fixed-map construction first places heterogeneous nodes in a common edge coordinate and then applies the same graph model.

Proposition 2. *Fix an edge $e = (i, j)$ and let $C_e \in \mathbb{R}^{r_e \times r_e}$ be invertible. Define*

$$A'_{e,i} = C_e A_{e,i}, \qquad A'_{e,j} = C_e A_{e,j}, \qquad W'_e = C_e^{-\top} W_e C_e^{-1}.$$

Then, for every pair of node states (y_i, y_j),

$$\big(A'_{e,i}y_i - A'_{e,j}y_j\big)^\top W'_e \big(A'_{e,i}y_i - A'_{e,j}y_j\big) = \big(A_{e,i}y_i - A_{e,j}y_j\big)^\top W_e \big(A_{e,i}y_i - A_{e,j}y_j\big).$$

Consequently, the edge contribution to the global precision matrix $Q_A(W)$ is invariant under this transformation.

Moreover, if $W_e^{(1)}$ and $W_e^{(2)}$ are two context-specific edge weights transformed by the same C_e, then

$$d_{\mathrm{AI}}\Big(C_e^{-\top} W_e^{(1)} C_e^{-1}, C_e^{-\top} W_e^{(2)} C_e^{-1}\Big) = d_{\mathrm{AI}}\Big(W_e^{(1)}, W_e^{(2)}\Big),$$

and the generalized eigenvalues of the pair $\big(W_e^{(1)}, W_e^{(2)}\big)$ are unchanged.

Proof. By (15), $\delta'_e(y) = C_e \delta_e(y)$, and hence

$$\delta'_e(y)^\top W'_e \delta'_e(y) = \delta_e(y)^\top W_e \delta_e(y).$$

Equivalently, $\widetilde{B}'_e = \widetilde{B}_e C_e^\top$, so $\widetilde{B}'_e W'_e \widetilde{B}'^\top_e = \widetilde{B}_e W_e \widetilde{B}_e^\top$. Thus the corresponding contribution to $Q_A(W)$ is unchanged. AIRM invariance follows from its congruence invariance (Bhatia, 2007; Moakher, 2005). Finally, if $W_e^{(2)} v = \lambda W_e^{(1)} v$, then with $v' = C_e v$,

$$(C_e^{-\top} W_e^{(2)} C_e^{-1}) v' = \lambda (C_e^{-\top} W_e^{(1)} C_e^{-1}) v',$$

so the generalized eigenvalues are also unchanged. □

Remark 2. The maps $A_{e,i}, A_{e,j}$ are fixed when the posterior over W is computed. If maps and weights were learned jointly, Proposition 2 would yield multiple parameterizations with the same edge energy, so additional normalization or orientation constraints would be required for identifiability.

In TCGA-BRCA, $d_i = 3$ and $r_e = 1$. The maps are leading regularized-CCA directions learned from pooled expression data without ER labels and frozen before edge-weight inference. Consequently $W_e = w_e > 0$. The general edge effect size $D_e = d_{\mathrm{AI}}(W_e^-, W_e^+)$ therefore reduces to the scalar log-ratio form in Equation (14); its signed counterpart η_e is defined there as well.

A.7 Identifiability of the edge blocks

Throughout, $\mathcal{S}^k$ denotes the space of real symmetric $k \times k$ matrices and $\mathcal{S}^k_{++}$ the open cone of symmetric positive-definite matrices, so $\dim \mathcal{S}^k = \binom{k+1}{2}$. Recall that the graph parameter is

$$W = (W_e)_{e \in E}, \qquad W_e \in \mathcal{S}^{r_e}_{++},$$

and that the corresponding graph precision is

$$Q(W) = R + \sum_{e \in E} B_e W_e B_e^\top, \qquad R \succ 0 \text{ fixed.}$$

Write N for the ambient dimension: $N = md$ in the incidence model of Section 2, and $N = p = \sum_{i \in V} d_i$ in the fixed-map model of Section A.6, where B_e is replaced by $\widetilde{B}_e = D_e^\top$.

Define the linear *edge-assembly map*

$$\mathcal{L} : \bigoplus_{e\in E} \mathcal{S}^{r_e} \longrightarrow \mathcal{S}^N, \qquad \mathcal{L}(W) = \sum_{e\in E} B_e W_e B_e^\top, \quad W = (W_e)_{e\in E}. \tag{20}$$

The admissible graph parameters form the open subset

$$\prod_{e\in E} \mathcal{S}^{r_e}_{++} \subset \bigoplus_{e\in E} \mathcal{S}^{r_e}.$$

Since R is fixed, $Q(W) = R + \mathcal{L}(W)$. Thus all statements below transfer verbatim to the fixed-map model upon replacing B_e by $\widetilde{B}_e$.

Lemma 1. *Let $R \succ 0$ be fixed and known, and let*

$$P_W = \mathcal{N}\big(0, Q(W)^{-1}\big), \qquad W = (W_e)_{e\in E} \in \prod_{e\in E} \mathcal{S}^{r_e}_{++}.$$

Then the parameterization $W \mapsto P_W$ is globally identifiable, i.e.,

$$P_W = P_{\widetilde{W}} \quad \Longrightarrow \quad W = \widetilde{W},$$

if and only if the assembly map $\mathcal{L}$ in (20) *is injective on $\bigoplus_{e\in E} \mathcal{S}^{r_e}$.*

Proof. For any $W \in \prod_{e\in E} \mathcal{S}^{r_e}_{++}$, each $B_e W_e B_e^\top$ is positive semidefinite. Since $R \succ 0$, $Q(W) = R + \mathcal{L}(W) \succ 0$, so $P_W = \mathcal{N}(0, Q(W)^{-1})$ is a nondegenerate centered Gaussian. A centered Gaussian law is uniquely determined by its covariance, and hence

$$P_W = P_{W'} \iff Q(W)^{-1} = Q(W')^{-1} \iff Q(W) = Q(W').$$

Because R is fixed and $\mathcal{L}$ is linear,

$$P_W = P_{W'} \iff \mathcal{L}(W) = \mathcal{L}(W') \iff \mathcal{L}(W - W') = 0.$$

If $\mathcal{L}$ is injective, the last equality implies $W = W'$, so the parameterization is globally identifiable.

Conversely, suppose $\mathcal{L}$ is not injective. Then there exists $H = (H_e) \neq 0$, with $H_e \in \mathcal{S}^{r_e}$, such that $\mathcal{L}(H) = 0$. Choose $\tau > \frac{1}{2} \max_{e\in E} \|H_e\|_2$ and define

$$W_e = \tau I_{r_e} + \frac{1}{2} H_e, \qquad W'_e = \tau I_{r_e} - \frac{1}{2} H_e.$$

Since every eigenvalue of $\pm H_e/2$ has magnitude at most $\|H_e\|_2/2$, this choice of τ ensures $W_e, W'_e \succ 0$ for every e. Moreover, $W \neq W'$ and

$$\mathcal{L}(W - W') = \mathcal{L}(H) = 0.$$

Thus $Q(W) = Q(W')$ and therefore $P_W = P_{W'}$, despite $W \neq W'$. Hence the parameterization is not globally identifiable. □

Theorem 3. *For each edge let $\mathcal{V}_e := \{\, B_e H B_e^\top : H \in \mathcal{S}^{r_e} \,\} \subseteq \mathcal{S}^N$. Then $\mathcal{L}$ is injective, equivalently, W is identifiable if and only if both of the following hold:*

(a) *each B_e has full column rank r_e; and*

(b) *the subspaces $\{\mathcal{V}_e\}_{e\in E}$ form a direct sum, i.e. $\mathcal{V}_e \cap \sum_{f\neq e} \mathcal{V}_f = \{0\}$ for every e; equivalently* $\dim \sum_e \mathcal{V}_e = \sum_e \binom{r_e+1}{2}$.

In particular, a necessary condition is the dimension count

$$\sum_{e\in E} \binom{r_e + 1}{2} \;\leq\; \binom{N+1}{2}. \tag{21}$$

Proof. The map $H \mapsto B_e H B_e^\top$ is injective on $\mathcal{S}^{r_e}$ if and only if B_e has full column rank. If B_e^+ is a left inverse ($B_e^+ B_e = I_{r_e}$) then $B_e H B_e^\top = 0 \Rightarrow H = B_e^+ (B_e H B_e^\top)(B_e^+)^\top = 0$. If instead $0 \neq u \in \ker B_e$, then $H = uu^\top \neq 0$ but $B_e H B_e^\top = (B_e u)(B_e u)^\top = 0$. When B_e has full column rank, $\dim \mathcal{V}_e = \dim \mathcal{S}^{r_e} = \binom{r_e+1}{2}$.

($\Leftarrow$). Assume (a) and (b) and suppose $\sum_e B_e H_e B_e^\top = 0$. Set $X_e := B_e H_e B_e^\top \in \mathcal{V}_e$; then $\sum_e X_e = 0$ with $X_e \in \mathcal{V}_e$, and the direct-sum property (b) forces $X_e = 0$ for every e. By (a) each $H_e = 0$, so $\mathcal{L}$ is injective.

($\Rightarrow$). Assume $\mathcal{L}$ is injective. If some B_e failed to have full column rank, the single-edge argument would produce $H_e \neq 0$ with $B_e H_e B_e^\top = 0$, hence $\mathcal{L}(0, \ldots, H_e, \ldots, 0) = 0$, contradicting injectivity; thus (a) holds. Given (a), suppose (b) failed, so $\sum_e X_e = 0$ with $X_e \in \mathcal{V}_e$ not all zero. Writing $X_e = B_e H_e B_e^\top$ and using (a) to recover H_e (unique, and nonzero exactly when $X_e \neq 0$) yields $H = (H_e) \neq 0$ with $\mathcal{L}(H) = 0$, again a contradiction; thus (b) holds.

Finally, (b) states $\dim \sum_e \mathcal{V}_e = \sum_e \binom{r_e+1}{2}$, and $\sum_e \mathcal{V}_e \subseteq \mathcal{S}^N$ gives (21). □

We now specialize to the two models used in the paper. The first is unconditional; the second has a genuine failure mode, certified away by a finite rank check.

Corollary 1. *Let $G = (V, E)$ be a simple graph with $m = |V|$ nodes, each carrying a d-dimensional feature vector. For each edge $e = \{i, j\}$, let*

$$b_e = \mathbf{e}_i - \mathbf{e}_j \in \mathbb{R}^m, \qquad B_e = b_e \otimes I_d,$$

where $\mathbf{e}_i$ is the ith standard basis vector of $\mathbb{R}^m$. Then, for every $d \geq 1$, the assembly map $\mathcal{L}$ is injective and the edge weights are globally identifiable.

Proof. By the Kronecker mixed-product rule (Horn & Johnson, 1991, Ch. 4),

$$B_e H_e B_e^\top = (b_e \otimes I_d) H_e (b_e \otimes I_d)^\top = (b_e b_e^\top) \otimes H_e.$$

Hence

$$\mathcal{L}(H) = \sum_{e \in E} (b_e b_e^\top) \otimes H_e,$$

which we view as an $m \times m$ array of $d \times d$ blocks.

For $e = \{i, j\}$, $b_e b_e^\top = (\mathbf{e}_i - \mathbf{e}_j)(\mathbf{e}_i - \mathbf{e}_j)^\top$, so $(b_e b_e^\top)_{ij} = -1$. For every other edge $f \neq e$, $(b_f b_f^\top)_{ij} = 0$, because a simple graph contains at most one edge with endpoints $\{i, j\}$. Therefore the (i, j) block of $\mathcal{L}(H)$ is exactly $-H_e$. If $\mathcal{L}(H) = 0$, this block must vanish, so $H_e = 0$. Repeating the argument for every edge gives $H_e = 0$ for all e, and hence $\ker \mathcal{L} = \{0\}$. Thus $\mathcal{L}$ is injective, and Lemma 1 gives global identifiability. □

This is the matrix-weighted analogue of the classical weighted-Laplacian identity: for an edge $\{i, j\}$ with scalar weight w_{ij}, the corresponding off-diagonal entry is $-w_{ij}$, so the edge weight is uniquely determined by the Laplacian (Godsil & Royle, 2001, Sec. 13.3); see also the Gaussian graphical-model parameterization in (Lauritzen, 1996). Consequently, in the incidence model, the edge-block parameterization is globally identifiable for every simple graph and any feature dimension d. This structural identifiability is a property of the parameterization itself and is distinct from the finite-sample recovery accuracy measured in the controlled simulations with known ground truth.

Corollary 2. *Consider the fixed-map model with $r_e = 1$. For each edge $e = (i, j)$, write $B_e = \widetilde{B}_e = v_e \in \mathbb{R}^p$, where v_e contains $a_{e,i}$ in the coordinates of node i, $-a_{e,j}$ in the coordinates of node j, and zeros elsewhere. Let $W_e = w_e > 0$. Then $\mathcal{L}(w) = \sum_{e \in E} w_e v_e v_e^\top$. The edge weights $w = (w_e)_{e \in E}$ are identifiable if and only if the rank-one symmetric matrices*

$$\{v_e v_e^\top : e \in E\} \subset \mathbb{S}^p$$

are linearly independent.

Equivalently, if

$$M = \begin{bmatrix} \operatorname{svec}(v_{e_1} v_{e_1}^\top) & \cdots & \operatorname{svec}(v_{e_{|E|}} v_{e_{|E|}}^\top) \end{bmatrix} \in \mathbb{R}^{\binom{p+1}{2} \times |E|},$$

then

$$w \text{ is identifiable} \quad \Longleftrightarrow \quad \operatorname{rank}(M) = |E|.$$

Proof. Apply Theorem 3 with $r_e = 1$. For each edge, $B_e = v_e$ is a $p \times 1$ matrix, so condition (a) is simply $v_e \neq 0$.

For condition (b), the contribution of edge e lies in the one-dimensional subspace $\mathcal{V}_e = \mathrm{span}\{v_e v_e^\top\} \subset \mathbb{S}^p$. The subspaces $\{\mathcal{V}_e\}_{e\in E}$ form a direct sum if and only if

$$\sum_{e\in E} \alpha_e v_e v_e^\top = 0 \quad \Longrightarrow \quad \alpha_e = 0 \text{ for every } e,$$

which is exactly linear independence of the matrices $\{v_e v_e^\top\}_{e\in E}$.

Since $\mathrm{svec} : \mathbb{S}^p \longrightarrow \mathbb{R}^{\binom{p+1}{2}}$ is a linear isomorphism, the matrices $\{v_e v_e^\top\}$ are linearly independent if and only if their vectorizations form linearly independent columns. Hence $\mathrm{rank}(M) = |E|$. □

Remark 3. In the TCGA-BRCA analysis, $p = 15$ and $|E| = 10 \leq \binom{16}{2} = 120$. For the specific frozen regularized-CCA maps used in the analysis, the matrix M in Corollary 2 has numerical rank 10. Hence the ten scalar edge weights are identifiable for these fixed maps. The smallest singular value of M is 1.335, with condition number approximately 3.10, so the full-rank conclusion is numerically well separated from rank deficiency.

Remark 4. Suppose Π_a and Π_b have densities on $\mathbb{S}^d_{++}$, and let $W_a \sim \Pi_a$ and $W_b \sim \Pi_b$ be independent. Then $\mathbb{P}(W_a = W_b) = 0$. Since AIRM is a metric,

$$d_{\mathrm{AI}}(W_a, W_b) = 0 \quad \Longleftrightarrow \quad W_a = W_b,$$

so

$$\mathbb{P}\{d_{\mathrm{AI}}(W_a, W_b) > 0\} = 1.$$

Thus posterior draws typically have positive distance even when the underlying data-generating edge matrices are identical. A positive draw-pair distance therefore reflects posterior variability as well as structural separation, motivating the same-truth reference used below.

B Posterior Inference Algorithm

ConeMALA requires the Riemannian gradient of the intrinsic posterior potential (5) to construct the Langevin drift. We first write the Euclidean gradient with respect to each edge block W_e as

$$G_e(W) := \nabla_{W_e}\Phi(W) = \frac{n}{2} B_e^\top \left(S - Q(W)^{-1}\right) B_e - \frac{\nu}{2} W_e^{-1} + \frac{1}{2}\Psi^{-1}. \tag{22}$$

Under the AIRM on $\mathbb{S}^d_{++}$, the corresponding Riemannian gradient is

$$\mathrm{grad}_e\, \Phi(W) = W_e G_e(W) W_e.$$

Thus $G_e(W)$ collects the likelihood and prior derivatives needed to form the intrinsic gradient for edge e.

The likelihood contribution requires $B_e^\top Q(W)^{-1} B_e$ for every edge. These quantities need not be computed by separately inverting $Q(W)$ for each edge. If $\mathcal{B} = [B_{e_1} \ \cdots \ B_{e_{|E|}}]$, then, after one Cholesky factorization of $Q(W)$, a single block solve

$$Q(W) Z = \mathcal{B}$$

provides all products $Q(W)^{-1} B_e$, and hence all $B_e^\top Q(W)^{-1} B_e$.

A posterior fit passes the prespecified diagnostic gate if the maximum split-$\widehat{R}$ is at most 1.05, the minimum ESS is at least 400, and no numerical failures occur. Acceptance rates are reported as sampler diagnostics but are not part of the pass/fail criterion. The affine-invariant exponential map is

$$\mathrm{Exp}_W(V) = W^{1/2} \exp(W^{-1/2} V W^{-1/2}) W^{1/2},$$

which remains SPD for every symmetric tangent vector V. Full proposal-density and reversibility derivations are given in (Dey, 2026).

Independent Wishart priors $W_e \sim \mathcal{W}_d(\nu, \Psi)$ are used throughout. For $d = 3$, $\nu = 5$ and $\Psi = I_3/5$; for Meteostat $d = 4$, $\nu = 6$ and $\Psi = I_4/6$, giving prior mean I_d.

Algorithm 1 Posterior inference for one graph context

Require: Centered observations $\{Y_r\}_{r=1}^n$; graph G; stabilizer R; prior (ν, Ψ); MCMC settings (step size, number of chains, burn-in, retained draws, and thinning).
Ensure: Posterior draws and multi-chain diagnostics.
1: Form
$$S = n^{-1} \sum_r Y_r Y_r^\top, \qquad Q(W) = R + \sum_e B_e W_e B_e^\top .$$
2: **for** each chain **do**
3: Initialize the SPD edge blocks.
4: **for** each MCMC iteration **do**
5: Evaluate the intrinsic posterior and product-AIRM gradient using the shared precision solve.
6: Generate a product-AIRM Langevin proposal and map each block with the affine-invariant exponential map.
7: Apply the exact Metropolis-Hastings correction.
8: **end for**
9: Discard burn-in and retain the prescribed production draws.
10: **end for**
11: Compute split-$\widehat{R}$, ESS, acceptance, and numerical-failure summaries.

C Additional Results for Known-Ground-Truth Structural Resolution

We use five independent generating seeds, $n \in \{200, 500, 1000\}$, and target distances $D^\star \in \{0.5, 0.75, 1.0, 1.5\}$. For each seed, a baseline truth is held fixed while matched one-direction and orientation deformations are constructed. Independent baseline, same-truth, and deformed datasets are generated, with smaller sample sizes obtained from nested prefixes. Each condition is fit independently using Algorithm 1 and must pass the diagnostic gate in Section B. Posterior draws are then paired edgewise to compute $D_{e,\mathrm{ref}}$, $D_{e,c}$, and $R_e(c)$, while generalized eigenvalues and eigenvectors assess deformation magnitude, sign, and direction recovery.

C.1 Detection boundary

We declare an edge change detectable when $R_e \geq 0.95$. For each deformation condition, one of the six graph edges is changed and the remaining five are unchanged. Across the five independent generating seeds, this gives five changed-edge and 25 unchanged-edge comparisons. Let TP and FN denote changed edges declared changed or unchanged, respectively, and let TN and FP denote unchanged edges declared unchanged or changed. We define

$$\text{Specificity} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}}.$$

No unchanged edge is declared changed in any reported one-direction or orientation condition, giving specificity 1.00 throughout.

At target deformation $D^\star = 1$, the orientation change is detected in $0/5$, $4/5$, and $5/5$ independent generating replicates at $n = 200$, 500, and 1000, respectively. The corresponding one-direction detection counts are $0/5$, $1/5$, and $3/5$. Thus deformations with the same AIRM magnitude can have different finite-sample detectability.

C.2 No-change separation and deformation recovery

Across the complete controlled experiment, all 165 posterior fits pass the prespecified diagnostic gate. These comprise five generating seeds and three sample sizes, with 11 fitted conditions for each seed–sample-size combination: one baseline condition, one independent same-truth condition, eight deformed conditions corresponding to two deformation families and four target magnitudes, and one auxiliary mixed-deformation condition. Thus, $5 \times 3 \times (1 + 1 + 8 + 1) = 165$. Table 3 reports the finite-sample separation under exactly unchanged truth. The mean draw-pair AIRM distance

n	Family	$D^\star = 0.50$	$D^\star = 0.75$	$D^\star = 1.00$	$D^\star = 1.50$
200	One-direction	$0/5$	$0/5$	$0/5$	$1/5$
200	Orientation	$0/5$	$0/5$	$0/5$	$4/5$
500	One-direction	$0/5$	$1/5$	$1/5$	$5/5$
500	Orientation	$0/5$	$2/5$	$4/5$	$5/5$
1000	One-direction	$0/5$	$1/5$	$3/5$	$5/5$
1000	Orientation	$0/5$	$4/5$	$5/5$	$5/5$

Table 2: Fraction of five independent replicates in which the changed edge satisfies $R_e \geq 0.95$.

decreases from 1.052 at $n = 200$ to 0.636 at $n = 500$ and 0.449 at $n = 1000$, closely following an empirical $n^{-1/2}$ scale. At $D^\star = 1$, recovery of the deformation magnitude approaches the target as n increases, while recovery of its sign and direction also improves with sample size.

n	Mean AIRM distance	Mean 95th percentile
200	1.052	1.726
500	0.636	0.991
1000	0.449	0.693

Table 3: Finite-sample separation between posteriors fitted to independent datasets generated from the same graph, averaged over six edges and five replicates.

n	One-direction	Orientation
200	1.561	1.446
500	1.320	1.363
1000	1.065	1.138

(a) Recovered AIRM magnitude.

Metric	$n = 200$	$n = 500$	$n = 1000$
One-direction sign	0.996	1.000	1.000
One-direction alignment	0.853	0.928	0.956
Orientation sign	0.991	0.999	1.000
Orientation alignment	0.928	0.963	0.978

(b) Directional recovery.

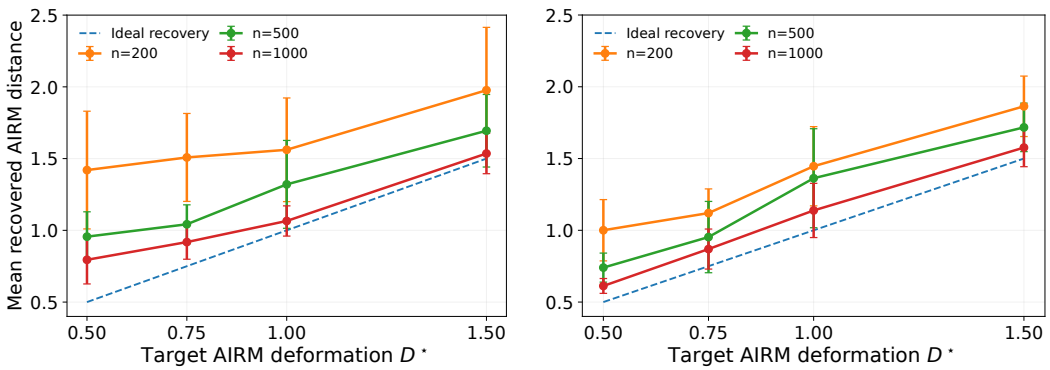
Table 4: Recovery at $D^\star = 1$. Sign entries give the probability of recovering the data-generating strengthening or weakening direction; alignment is the sign-invariant baseline-metric alignment with the data-generating generalized direction.

All 165 posterior fits pass the prespecified diagnostic gate: the largest split-$\widehat{R}$ is 1.0012, the smallest ESS is 6448, and no numerical failures occur. Production acceptance rates lie in $[0.644, 0.746]$.

C.3 Native-comparator details and localization

This appendix provides additional details for the comparison in Section 4.2. FGL is fitted separately to the two 15-dimensional contexts, with its sparsity and fusion parameters selected using validation Gaussian NLL. The Bayesian multi-GGM model is fitted using G-Wishart MCMC with four chains, 5000 burn-in iterations, and 2000 retained draws per chain. CPC is fitted to the three-dimensional contrast associated with the changed edge. All methods use only the training data for fitting; the known data-generating parameters and held-out test observations are used only for evaluation.

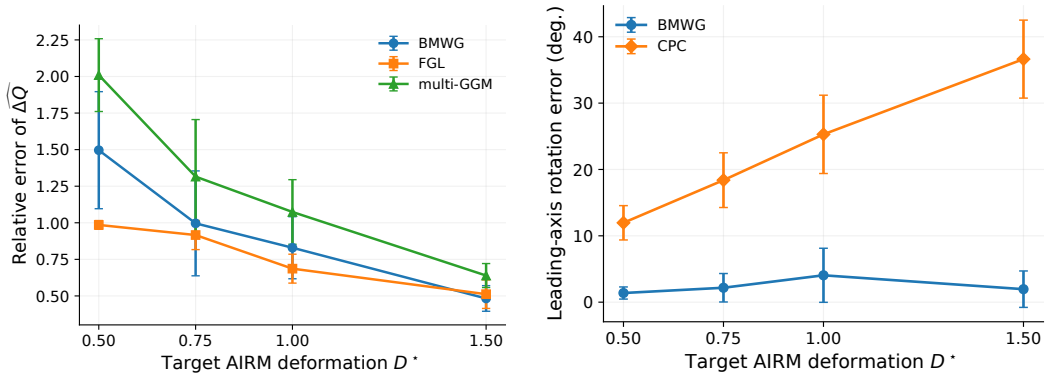
The controlled experiment in BMWG changes the strength or orientation of an existing multivariate interaction rather than adding or removing an edge. Multi-GGM's support-change probability instead measures whether individual scalar precision entries are present in one context but absent in the other. Therefore, a substantial change in a 3×3 module-pair interaction can occur even when its scalar precision entries remain nonzero in both contexts. BMWG ranks the six edges in its fixed graph, whereas FGL and multi-GGM consider all ten possible module pairs. For a more direct comparison with BMWG, we also evaluate FGL after restricting its fitted result to the same six graph edges. Table 5 reports averages over all 120 experimental conditions.



(a) One-direction deformation.

(b) Orientation deformation.

Figure 5: Recovery of the changed-edge AIRM deformation magnitude across target deformations and sample sizes. Points show means across five independent generating replicates and error bars show one sample standard deviation. The dashed diagonal denotes ideal recovery. Finite-sample posterior separation inflates the recovered magnitude at smaller n, while the estimates approach the target as the sample size increases.



(a) Global precision-change recovery.

(b) Changed-edge directional recovery.

Figure 6: Complementary recovery measures for the orientation family at $n = 1000$. (a) Relative Frobenius error $\|\widehat{\Delta Q} - \Delta Q^\star\|_F / \|\Delta Q^\star\|_F$ of the global precision change. (b) Absolute leading-axis rotation error for the changed-edge deformation. The comparison illustrates that accurate recovery of the aggregate precision change and accurate recovery of the local multivariate deformation direction are distinct objectives.

For changed-pair localization, we rank candidate module pairs using the Frobenius norm of their estimated precision change. Using Eq (12), for a candidate module pair (i, j), let

$$\widehat{\Delta Q}_{ij} := (\widehat{\Delta Q})_{ij} = \widehat{Q}_{1,ij} - \widehat{Q}_{0,ij}$$

Method	Localization measure / pairs considered	Mean rank	AUROC	AP
BMWG	6 graph edges	1.258	0.948	0.914
FGL	all 10 pairs	3.625	0.708	0.480
FGL	same 6 graph edges	2.454	0.709	0.525
multi-GGM	all 10 pairs; $\|\Delta Q_{ij}\|_F$	1.433	0.952	0.882
multi-GGM	all 10 pairs; probability of support change	7.325	0.297	0.143

Table 5: Changed-pair localization averaged over five generating seeds, three sample sizes, four target deformation magnitudes, and both deformation families. BMWG considers the six edges in its fixed graph, whereas FGL and multi-GGM consider all ten module pairs. The additional FGL row restricts its fitted result to the same six edges used by BMWG.

denote the corresponding $d \times d$ off-diagonal precision block, and define its localization score by

$$s_{ij} = \|\widehat{\Delta Q}_{ij}\|_F.$$

Therefore this quantity provides a more appropriate localization measure for this experiment (Table 5). Let $y_{ij} = 1$ for the truly changed module pair and $y_{ij} = 0$ otherwise, and let

$$N_+ = \sum_{(i,j)} \mathbf{1}(y_{ij} = 1), \qquad N_- = \sum_{(i,j)} \mathbf{1}(y_{ij} = 0)$$

denote the numbers of changed and unchanged candidate pairs, respectively. Then

$$\mathrm{AUROC} = \frac{1}{N_+ N_-} \sum_{(i,j):y_{ij}=1} \sum_{(k,\ell):y_{k\ell}=0} \left[\mathbf{1}(s_{ij} > s_{k\ell}) + \frac{1}{2}\mathbf{1}(s_{ij} = s_{k\ell}) \right].$$

After ranking the candidates by decreasing score, let TP_k and FP_k denote the numbers of positive and negative items among the first k ranks, and let N_+ be the total number of positive items. Then

$$P_k = \frac{\mathrm{TP}_k}{k}, \qquad R_k = \frac{\mathrm{TP}_k}{N_+},$$

since $\mathrm{TP}_k + \mathrm{FP}_k = k$ and $\mathrm{TP}_k + \mathrm{FN}_k = N_+$. Average precision (AP) is

$$\mathrm{AP} = \sum_k (R_k - R_{k-1}) P_k. \tag{23}$$

Because each controlled condition contains one changed pair, AP is the precision when that pair enters the ranked list. Thus larger AUROC and AP, with a maximum value of one, indicate better localization.

C.4 Isotropic BMWG representation ablation

To isolate the contribution of matrix-valued edges, we refit the controlled experiment with an *Isotropic BMWG* model in which

$$W_e = w_e I_d, \qquad w_e > 0.$$

The graph scaffold, Gaussian likelihood, training and evaluation data, and global precision construction are unchanged:

$$Q(w) = R + \sum_{e \in E} B_e (w_e I_d) B_e^\top. \tag{24}$$

Thus the ablation removes only the ability of an edge to assign different strengths to different multivariate directions or to rotate its interaction geometry. Empirically, the restricted model exhibits an error floor that does not disappear with increasing sample size, despite satisfactory MCMC diagnostics, indicating a limitation of the representation rather than of posterior sampling.

For each scalar edge parameter we use

$$w_e \sim \mathrm{Gamma}\left(\frac{\nu}{2}, \mathrm{scale} = 2\psi_{\mathrm{scale}}\right), \qquad \mathbb{E}[w_e] = \nu\psi_{\mathrm{scale}} = 1.$$

This is the one-dimensional Wishart analogue of the BMWG edge prior and matches its prior mean, so the ablation changes the edge representation without introducing a different prior location.

Sampling is performed in the log-coordinate $z_e = \log w_e$ with Metropolis-adjusted Langevin updates. The paper-profile ablation uses four chains, 2000 burn-in transitions, 4000 retained draws per chain, and thinning by five. Across all 120 isotropic fits, the largest rank-normalized split-$\widehat{R}$ is 1.0012, the smallest ESS is 4439, and production acceptance rates lie in $[0.533, 0.606]$.

To separate finite-sample estimation error from an intrinsic limitation of the isotropic model, we compute the *oracle representation gap*: the smallest relative error achievable when the true matrix-valued deformation is approximated by any isotropic deformation αI_d.

For a data-generating edge deformation $\Delta W_e^\star = W_{e,1}^\star - W_{e,0}^\star$, the best Frobenius isotropic approximation solves

$$\min_{\alpha \in \mathbb{R}} \|\Delta W_e^\star - \alpha I_d\|_F^2.$$

Its solution is

$$\alpha^\star = \frac{1}{d}\,\mathrm{tr}(\Delta W_e^\star) \tag{25}$$

giving the oracle relative residual

$$G_e^{\mathrm{iso}} = \frac{\|\Delta W_e^\star - \alpha^\star I_d\|_F}{\|\Delta W_e^\star\|_F}.$$

For the isospectral orientation perturbation, $\mathrm{tr}(\Delta W_e^\star) = 0$, so

$$\alpha^\star = 0, \qquad G_e^{\mathrm{iso}} = 1.$$

Hence no component of the additive orientation deformation lies in the isotropic subspace. For the one-direction perturbation used in known-ground-truth experiment, $\eta = (0, 0, D^\star)$, so

$$\Delta W_e^\star = (e^{D^\star} - 1)vv^\top, \qquad v = (W_{e,0}^\star)^{1/2} e_3,$$

which is rank one. Hence

$$\|\Delta W_e^\star\|_F^2 = \{\mathrm{tr}(\Delta W_e^\star)\}^2.$$

With $d = 3$ and $\alpha^\star = \mathrm{tr}(\Delta W_e^\star)/3$, orthogonal projection onto $\mathrm{span}\{I_3\}$ gives

$$\frac{\|\Delta W_e^\star - \alpha^\star I_3\|_F^2}{\|\Delta W_e^\star\|_F^2} = 1 - \frac{1}{3} = \frac{2}{3}.$$

Therefore, $G_e^{\mathrm{iso}} = \sqrt{\frac{2}{3}} \approx 0.816$. To quantify recovery of the changed-edge deformation, let

$$\Delta W_e^\star = W_{e,1}^\star - W_{e,0}^\star, \qquad \widehat{\Delta W}_e = \widehat{W}_{e,1} - \widehat{W}_{e,0},$$

where $\widehat{W}_{e,c}$ denotes the posterior-mean edge matrix in context c. We define the relative deformation error and the candidate-edge error by

$$E_{\Delta,e} = \frac{\|\widehat{\Delta W}_e - \Delta W_e^\star\|_F}{\|\Delta W_e^\star\|_F}, \qquad E_{W,e} = \frac{\|\widehat{W}_{e,1} - W_{e,1}^\star\|_F}{\|W_{e,1}^\star\|_F}$$

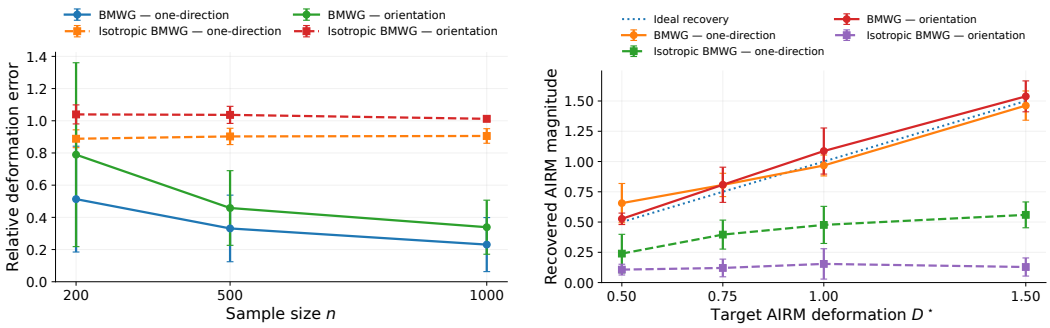
respectively. Thus $E_{\Delta,e} = 0$ corresponds to exact recovery, while larger values indicate greater discrepancy relative to the magnitude of the true deformation. The former evaluates recovery of the *change* between contexts, whereas the latter evaluates recovery of the candidate-context edge matrix itself.

Table 6: BMWG-Isotropic BMWG comparison averaged over all four target deformation magnitudes and five independent generating seeds. Entries are relative Frobenius errors on the true changed edge.

Family	n	Candidate edge $W_{e,1}$		Change ΔW_e		G_e^{iso}
		BMWG	Iso.	BMWG	Iso.	
Orientation	200	0.181	0.582	0.789	1.040	1.000
	500	0.116	0.585	0.458	1.037	1.000
	1000	0.081	0.582	0.339	1.012	1.000
One-direction	200	0.168	0.696	0.513	0.888	0.816
	500	0.115	0.699	0.331	0.903	0.816
	1000	0.076	0.698	0.231	0.906	0.816

Table 6 shows the distinction between estimation error and representation error. As n increases from 200 to 1000, BMWG deformation error falls from 0.789 to 0.339 for the orientation family and from 0.513 to 0.231 for the one-direction family. In contrast, the isotropic errors remain near 1.0 and 0.9, respectively, consistent with the oracle gaps above.

Figure 7 makes this distinction visually. Panel (a) shows that the relative deformation error of BMWG decreases with sample size for both the one-direction and orientation families, whereas the Isotropic BMWG error remains nearly flat, consistent with a persistent representation error. Panel (b) shows the same effect on the intrinsic AIRM scale at $n = 1000$: BMWG closely tracks the target deformation magnitude across both families, while Isotropic BMWG systematically attenuates the recovered change. The limitation is especially severe for orientation changes. For $D^\star \in \{0.50, 0.75, 1.00, 1.50\}$,



(a) Relative deformation error.

(b) AIRM deformation recovery at $n = 1000$.

Figure 7: BMWG versus Isotropic BMWG under the controlled one-direction and orientation deformations. (a) Relative changed-edge deformation error as a function of sample size, averaged over the four target deformation magnitudes and five independent generating replicates; error bars show one sample standard deviation. BMWG error decreases with sample size, whereas the isotropic restriction retains a substantial error floor. (b) Recovered changed-edge AIRM magnitude at $n = 1000$ as a function of the target deformation $D^\star$. Points and error bars are the mean and one sample standard deviation across five independent generating replicates. The dotted diagonal denotes ideal recovery.

the corresponding orientation-family BMWG estimates are 0.527, 0.808, 1.087, 1.538, compared with 0.107, 0.121, 0.154, 0.129 under the isotropic restriction. The advantage is also consistent across individual experimental conditions, rather than arising only after averaging. Across the complete $2 \times 4 \times 5 \times 3 = 120$ condition grid, BMWG has smaller candidate-edge error in all $120/120$ matched comparisons and smaller deformation error in $111/120$. All nine exceptions occur at $n = 200$; for $n \geq 500$, BMWG has smaller deformation error in all $80/80$ matched comparisons. Together with the oracle representation gaps and successful diagnostics above, these results indicate that the persistent isotropic error is due to the restriction $W_e = w_e I_d$, rather than failed posterior sampling.

Comparison	All conditions	$n = 200$	$n \geq 500$
Candidate-edge error	$120/120$	$40/40$	$80/80$
Deformation error	$111/120$	$31/40$	$80/80$

Table 7: Condition-by-condition BMWG wins over Isotropic BMWG. A win means strictly smaller relative Frobenius error for the same deformation family, target magnitude, generating seed, and sample size.

D Additional Meteostat Results

This appendix examines the Meteostat results in greater detail. We first compare graph geometry across weather regimes and seasons, and then identify which geographic edges and meteorological directions account for the anomaly-innovation difference. Finally, we ask whether structural energy under the training-period anomaly graph distinguishes later days whose weather features depart from the training distribution. The results show that the anomaly-innovation change varies substantially across both edges and multivariate directions, and that daily structural energy helps distinguish later out of distribution (OOD) from in-distribution (ID) atmospheric conditions.

The labels $R0, \ldots, R3$ denote four weather regimes learned from the training period only. For each 3-hour anomaly field, regime features summarize spatial means, standard deviations, ranges, 12-hour tendencies, maximum wind anomaly, minimum pressure anomaly, and the norm of the tendency vector. Features are averaged over a trailing 12-hour window using only current and past observations, standardized using training-period statistics, and clustered by k-means with $K = 4$. The labels are

ordered by increasing spatial-mean temperature anomaly at the cluster center and do not correspond to predefined meteorological categories.

D.1 Regime and seasonal separation

To place the anomaly-innovation difference on a natural atmospheric scale, we compare posterior-mean graph geometries across the four learned regimes and across seasons. For two contexts a and b, we average the AIRM distance between corresponding posterior-mean edge matrices over the 16 fixed geographic edges.

Table 8 provides a natural scale for interpreting the global anomaly-innovation separation. Across regime pairs, the mean edgewise AIRM separation ranges from 2.453 to 3.011, with mean 2.800. Across seasons it ranges from 2.347 to 4.168, with mean 3.511. The anomaly-innovation separation, $\bar{D}_{\text{anomaly,innovation}} = 3.576$, exceeds every regime-pair separation and is very close to the mean of seasonal separations. Thus replacing atmospheric state by its 12-hour evolution changes the learned spatial graph geometry on a scale comparable to changing seasons.

Context pair	Mean edge AIRM
R1–R2	3.011
R2–R3	2.917
R0–R1	2.847
R1–R3	2.795
R0–R2	2.780
R0–R3	2.453

(a) Regime pairs.

Context pair	Mean edge AIRM
Autumn–Spring	4.168
Spring–Winter	3.949
Summer–Winter	3.923
Autumn–Summer	3.609
Autumn–Winter	3.067
Spring–Summer	2.347

(b) Season pairs.

Table 8: Pairwise separation of anomaly-graph geometries across learned weather regimes and seasons. Entries are mean AIRM distances between corresponding posterior-mean edge matrices over the 16 fixed geographic edges.

D.2 Global variable and directional structure

We next ask which weather variables account for the large anomaly-innovation change. Two complementary summaries are useful. First, the edge matrices themselves describe which multivariate directions are most strongly coupled. For edge e, define $L_e = \frac{\lambda_{\max}(\bar{W}_e)}{\operatorname{tr}(\bar{W}_e)}$. A value near one means that most of the edge strength is concentrated along one multivariate direction. Averaged over the 16 edges, L_e is 0.846, 0.848, and 0.892 for raw, anomaly, and 12-hour innovation representations, respectively. The corresponding leading eigenvector is pressure-dominated on essentially every edge. Thus the strong pressure-associated component of the spatial coupling persists across all three representations.

Let $Y_i \in \mathbb{R}^4$ denote the weather vector at station i, and let Y_{-i} collect the weather vectors at all remaining stations. Writing the full weather field as $Y = \begin{pmatrix} Y_i \\ Y_{-i} \end{pmatrix}$, partition the global Gaussian precision matrix Q conformably as $Q = \begin{pmatrix} Q_{ii} & Q_{i,-i} \\ Q_{-i,i} & Q_{-i,-i} \end{pmatrix}$. Here $Q_{ii} \in \mathbb{R}^{4\times 4}$ is the precision block corresponding to station i, while $Q_{i,-i}$ contains its precision couplings with all remaining stations. The standard Gaussian conditional-distribution formula then gives

$$Y_i \mid Y_{-i} \sim \mathcal{N}\big(-Q_{ii}^{-1}Q_{i,-i}Y_{-i},\, Q_{ii}^{-1}\big)\,,$$

and hence

$$\operatorname{Cov}(Y_i \mid Y_{-i}) = Q_{ii}^{-1}$$

see (Lauritzen, 1996; Rue & Held, 2005) for details. Therefore, the diagonal entries of Q_{ii}^{-1} quantify the residual variance of each weather variable at station i after conditioning on all other stations. Larger values mean that a variable at one station is less tightly determined by the remaining stations.

We also summarize the model-implied variation of the same variable between two stations. With $\Sigma = Q^{-1}$, define

$$R_{ij,v} = \mathrm{Var}(Y_{i,v} - Y_{j,v}) = \Sigma_{iv,iv} + \Sigma_{jv,jv} - 2\Sigma_{iv,jv}.$$

This is the effective-resistance analogue for the stabilized Gaussian graph; we refer to it here simply as the *spatial contrast variance*. Larger values mean that the fitted model permits greater station-to-station differences in variable v. Table 9 reports averages over all station pairs and posterior draws.

Geometry	Lead share	Temp.	RH	Wind	Press.
Anomaly	0.848	0.066	0.262	0.380	0.007
Innovation	0.892	0.125	0.436	0.687	0.010
Raw	0.846	0.072	0.276	0.414	0.008

(a) Leading-eigenvalue share and mean conditional variance.

Variable	Anomaly	Innovation
Pressure	0.018	0.024
Temperature	0.154	0.285
Relative humidity	0.549	0.897
Wind speed	0.786	1.400

(b) Mean spatial contrast variance.

Table 9: Global variable-level summaries of the fitted Meteostat geometries. The leading share is $\lambda_{\max}(\bar{W}_e)/\operatorname{tr}(\bar{W}_e)$, averaged over the 16 graph edges. Conditional variances are averaged over stations and posterior draws; spatial contrast variances are averaged over all station pairs and posterior draws.

Table 9 shows a clear separation between the persistent and changing parts of the spatial structure. Pressure remains strongly coupled: it dominates the leading edge direction and has very small conditional and between-station variances in both representations. In contrast, from anomaly to 12-hour innovation the conditional variances of temperature, relative humidity, and wind increase from $0.066, 0.262, 0.380$ to $0.125, 0.436, 0.687$, respectively. Their spatial contrast variances likewise increase from $0.154, 0.549, 0.786$ to $0.285, 0.897, 1.400$. Thus short-horizon evolution preserves the dominant pressure-associated spatial structure but allows substantially greater station-to-station variation in temperature, humidity, and wind.

Figure 8 gives the corresponding edge-level view. Panel (a) shows the persistence of the pressure-dominated leading direction across individual edges. Panel (b) shows that the strongest sign-certain anomaly-to-innovation changes are mostly negative ($\eta = \log\lambda < 0$), indicating weaker coupling, and are dominated primarily by wind speed and relative humidity. Hence the global summaries in Table 9 and the edgewise generalized-eigenvalue analysis give the same qualitative conclusion.

D.3 Edgewise reconfiguration

The anomaly-innovation change is also uneven across geographic edges. Table 10 lists the six largest changes measured by the AIRM distance between posterior-mean edge matrices. The largest occurs on Half Moon Bay–San Carlos, while several other geographic edges also show substantial reconfiguration.

Edge	D_{mean}	Draw mean	$q_{.05}$	$q_{.95}$
Half Moon Bay–San Carlos	5.351	6.850	5.454	8.740
San Francisco–Half Moon Bay	4.609	5.819	4.461	7.497
Oakland–Novato	4.464	5.943	4.437	8.009
Palo Alto–San Carlos	4.335	5.089	3.802	6.901
Oakland–Concord	4.192	5.647	4.237	7.175
Oakland–Half Moon Bay	3.890	4.744	3.510	6.314

Table 10: Largest anomaly–innovation edge changes. D_{mean} is the AIRM distance between posterior-mean edge matrices. The remaining columns give the mean and 5th–95th percentiles of distances between independently paired posterior draws.

Together, Table 10 and Figure 8 show two complementary features of the change: its magnitude differs substantially across geographic edges, and its strongest signed directions primarily correspond to weakening of wind- and relative-humidity-associated coupling.

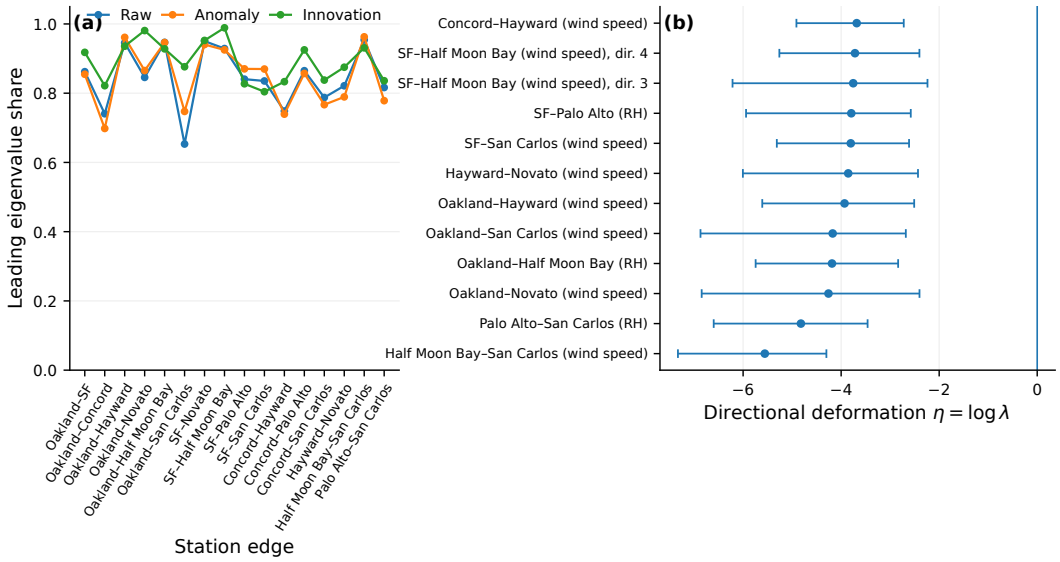


Figure 8: Edge-level structure across the fitted Meteostat geometries. (A) Fraction of each posterior-mean edge matrix accounted for by its largest ordinary eigenvalue across raw, anomaly, and 12-hour innovation geometries; the corresponding leading eigenvectors are pressure-dominated on essentially every edge. (B) Strongest sign-certain anomaly-to-innovation generalized directional changes. Horizontal intervals are posterior 5–95% quantiles for $\eta = \log \lambda$. Negative values denote weakening and positive values strengthening; labels identify the dominant meteorological variable in the corresponding generalized direction.

D.4 Atmospheric distribution-shift detection

We finally ask whether structural energy under the training-period anomaly geometry identifies later atmospheric conditions that differ from those seen during training. For each 3-hour anomaly field y_t, we compute

$$E_t = \sum_{e=(i,j)\in E} (y_{t,i} - y_{t,j})^\top \bar{W}_e (y_{t,i} - y_{t,j}),$$

where $\bar{W}_e$ is the posterior-mean anomaly edge matrix. To remove ordinary seasonal and regime-dependent variation, E_t is standardized using training-only median and median-absolute-deviation references for the corresponding season–regime group, with broader training groups used when necessary.

OOD labels are defined independently of graph energy. A 3-hour field is called novel when its standardized weather-regime feature distance exceeds the corresponding training 95th percentile, and a day is labelled OOD when at least 25% of its fields are novel. The resulting labels therefore measure feature-space distribution shift rather than being constructed from the BMWG score itself.

Each day is summarized from its eight 3-hour structural scores using one of six prespecified statistics: maximum, mean, median, mean after removing the largest score, 75th percentile, or mean of the two largest scores. Post-training days are divided chronologically into calibration, validation, and final-test periods. The validation data are used to select the daily summary and its alarm threshold; the final test period is not used for this selection.

Let S_d be a daily score, $Y_d \in \{0, 1\}$ its ID/OOD label, and $A_d(\tau) = \mathbf{1}\{S_d \geq \tau\}$ the alarm at threshold τ. We define

$$\mathrm{TPR}(\tau) = \frac{\sum_d Y_d A_d(\tau)}{\sum_d Y_d}, \qquad \mathrm{FPR}(\tau) = \frac{\sum_d (1 - Y_d) A_d(\tau)}{\sum_d (1 - Y_d)}.$$

Thus TPR is the fraction of OOD days detected and FPR is the fraction of ID days incorrectly flagged. For each candidate summary, the validation threshold is chosen subject to $\mathrm{FPR}_{\mathrm{val}} \leq 0.05$, and among the feasible summaries we select the one with largest validation TPR.

We also report two threshold-independent ranking measures. AUROC measures how often an OOD day receives a larger structural score than an ID day, with ties receiving half credit. For average precision (AP), order days by decreasing structural score. Let TP_k and FP_k denote the numbers of OOD and ID days, respectively, among the first k ranks, and let $\mathrm{FN}_k = N_+ - \mathrm{TP}_k$, where N_+ is the total number of OOD days. AUROC is

$$\mathrm{AUROC} = \Pr(S^+ > S^-) + \frac{1}{2}\Pr(S^+ = S^-),$$

where S^+ and S^- denote scores from OOD and ID days, respectively. Average precision (AP) is defined as in Eq (23). Therefore, AUROC measures OOD-versus-ID ranking over all score thresholds, whereas AP emphasizes how concentrated the OOD days are near the top of the ranking.

Statistic	Val. AUROC	Val. AP	Val. TPR	Test AUROC	Test AP
Maximum	.703	.697	.250	.751	.794
Mean	.841	.827	.438	.823	.840
Median	.891	.861	.563	.787	.810
Trimmed mean	.888	.883	.625	.835	.848
75th percentile	.800	.794	.438	.799	.834
Top-two mean	.753	.773	.438	.777	.813

Table 11: Performance of the six daily structural-score summaries. AUROC and AP measure how well the continuous score ranks OOD days above ID days. Validation TPR is evaluated at a threshold chosen to keep validation FPR at or below 0.05. The bold row is the summary selected using validation data; its test metrics are reported without further tuning.

The trimmed mean is selected because it detects the largest fraction of OOD validation days while keeping the validation FPR at or below 0.05: its validation TPR is 0.625. On the final test period, the same frozen score gives AUROC 0.835 and AP 0.848. At the corresponding threshold, test TPR is 0.36 and test FPR is 0, meaning that 36% of labelled OOD test days are detected and none of the labelled ID test days are falsely flagged.

As a sensitivity check, if the allowed validation FPR is increased from 0.05 to 0.15, the resulting frozen rule gives test TPR 0.52 and test FPR 0.033. Increasing it further to 0.20 produces only a small additional gain in detection relative to the increase in false alarms. Structural departure is also stronger in December than in November. We therefore interpret this experiment narrowly: the learned graph score detects some later atmospheric conditions whose feature distribution differs from the training period, but it is not intended as a universal detector of every unusual weather event.

E Additional TCGA-BRCA Results

This appendix provides additional validation for the TCGA-BRCA experiment in Section 4.4. We first document the label-free construction of the frozen scalar edge maps, then test whether the inferred ER-associated reconfiguration persists under scaffold changes and removal of ER-specific mean shifts.

E.1 Label-free construction of the frozen scalar edge maps

The population comparison uses 1097 unique primary TCGA-BRCA tumors and the 15 genes listed in Section 4.4. The source expression matrix uses the `PKM2` field for PKM, so the loader applies PKM←PKM2. For the population-level geometry analysis, gene-wise means and standard deviations are computed once on the primary cohort and then reused for the ER− and ER+ comparisons.

For each of the ten module pairs, we construct a one-dimensional coordinate before loading the ER labels. Specifically, the leading regularized-CCA coordinate is estimated from the pooled expression data using regularization 0.05, after which the resulting map is frozen and applied unchanged to both ER groups. Thus the coordinate construction cannot preferentially select directions because they separate ER− from ER+. The projected endpoint correlations range from 0.172 to 0.567, with mean 0.403.

Figure 9 provides an audit of these frozen coordinates. Its purpose is not to establish ER-associated reconfiguration, rather it documents the label-free representation on which the subsequent comparison is performed.

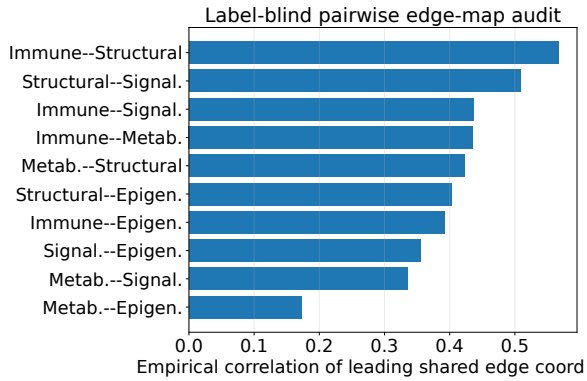


Figure 9: Empirical correlations between the projected endpoint scores for the ten module-pair coordinates. Each coordinate is learned from pooled expression without ER labels and is frozen before the ER− versus ER+ posterior comparison.

E.2 ER-ASSOCIATED EDGE RECONFIGURATION ON THE COMPLETE SCAFFOLD

We first fit the scalar BMWG model on the complete K_5 module graph, so that all ten module pairs are represented and the resulting ranking does not depend on selecting a sparse scaffold. Four chains are run separately for each ER group. The largest split-$\widehat{R}$ is 1.000255, the smallest ESS is 11042.98, and no numerical failures occur, providing satisfactory sampling diagnostics for the comparison.

Let $\bar{w}_e^-$ and $\bar{w}_e^+$ denote the posterior-mean scalar weights for edge e in the ER− and ER+ groups, respectively. Because a positive scalar is a one-dimensional SPD matrix, the AIRM distance reduces to

$$D_e = d_{\mathrm{AI}}(\bar{w}_e^-, \bar{w}_e^+) = \left| \log \frac{\bar{w}_e^+}{\bar{w}_e^-} \right|.$$

To retain the direction of change, we also report the signed log-ratio $\eta_e = \log \frac{w_e^+}{w_e^-}$

at the posterior-draw level. Positive η_e denotes a larger edge weight in ER+ and negative η_e denotes a smaller one. The intervals in Table 12 are paired-draw 5–95% posterior intervals for η_e.

Rank	Module pair	$\bar{w}_e^-$	$\bar{w}_e^+$	D_e	90% interval for η_e	Sign-certain
1	Structural/EMT–Signalling	0.182	0.924	1.625	$[1.173, 2.314]$	Yes
2	Metabolic–Epigenetic	0.036	0.113	1.151	$[0.014, 3.142]$	Yes
3	Metabolic–Signalling	0.266	0.620	0.845	$[0.541, 1.211]$	Yes
4	Immune–Structural/EMT	0.544	0.983	0.591	$[0.300, 0.928]$	Yes
5	Signalling–Epigenetic	0.351	0.523	0.398	$[0.069, 0.790]$	Yes
6	Metabolic–Structural/EMT	0.508	0.667	0.272	$[0.003, 0.572]$	Yes
7	Structural/EMT–Epigenetic	0.537	0.669	0.221	$[-0.054, 0.532]$	No
8	Immune–Epigenetic	0.325	0.392	0.189	$[-0.222, 0.715]$	No
9	Immune–Metabolic	0.580	0.618	0.063	$[-0.210, 0.364]$	No
10	Immune–Signalling	0.507	0.513	0.013	$[-0.316, 0.364]$	No

Table 12: ER− versus ER+ comparison on the complete K_5 scaffold. D_e is the AIRM distance between posterior-mean scalar edge weights. Intervals are paired-draw 5–95% posterior intervals for the signed log-ratio η_e.

The strongest reconfiguration occurs on the Structural/EMT–Signalling edge, with $D_e = 1.625$, $\eta_e \in [1.173, 2.314]$ for the central 90% posterior interval. Its posterior-mean weight changes from 0.182 in ER− to 0.924 in ER+, corresponding to a ratio of approximately $\frac{0.924}{0.182} \approx 5.08$. Six of the ten edges have sign-certain 90% intervals, whereas the four weakest changes have intervals containing zero. Thus the complete-graph analysis identifies a reconfiguration that is heterogeneous in magnitude across module pairs rather than uniform across the network.

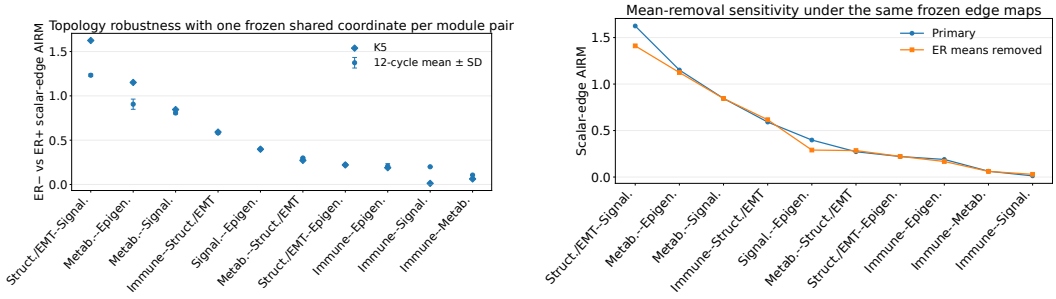
E.3 Robustness to graph scaffold and ER-specific mean shifts

The K_5 result could in principle be sensitive either to the chosen graph scaffold or to group-specific mean expression rather than dependence structure. We therefore perform two separate robustness checks.

First, we refit the analysis on all 12 labeled five-cycle scaffolds on the five modules. Each of the ten possible edges appears in six of these cycles. Structural/EMT–Signalling ranks first in all six cycles in which it appears. Moreover, the K_5 edge distances and the corresponding cycle-averaged distances agree closely, with Pearson $r = 0.990$ and Spearman $\rho = 0.988$. Thus the leading edge reconfiguration is not an artifact of fitting the complete graph.

Second, we remove ER-specific means before refitting the comparison. The mean edge AIRM distance decreases only from 0.536816 to 0.505015, retaining 94.08% of the original separation. All ten edge ranks are preserved, with Spearman $\rho = 1.000$ and Pearson $r = 0.9934$. Therefore the observed edge ordering is not explained solely by ER-associated shifts in the group means.

Figure 10 summarizes these two checks.



(a) Sensitivity to the graph scaffold.

(b) Sensitivity to ER-specific centering.

Figure 10: Robustness of the ER− versus ER+ edge reconfiguration. (A) Complete-graph edge distances are compared with the mean and standard deviation over the six five-cycle scaffolds containing each edge. (B) Edge distances before and after removal of ER-specific means. The close agreement in both panels shows that the leading module-pair reconfiguration is stable to graph topology and is not driven solely by between-group mean shifts.