COMPARING DYNAMICAL OPERATORS WITH FLAGS

Anonymous authors
Paper under double-blind review

ABSTRACT

Many scientific and engineering domains, including neuroscience, fluid dynamics, mechanics, and climate science, study ensembles of multidimensional spatiotemporal fields through spectral or modal decompositions. A recurring challenge is to compare such decompositions across many realizations while respecting both their intrinsic geometric characteristics and their hierarchy of dynamical timescales. Euclidean distances are common across these disciplines, but they are sensitive to arbitrary basis choices, whereas representing all modes by a single subspace discards their timescale organization. We represent each realization by a *flag*: a nested sequence of subspaces formed by grouping dynamical modes by spectral conjugacy and ordering them by decay time. We equip the space of flags with a chordal distance that lets us compare decompositions across realizations. The resulting distance is gauge-invariant by construction. We prove that it defines a metric, admits an isometric Euclidean embedding (thus enabling standard kernel methods), and is stable under perturbations of the fitted operator away from decay-time crossings. We illustrate the approach in climate science on 216 realizations from four CMIP6 models and a reanalysis across three variables, as well as synthetic ensembles with known ground truth. The flag geometry separates dynamical structures that the Euclidean baseline fails to distinguish and regularly outperforms a flat-subspace representation.

1 INTRODUCTION

Across the sciences, a system's dynamics are rarely captured by a single realization. Numerical simulators are run under perturbed initial conditions, competing models are developed in parallel, and observational records are assimilated into products that constrain them. Climate science is paradigmatic: a single observed trajectory is studied through ensembles of simulated fields from multiple Earth System Models (ESMs), each with various initial-condition members, alongside other fields from observation-constrained reanalyses (Maher et al., 2025). Evaluating such ensembles poses a structural question: which realizations share the same underlying modes of variability, and which differ? Answering it at the ensemble scale needs a representation of each realization compact enough to compare directly, rather than mode by mode. This is the evaluation counterpart of learning dynamics from data, where recovered equations or causal structure are useful only if they can be compared across realizations and against observations (Camps-Valls et al., 2023; Runge et al., 2023).

Most such comparisons begin by reducing each realization to an ordered *family* of modes of variability, which can be done in various ways. Principal component analysis, or empirical orthogonal function (PCA/EOF) analysis (Hotelling, 1933; North et al., 1982), orders modes by explained variance, and dynamic mode decomposition (DMD) (Schmid, 2010) is a linear example of Koopman mode decomposition (Brunton et al., 2022) where modes are often ordered by persistence. In climate science, a range of methods characterize the dynamics of the system. Methods organizing modes by temporal information include linear dynamical mode decomposition, which models the observable as a function of a latent state and orders modes by characteristic timescales (Gavrilov et al., 2020; 2024), average predictability time decomposition, which ranks modes by predictability time (DelSole & Tippett, 2009b;a), and Koopman-based approaches (Navarra et al., 2021; Lorenzo-Sánchez et al., 2025), including DMD with control (DMDc) (Mankovich et al., 2025a; 2026), which treat the observable as the state variable and, as stated above, order modes by persistence.

Comparing modal families across an ensemble is challenging, and fundamentally constrained by gauge ambiguities and ordering requirements. Pointwise Euclidean distances or correlations are

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.

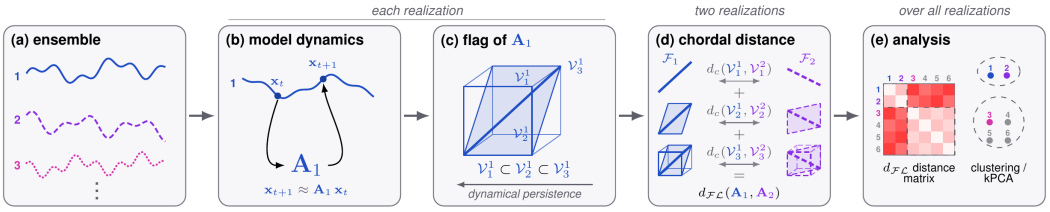


Figure 1: From ensembles of realizations to a single map of its structure. (a) Each realization is fitted with an operator $\boldsymbol{A}$, (b) whose spectrum is grouped by conjugacy and ordered by decreasing persistence. (c) A flag is generated for each realization so that each subspace k spans the dynamic modes corresponding to the top k decay times. (d) Two realizations are compared via measuring dissimilarity between their corresponding flags. Specifically, flags are compared using the sum of a chordal distance between corresponding subspace levels of the flag. (e) Then we analyze the resulting dissimilarity matrix between all realizations using clustering and kernel methods.

ill-posed because eigenvectors are defined only up to sign or complex phase. Conversely, aggregating modes into a Grassmannian subspace resolves the gauge ambiguity but discards the modal hierarchy entirely. That is, two DMD operators with identical modes but reversed timescale orderings may become indistinguishable. We therefore require a geometric representation that quotients out phase and sign symmetries without collapsing the timescale hierarchy. While spectral-Grassmann geometries combined with optimal transport compare operators through their spectra (Germain et al., 2026), our approach builds on the ordered Schur decomposition introduced by Drmač & Mezić (2026). Whereas they construct a flag of nested invariant subspaces for an individual realization, we define a distance on flag manifolds to systematically compare these structures across an ensemble.

In this work, we order the modes into a persistence hierarchy. Ordering by decay time rather than variance makes the hierarchy a statement about the operator: explained variance depends on the initial condition and the forcing realization, whereas decay times are properties of $\boldsymbol{A}$ alone. We encode it as a *flag*, a nested sequence of subspaces that folds each pair into one two-dimensional step (Ye et al., 2022), and compare two flags by summing, level by level, the chordal distance (Ye & Lim, 2016) between corresponding subspaces. Concretely, if the two most persistent modes are a conjugate pair and the next is a real mode, the flag is the plane they span sitting inside the three-dimensional subspace that adds the real mode: a line within a plane within $\mathbb{R}^d$, indexed by decay time. Each level is invariant, so the dynamics restricted to it form a closed system. Because this distance depends only on the subspace each level spans, it is by construction invariant to the gauge freedoms above. Kernel PCA (kPCA) of the ensemble distance matrix then maps the whole comparison into two dimensions. Such flag representations have proven useful well beyond dynamical systems, in multivariate analysis (Nishimori et al., 2006; Szwagier & Pennec, 2023; 2026a;b; Mankovich et al., 2024; 2025b), computer vision (Draper et al., 2014; Mankovich et al., 2022; Mankovich & Birdal, 2023), and signal processing (Santamaría et al., 2019).

Contributions. (1) We introduce the *persistence-ordered invariant flag* of a fitted operator: invariant real subspaces grouped and nested by decreasing persistence. The grouping quotients out the eigenvector phase. The ordering keeps the timescale hierarchy that a single subspace discards. (2) We compare flags with a chordal distance that admits levels of differing dimension, so realizations with different conjugacy composition remain comparable. We prove it is a metric on flags with an isometric Euclidean embedding, making the downstream kernel methods well posed, and that it is Lipschitz in the fitted operator away from decay-time crossings. (3) We validate on synthetic ensembles with known ground truth and apply the method to 216 realizations spanning four CMIP6 ESMs and the 20CRv3 reanalysis across three variables, extending single realization evaluation by Mankovich et al. (2026) to a quantitative ensemble comparison, where the flag distance separates the four models that a Euclidean baseline cannot and improves over a flat-subspace representation on every variable.

We develop the representation, distance, and its guarantees in sec. 2, then validate on synthetic ensembles and apply the method to CMIP6 climate models and reanalysis data in sec. 3 using the pipeline outlined in fig. 1. The appendices provide more theoretical details, further description of the experimental data and setups, and additional empirical evidence of performance.

2 METHODS

2.1 OPERATORS FROM DATA: A KOOPMAN VIEW

Each realization gives us a linear operator to compare. For dynamics $\boldsymbol{x}_{t+1} = F(\boldsymbol{x}_t)$ on $\mathcal{X} \subseteq \mathbb{R}^D$, the Koopman operator $\mathcal{K}$ acts on observables $g : \mathcal{X} \to \mathbb{R}$ or $g : \mathcal{X} \to \mathbb{C}$ by composition, $(\mathcal{K}g)(\boldsymbol{x}) = g\big(F(\boldsymbol{x})\big)$, and is linear even when F is not (Brunton et al., 2022).[1] Assuming the full-state observable lies in the span of the eigenfunctions, and a countable spectrum of $\mathcal{K}$, we obtain the Koopman mode decomposition of the state, $\boldsymbol{x}_t = \sum_j \lambda_j^t \, \boldsymbol{v}_j \, \varphi_j(\boldsymbol{x}_0)$, where φ_j is the jth eigenfunction, λ_j the corresponding eigenvalue, and $\boldsymbol{v}_j$ the Koopman mode. Each mode contributes a term whose temporal evolution is governed by λ_j^t. So $|\lambda_j|$ sets its rate of growth or decay over time. This notion of modal persistence is central to our construction in sec. 2.2.

Many methods produce a finite-dimensional operator $\boldsymbol{A}$ whose eigenvalues approximate Koopman eigenvalues and whose eigenvectors give the corresponding modes in whichever basis the operator was fitted on (Brunton et al., 2022; Colbrook, 2024). In our experiments, we estimate $\boldsymbol{A}$ using the DMD with control (DMDc) (Proctor et al., 2016) $\boldsymbol{x}_{t+1} \approx \boldsymbol{A}\boldsymbol{x}_t + \boldsymbol{B}\boldsymbol{y}_t$. In general, *our construction is estimator-agnostic, depending only on a particular* $\boldsymbol{A}$.

2.2 PERSISTENCE-ORDERED INVARIANT FLAG OF AN OPERATOR

In the setting we consider here, each realization supplies a *real* operator $\boldsymbol{A} \in \mathbb{R}^{d\times d}$, typically obtained by solving a regression problem, that acts on the real space of observables $\mathbb{R}^d$, as in the DMD methods above. Over $\mathbb{R}$, only the real roots of the characteristic polynomial of $\boldsymbol{A}$ are eigenvalues with eigenvectors in $\mathbb{R}^d$. The remaining roots come in conjugate pairs $(\lambda, \bar{\lambda})$ associated with oscillatory modes, and they become eigenvalues only after extending $\boldsymbol{A}$ to $\mathbb{C}^d$, which is why DMD eigendecompositions, like Koopman mode decompositions, are computed over $\mathbb{C}$. The price is that the resulting complex eigenvectors are not in the real space of observables. Each conjugate pair is nevertheless represented in $\mathbb{R}^d$, free of this ambiguity, by a two-dimensional real subspace invariant under $\boldsymbol{A}$ (see Example 2.1 below).

We collect all real and complex roots of the characteristic polynomial of $\boldsymbol{A}$ in the set of eigenvalues, a.k.a., the spectrum $\sigma(\boldsymbol{A}) \subset \mathbb{C}$. We assume throughout that $\boldsymbol{A}$ is stable, meaning $|\lambda| < 1$ for all $\lambda \in \sigma(\boldsymbol{A})$.

Example 2.1 (Complex conjugate eigenpairs)**.** *Suppose* $\boldsymbol{A}$ *is diagonalizable over* $\mathbb{C}$ *and has a simple, non-real eigenvalue* λ*. We separate the eigenpair* $(\lambda, \boldsymbol{w})$ *into*

$$a = \Re(\lambda),\ b = \Im(\lambda); \quad \boldsymbol{a} = \Re(\boldsymbol{w}),\ \boldsymbol{b} = \Im(\boldsymbol{w}).$$

Since $\boldsymbol{A}$ *is real,* $(\bar{\lambda}, \bar{\boldsymbol{w}})$ *is also an eigenpair. The complex conjugate eigenpairs can be represented over* $\mathbb{R}$ *by a real matrix for the eigenvalues and a real subspace for the eigenvectors*

$$\boldsymbol{M} = \begin{bmatrix} a & b \\ -b & a \end{bmatrix}, \qquad \mathcal{V} = \operatorname{span}_{\mathbb{R}}\{\boldsymbol{a}, \boldsymbol{b}\}. \tag{1}$$

In the coordinates given by $\{\boldsymbol{a}, \boldsymbol{b}\}$*,* $\boldsymbol{A}$ *acts on* $\mathcal{V}$ *as* $\boldsymbol{M}$*, a scaling by* $|\lambda|$ *and a rotation by* $-\arg(\lambda)$ *while fixing* $\mathcal{V}$*. Moreover, replacing* $\boldsymbol{w}$ *by* $\alpha\boldsymbol{w}$ *for any* $\alpha \in \mathbb{C}^*$ *rotates and rescales the real basis* $(\boldsymbol{a}, \boldsymbol{b})$ *while leaving* $\mathcal{V}$ *unchanged. See sec. B for more details.*

Although this subspace admits a real-valued representation, its basis is not unique. We therefore represent the learned dynamics through invariant subspaces (Takeishi et al., 2017), thus restricting to the linear case. Future work will consider invariant manifolds (Katok et al., 1995).

Definition 1 (Invariant (real) subspace of $\boldsymbol{A}$)**.** $\mathcal{V} \subset \mathbb{R}^d$ *is an invariant subspace of* $\boldsymbol{A}$ *if* $\boldsymbol{A}\boldsymbol{v} \in \mathcal{V}\ \forall\, \boldsymbol{v} \in \mathcal{V}$*. Thus, a state initially in* $\mathcal{V}$ *remains in* $\mathcal{V}$ *under repeated application of* $\boldsymbol{A}$*.*

Remark 1 (Invariant subspaces vs eigenspaces)**.** *A real eigenvector spans a one-dimensional invariant subspace of* $\boldsymbol{A}$*, while the real and imaginary parts of a complex eigenvector span a two-dimensional invariant subspace associated with its complex-conjugate eigenpair. More generally, the span of eigenvectors associated with a collection of eigenvalues is invariant when it is closed under conjugation. Invariant subspaces are more general and need not admit a basis of eigenvectors.*

[1] For stochastic dynamics $\mathcal{K}$ is the transfer operator $(\mathcal{K}g)(x) = \mathbb{E}[g(x_{t+1}) \mid x_t = x]$ (Mezić, 2005). The construction here applies to the fitted operator.

We consider *nested invariant subspaces* of $\boldsymbol{A}$ in order of decreasing persistence (decay times). To access this, we sort the eigenvalues into subsets based on magnitude, and we then construct our flag where the ℓth invariant subspace corresponds to the top ℓ distinct magnitudes.

Definition 2 (Persistence-ordered invariant flag of $\boldsymbol{A}$)**.** *Choose L to be at most the number of distinct eigenvalue magnitudes of $\boldsymbol{A}$. Let $\mathcal{V}_\ell$ be the invariant subspace of $\boldsymbol{A}$ associated with the eigenvalues with the top ℓ distinct magnitude groups. The nested subspaces $\mathcal{V}_1 \subset \cdots \subset \mathcal{V}_L$ are the* persistence-ordered invariant flag *of $\boldsymbol{A}$, with* signature $\boldsymbol{n} = (n_1, \ldots, n_L)$ [2]*. Pick an orthonormal basis for $\mathcal{V}_\ell$ and call its matrix $\boldsymbol{V}_\ell$. Writing $\boldsymbol{\Pi}_\ell = \boldsymbol{V}_\ell \boldsymbol{V}_\ell^\top$ for the orthogonal projector onto $\mathcal{V}_\ell$, we have the projection matrix representation of the flag:*

$$\boldsymbol{\Phi}(\boldsymbol{A}) = \tfrac{1}{\sqrt{2}}\big(\boldsymbol{\Pi}_1, \ldots, \boldsymbol{\Pi}_L\big). \tag{2}$$

Note: we choose to scale by $\frac{1}{\sqrt{2}}$ to directly define a distance between flags Frobenius norm in sec. 2.3.

Example 2.2. *An operator $\boldsymbol{A}$ whose most persistent modes are a conjugate pair, then a real eigenvalue, then two conjugate pairs of equal modulus, has signature $(2, 3, 7)$: a plane inside a 3-dimensional subspace inside a 7-dimensional subspace of $\mathbb{R}^d$. Furthermore, the dynamics of $\boldsymbol{A}$ are closed over each of these subspaces.*

While the invariant flag representation of linear dynamics (Drmač & Mezić, 2026) and the projection matrix embedding of flags (Ye et al., 2022) are established tools, Dfn. 2 additionally enforces that the flag is ordered by dynamical persistence. We now characterize the set of all matrices $\boldsymbol{A}$ that admit such persistence-ordered invariant flags with a prescribed signature of length L.

Definition 3 (Signature of $\boldsymbol{A}$ at level L)**.** *Let $\rho_1 \geq \cdots \geq \rho_d$ be the eigenvalue magnitudes of $\boldsymbol{A}$, listed with algebraic multiplicity. Let $n_1 < \cdots < n_m$ enumerate the dimensions $n \in \{1, \ldots, d-1\}$ at which $\rho_n > \rho_{n+1}$. For $L \leq m$, the* signature of $\boldsymbol{A}$ at level L *is* $\mathrm{sig}_L(\boldsymbol{A}) = \boldsymbol{n} = (n_1, \ldots, n_L)$.

Definition 4 (L-level signature region)**.** *The L-level signature region is the set of operators admitting a signature of length L,*

$$U_L = \bigcup_{\boldsymbol{n}:\, \mathrm{len}(\boldsymbol{n})=L} \big\{\boldsymbol{A} \in \mathbb{R}^{d\times d} : \mathrm{sig}_L(\boldsymbol{A}) = \boldsymbol{n}\big\}.$$

2.3 A FLAG DISTANCE BETWEEN DYNAMICAL OPERATORS: $d_{\mathcal{FL}}$

We compare operators by summing the *chordal distance between subspaces* (Ye & Lim, 2016),

$$d_c(\mathcal{V}_\ell, \mathcal{V}'_\ell)^2 = \tfrac{1}{2}\|\boldsymbol{\Pi}_{\mathcal{V}_\ell} - \boldsymbol{\Pi}_{\mathcal{V}'_\ell}\|_F^2, \tag{3}$$

at each level. The chordal distance has a geometric interpretation as a norm of the sines of principal angles between subspaces (Ye & Lim, 2016) and is normally used to compare subspaces of the same dimension. In our case, we allow subspaces of different dimensions.

Definition 5 (Chordal distance between flags)**.** *Suppose $\boldsymbol{A}, \boldsymbol{A}' \in U_L$, this distance amounts to*

$$d_{\mathcal{FL}}(\boldsymbol{A}, \boldsymbol{A}') = \sqrt{\sum_{\ell=1}^{L} d_c\big(\mathcal{V}_\ell, \mathcal{V}'_\ell\big)^2} = \|\boldsymbol{\Phi}(\boldsymbol{A}) - \boldsymbol{\Phi}(\boldsymbol{A}')\|_F. \tag{4}$$

Equation (4) differs from the chordal distance introduced by Pitaval & Tirkkonen (2013) because it measures the distance between each subspace in the hierarchy ($d_c\big(\mathcal{V}_\ell, \mathcal{V}'_\ell\big)$), not the gaps between them ($d_c\big(\mathcal{V}_\ell \cap \mathcal{V}_{\ell-1}^{\perp}, \mathcal{V}'_\ell \cap \mathcal{V}'^{\perp}_{\ell-1}\big)$). In our distance, leading subspaces enter at every level ℓ.

This flag distance is attractive because it embeds the flag in Euclidean space (Prop. 1), and is stable under perturbations of the operator $\boldsymbol{A}$ (Prop. 2).

Proposition 1 (Flags embed in Euclidean space)**.** *$d_{\mathcal{FL}}$ is a metric on flags and a pseudometric on U_L, and $\exp(-\gamma\, d^2_{\mathcal{FL}})$ is a positive definite kernel on both for every $\gamma > 0$.*

[2] In previous works, this has been referred to as the flag "type" and is sometimes denoted as $(n_1, \ldots, n_L)$ to include the ambient space dimension.

Proof sketch. A subspace is completely encoded by its orthogonal projector, so this mapping turns each Grassmannian into a *subset* of $(\mathrm{Sym}(d), \|\cdot\|_F)$, and stacking the L levels of a flag turns flags of L subspaces into $\mathrm{Sym}(d)^L$. Once our objects live inside a Euclidean space, we get everything we want for free. Details in sec. E. □

Proposition 2 (Stability of the flag). *Let $\boldsymbol{A} \in U_L$ be stable and diagonalizable, with $\mathrm{sig}_L(\boldsymbol{A}) = \boldsymbol{n} = (n_1, \ldots, n_L)$, and fix a diagonalization $\boldsymbol{A} = \boldsymbol{S}\boldsymbol{\Lambda}\boldsymbol{S}^{-1}$ over $\mathbb{C}$. Set*

$$\kappa = \|\boldsymbol{S}\|_2\|\boldsymbol{S}^{-1}\|_2, \quad g_\ell = \rho_{n_\ell} - \rho_{n_\ell+1}, \quad \delta = \min_{1\le\ell\le L} g_\ell > 0, \quad r_\ell = \frac{\rho_{n_\ell} + \rho_{n_\ell+1}}{2}.$$

Thus g_ℓ is the modulus gap across the ℓth cut and r_ℓ is its midpoint. If $\boldsymbol{A}' = \boldsymbol{A} + \boldsymbol{E} \in \mathbb{R}^{d\times d}$ satisfies $\mathrm{sig}_L(\boldsymbol{A}') = \boldsymbol{n}$ and $\|\boldsymbol{E}\|_F \le \delta/(4\kappa)$, then, for every ℓ, $\boldsymbol{A}'$ has exactly n_ℓ eigenvalues outside $|z| = r_\ell$ and none on that circle, and

$$d_{\mathcal{FL}}(\boldsymbol{A}, \boldsymbol{A}') \le 8\kappa^2 \left(\sum_{\ell=1}^{L} \frac{r_\ell^2}{g_\ell^4}\right)^{1/2} \|\boldsymbol{E}\|_F \le \frac{8\sqrt{L}\,\kappa^2}{\delta^2}\|\boldsymbol{E}\|_F. \tag{5}$$

Neither stability nor diagonalizability of $\boldsymbol{A}'$ is required.

Proof sketch. Analysis of the Riesz projector and linking it to orthogonal projectors yield the desired result. Details are in sec. F. □

In practice, computing equation 4 is sped up using basis representations and the identity (Jayasumana et al., 2015),

$$d_c(\mathcal{V}_\ell, \mathcal{V}'_\ell) = \sqrt{\tfrac{1}{2}(n_\ell + n'_\ell) - \left\|\boldsymbol{V}_\ell^\top \boldsymbol{V}'_\ell\right\|_F^2}. \tag{6}$$

2.4 Finding persistence-ordered invariant flags

Our tool for finding persistence-ordered invariant flags is the *real Schur decomposition* because it extracts invariant subspaces from $\boldsymbol{A}$.

Definition 6 (Real Schur Decomposition). *For a real $\boldsymbol{A} \in \mathbb{R}^{d\times d}$, there exists an orthogonal $\boldsymbol{Z} \in \mathbb{R}^{d\times d}$ so that $\boldsymbol{A} = \boldsymbol{Z}\boldsymbol{T}\boldsymbol{Z}^\top$, where $\boldsymbol{T}$ is block upper triangular with 1×1 diagonal blocks corresponding to real eigenvalues of $\boldsymbol{A}$ and 2×2 diagonal blocks corresponding to complex conjugate eigenvalue pairs (Golub & Van Loan, 2013).*

When $\boldsymbol{A}$ is diagonalizable over $\mathbb{C}$, these Schur invariant subspaces coincide with the subspaces spanned by the corresponding eigenvectors. In general, the span of the Schur vectors in $\boldsymbol{Z}$ corresponding to any leading collection of diagonal blocks of $\boldsymbol{T}$ is an invariant subspace of $\boldsymbol{A}$ (Golub & Van Loan, 2013). The Schur decomposition is often used as an alternative to the eigenvalue decomposition in DMD-related methods (Drmac et al., 2018; Drmač, 2024).

The Schur decomposition is limited to constructing invariant subspaces from its leading blocks, with no guarantee that blocks are ordered by decreasing eigenvalue magnitude. Since we aim to use the Schur decomposition as a tool to extract a persistence-ordered invariant flag for any $\boldsymbol{A} \in U_L$, we must find a way to order $\boldsymbol{V}$ by persistence. This decomposes into three problems: (i) extracting eigenvalue magnitudes from $\boldsymbol{T}$; (ii) grouping these magnitudes with numerical precision; (iii) reordering a Schur decomposition. We address these one by one.

Remark 2 ((i) Eigenvalue magnitudes from the Schur form). *The diagonal blocks of $\boldsymbol{T}$ in the real Schur form determine the eigenvalue magnitudes of $\boldsymbol{A}$ directly. A 1×1 block t corresponds to a real eigenvalue and contributes the single magnitude $\rho = |t_{jj}|$. A 2×2 block $\boldsymbol{T}_{jj}$ corresponds to a complex-conjugate pair $\lambda, \bar{\lambda}$ with $\lambda = a + bi$. $\boldsymbol{T}_{jj}$ is similar to $\boldsymbol{M}$ (defined in eq. (1)) by a rotation $\boldsymbol{Q}$, so $\det(\boldsymbol{T}_{jj}) = \det(\boldsymbol{Q}\boldsymbol{M}\boldsymbol{Q}^\top) = \det(\boldsymbol{M}) = a^2 + b^2 = |\lambda|^2$ and $\rho = \sqrt{\det(\boldsymbol{T}_{jj})}$. Since $|\lambda| = |\bar{\lambda}|$, a 2×2 block contributes this value* twice. *Reading the blocks of $\boldsymbol{T}$ in this way produces exactly d magnitudes, counted with algebraic multiplicity, which can be sorted into $\rho_1 \ge \cdots \ge \rho_d$ of Dfn. 3.*

Remark 3 ((ii) Grouping moduli in finite precision). *Dfn. 3 separates levels by the strict inequality $\rho_n > \rho_{n+1}$, which a computed spectrum realizes only up to rounding. We therefore decide equality of moduli by a relative tolerance. Discarding eigenvalues with modulus at most $\tau_0 = 10^{-6}$, we sort the remaining moduli $\rho_1 \ge \cdots \ge \rho_{d'} > \tau_0$ and cluster them greedily from the largest: a cluster opened at ρ_k absorbs each subsequent $\rho_{k'}$ with $\rho_k - \rho_{k'} \le \tau\rho_k$, where $\tau = 10^{-8}$, and $\rho_{k'}$ opens a new group otherwise. The stability of alg. 1 with respect to this tolerance is a direction for future work.*

Remark 4 ((iii) Schur reorderings)**.** *Two* `LAPACK` *routines save the day. We use* `dtrsen` *once per level ℓ to move the selected blocks to the leading positions, thus generating a basis $\boldsymbol{V}_\ell$ for each level. Another option, left for future implementations, is* `dtrexc`*, which instead moves a single block of $\boldsymbol{T}$ to any position.*

Thus, we propose alg. 1 for finding a persistence ordered flag of an operator $\boldsymbol{A} \in U_L$.

Algorithm 1 Flag of an operator $\boldsymbol{A}$

Require: operator $\boldsymbol{A} \in \mathbb{R}^{d\times d}$, levels L
1: $(\boldsymbol{Z}, \boldsymbol{T}) \leftarrow \textsc{RealSchur}(\boldsymbol{A})$
2: group diagonal blocks of $\boldsymbol{T}$ by eigenvalue modulus into L levels ▷ Remark 2 & remark 3
3: $n_\ell \leftarrow$ total dimension of the blocks in levels $1, \dots, \ell$
4: **for** $\ell = 1$ to L **do**
5: $\quad \pi_\ell \leftarrow$ selection mask for the diagonal blocks in levels $1, \dots, \ell$
6: $\quad \boldsymbol{Z}_\ell \leftarrow \textsc{ReorderSchur}(\boldsymbol{Z}, \boldsymbol{T}, \pi_\ell)$ ▷ Remark 4
7: $\quad \boldsymbol{V}_\ell \leftarrow \boldsymbol{Z}_\ell[:, 1{:}n_\ell]$
8: **end for**
9: **return** $(\boldsymbol{V}_1, \dots, \boldsymbol{V}_L)$, $\boldsymbol{n} = (n_1, \dots, n_L)$

The operations count of alg. 1 is $O(d^3)$ for the Schur decomposition and $O(Ln_L d^2)$ for the reordering at each level. So we have come full circle: we proposed a flag representation for a dynamical system and presented a method for finding it. It remains to connect the theoretical object of Dfn. 2 with the algorithm of alg. 1.

Remark 5 (The flag does not depend on the Schur form)**.** *The real Schur decomposition of $\boldsymbol{A}$ is not unique, so alg. 1 requires justification. For each level ℓ the selected eigenvalues are cut out by a condition on modulus alone, hence closed under conjugation, and are separated from the rest of $\sigma(\boldsymbol{A})$ by a strict gap since $\boldsymbol{A} \in U_L$. The span of the corresponding leading Schur vectors is therefore the kernel of the factor of the characteristic polynomial carrying the selected eigenvalues, which is determined by $\boldsymbol{A}$ and ℓ alone, and any orthonormal basis of it gives the same projector, since the orthogonal projector onto a subspace is unique. Because the selected sets are nested across levels, so are these spans, and $\boldsymbol{\Phi}(\boldsymbol{A})$ parametrizes a well-defined flag of signature $\operatorname{sig}_L(\boldsymbol{A})$. Although these are standard linear algebra results, we recapitulate them for completeness in sec. C with formal statements and proofs.*

Prop. 2 certifies how far this flag moves, measured by $d_{\mathcal{FL}}$ (Dfn. 5), when $\boldsymbol{A}$ is perturbed. With exact grouping, alg. 1 returns the spectral flag at the signature indices (Prop. 3), so equation 5 bounds how far its output moves whenever the perturbation preserves the signature. Perturbations that change the signature are not covered: for them, $d_{\mathcal{FL}}$ need not tend to zero with $\boldsymbol{E}$. Neither is the finite-precision grouping of remark 3. Studying both is a direction for future work.

A final detail must be cleared before moving to experiments. In practice, $\boldsymbol{A}$ is often fitted on a reduced space, such as the span of the leading principal components, so alg. 1 runs there. Realizations are fitted independently and carry different reduced bases, so their flags are not directly comparable.

Remark 6 (Lifting the flag)**.** *We lift each flag into the shared ambient space before comparing. For area-weighted PCA (used in our climate experiments), the transform $\boldsymbol{U} \in \mathbb{R}^{D\times d}$ is orthonormal in the area-weighted inner product, and level ℓ is represented by $\boldsymbol{\Pi}_\ell = \boldsymbol{U}\boldsymbol{V}_\ell\boldsymbol{V}_\ell^\top\boldsymbol{U}^\top$. As $\boldsymbol{U}$ is an isometry onto its range, everything in sec. 2.3 and Prop. 1 holds with ambient dimension p, and the chordal distance can be evaluated on $\boldsymbol{V}_\ell^\top(\boldsymbol{U}^\top\boldsymbol{U}')\boldsymbol{V}_\ell'$, so the two reduced bases meet only through $\boldsymbol{U}^\top\boldsymbol{U}'$. See sec. D for details.*

3 Experiments

Two things must hold for the flag distance to be useful: on synthetic data, it should recover dynamical structure that the baseline methods miss, and on real fields, it should separate distinct climate models. We use $L = 3$ in both experiments. Our experiments follow the procedure outlined in fig. 1. We begin with a large collection of realizations, each of which is a timeseries in $\mathbb{R}^d$. Our goal is to analyze the relationship between these trajectories. We fit real operator $\boldsymbol{A}$ on each realization (e.g., cluster

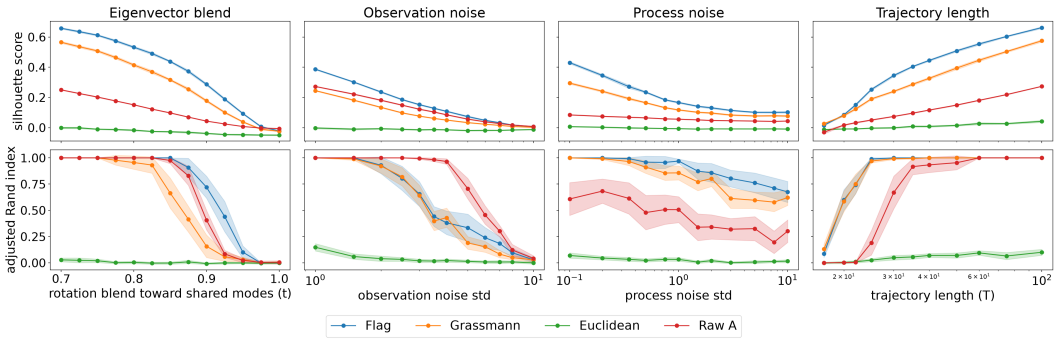


Figure 2: Synthetic ablations (3 clusters of 20 trajectories, 20 trials, shading gives 95% confidence intervals). In (top), each distance matrix is scored directly by silhouette score. In (bottom), complete-linkage agglomerative clustering, cut at three clusters, and scored against the ground-truth clustering.

member in synthetic experiment, and ensemble member or reanalysis in climate using DMDc model with least squares regression. Then we use alg. 1 to build a persistence-ordered dynamical flag of $\boldsymbol{A}$ of length L. Next we use the chordal distance (eq. (4)) generate a distance matrix between all realizations. Finally, we evaluate this distance with known structure in the data (e.g., via clustering). We call `Flag` the method that uses alg. 1 for data representation and eq. (4) for distance. Other baselines replace these steps with more standard representations of $\boldsymbol{A}$.

Baselines. We compare our method against *three baselines*. First, `Grassmannian` measures chordal distance on the flag's largest subspace before the ambient space, which drops the hierarchy but keeps the invariance. Next is `Euclidean`, where the Frobenius distance is used on raw eigenvector components. For Euclidean distance, we use 4 eigenvectors in synthetic experiments, take the top 3 eigenvalue-magnitude groups, and then pad the eigenvector matrix to 6 columns for the real-world experiment. `Raw A` is our final baseline, which compares the $\boldsymbol{A}$ matrices directly as using the Frobenius norm. This baseline is not used in the climate science experiments because the $\boldsymbol{A}$ matrices live in different EOF bases.

Whereas the previous baselines sacrifice hierarchy, gauge invariance, or basis independence, SGOT (Germain et al., 2026) preserves all three. The distinction lies in the coupling mechanism: SGOT matches modes by spatial proximity rather than rank, meaning operators with inverted persistence orderings remain close under SGOT but are separated by d_{FL}. Additionally, SGOT measures continuous eigenvalue differences, whereas the flag representation isolates their pesistence hierarchy. A careful, fair quantitative comparison is a matter of ongoing work.

Evaluation metrics. We use two metrics to score the resulting dissimilarity matrices with respect to known cluster labels. This strategy is a geometric, cluster-based comparison, allowing for a fair comparison between the geometries induced by different distance choices between dynamical operators. First, we use silhouette score (Rousseeuw, 1987) to evaluate cluster structure and compactness. Next, we use the adjusted Rand index (ARI) of an agglomerative (complete linkage) hierarchical clustering. Finally, we use kernel PCA (kPCA) (Schölkopf et al., 1997) with a Gaussian kernel for visualization.

3.1 Synthetic validation

We simulate 14-dimensional stochastic linear systems $\boldsymbol{x}_{t+1} = \boldsymbol{A}\boldsymbol{x}_t + \boldsymbol{B}\boldsymbol{y}_t + \boldsymbol{\xi}_t$, observed with noise $\hat{\boldsymbol{x}}_t = \boldsymbol{x}_t + \boldsymbol{\epsilon}_t$ and with $\boldsymbol{A}$ in a canonical form (one slow real mode, one oscillatory conjugate pair, one faster real mode). Each dataset holds 3 clusters of 20 trajectories each, whose centers share eigenvalues but differ in the eigenvectors of $\boldsymbol{A}$. Each cluster has its own persistence-ordered invariant flag encoded in $\boldsymbol{A}$. Within a dataset, the control matrix $\boldsymbol{B}$ and the control signal $\boldsymbol{y}_t$ are fixed, and the initial condition $\boldsymbol{x}_0$ varies per trajectory, so the three clusters differ only through the eigenvectors of $\boldsymbol{A}$. We run 5 ablations with 20 initial conditions per trial: perturbing the eigenvectors, the eigenvalues, the process noise ($\boldsymbol{\xi}_t$), the observation noise ($\boldsymbol{\epsilon}_t$), and the trajectory length. Further details on dataset generation are in sec. G.

Figure 2 reports the silhouette score (top row) and the ARI (bottom row) for each ablation (columns), averaged over 20 trials. These two scores answer different questions. The silhouette averages over all

trajectories, whereas complete linkage merges clusters by their largest dissimilarity, so a few badly estimated trajectories can lower the ARI while barely moving the silhouette. Conversely, clusters that are diffuse but outlier-free can reach a perfect ARI with a modest silhouette.

The flag representation achieves the highest silhouette score across all regimes where clustering structure is recoverable. The performance hierarchy—Flag, Grassmann, Raw $\mathbf{A}$, Euclidean—is preserved across every evaluation condition except observation noise. As expected, the Euclidean baseline yields low silhouette scores because the underlying clusters share no discriminative signal in the eigenvalues, while sign indeterminacy in the eigenvectors prevents meaningful alignment. The *Eigenvector blend* column tests eigenvector discrimination most directly: as we approach shared modes, the clusters converge to identical systems and every score falls to zero. As the clusters approach a common set of modes, Grassmann sees only how the signal subspace tilts, whereas Flag also sees how the modes are arranged within it.

Next, the *Trajectory length* and *Process noise* columns show how estimation error limits each method. Flag and Grassmann saturate to $\mathrm{ARI} = 1$ faster than Raw A, consistent with its poorly estimated nuisance columns. In silhouette, Flag separates from Grassmann once T suffices to resolve the inner levels of the flag.

Finally, the *Observation noise* column is the one place where the ARI ranking reverses. Observation noise corrupts the regressors, which can reorder persistence and admit nuisance modes. The resulting outliers merge clusters under complete linkage, lowering ARI while silhouette holds. This likely explains the lower ARIs for both Flag and Grassmann than Raw A, while they maintain higher silhouette scores.

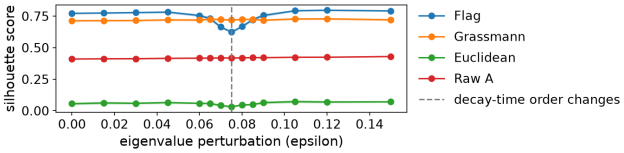


Figure 3: Eigenvalue-perturbation ablation. The flag silhouette dips sharply at the perturbation where two decay times cross and the level ordering flips (dashed line), and recovers on either side. The Grassmann and Raw A baselines do not react.

The fifth ablation perturbs the eigenvalues of $\boldsymbol{A}$, changing the ordering of subspaces in the persistence hierarchy used to build the flag. Figure 3 shows that Flag detects this reordering strongly, whereas the Grassmann and Raw A baselines are blind to it. The Euclidean baseline responds only weakly.

3.2 CMIP6 AND REANALYSIS

Configuration. We use the DMDc fits of Mankovich et al. (2026): area-weighted EOF truncation at 20 components for the temperature variables and 200 for SLP, a lag of 3 months, and the forcing variant selected there for each variable, namely three forcing predictors (greenhouse gases, volcanic, aerosols) for temperature and two (CO_2, volcanic) for SLP.

We apply the method with $L = 3$ to 216 realizations per variable: 215 members of four CMIP6 (Eyring et al., 2016) ESMs (CanESM5, CESM2, MIROC6, MPI-ESM1-2-LR; 50–65 each) (Swart et al., 2019; Tatebe et al., 2019; Mauritsen et al., 2019) from MMLEA-v2 (Maher et al., 2025), plus one 20CRv3 member (Slivinski et al., 2019). We use the DMDc fits for three monthly variables: ocean near-surface air temperature between 40°S and 65°N (OSAT), global near-surface air temperature (SAT), and sea-level pressure (SLP).

Results. In fig. 4b, the flag-based comparison separates models clearly for both temperature variables, while the Euclidean baseline collapses to near chance. The Grassmann baseline is well below the flag for all three variables, so the nested hierarchy adds signal. Separation is much weaker for SLP, whose fits use a far larger EOF truncation (200 against 20) and a lower-dimensional forcing. Unlike synthetic experiments, we extract flags of different signatures and thus leverage the full power of the chordal distance in eq. (4) for comparing them (details in fig. 8).

We proceed with our analysis on the climate science data, focusing on the results from our `Flag` pipeline (see fig. 1). In kPCA embedding of OSAT, the four models form well-separated groups with the reanalysis between them. Their distances to 20CRv3 overlap substantially (fig. 4a). In SAT and SLP, CESM2 is a clear SAT outlier and shows within-ESM distances to be much smaller than

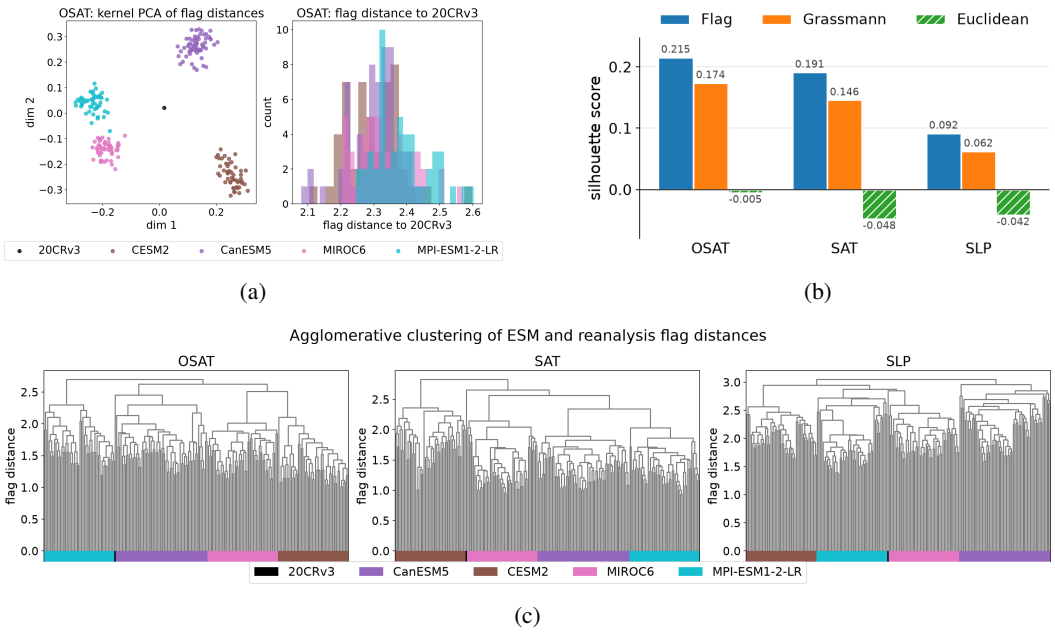


Figure 4: ESM separation on real climate data. a OSAT, 215 ESM members and one 20CRv3 member. *Left*: kPCA embedding of the flag distance matrix. *Right*: each realization's flag distance to 20CRv3. b Silhouette score of ESM separation (four CMIP6 models with 215 members total) for OSAT, SAT, and SLP; higher is better. c Hierarchical agglomerative clustering (complete linkage) of the ESM and reanalysis flag distances for OSAT, SAT, and SLP, with leaves colored by source model. Members of the same ESM largely merge at small flag distances before different models join at larger distances. The leaf colors confirm that the clustering recovers the model identities.

cross-ESM and reanalysis distances fig. 7. Figure 4c looks at these results in more detail. Members of the same ESM tend to merge at a small flag distance before distinct models join at a larger distance, a structure that the silhouette scores in fig. 4b and the median ranking in fig. 6 quantify.

Beyond separating the models, the flag distance ranks each ESM by its proximity to the reanalysis. Figure 6 orders the four ESMs by their median flag distance to 20CRv3 in each variable. The ranking is variable-dependent: CESM2 is closest to the reanalysis for OSAT yet farthest for SAT, MIROC6 is closest for SAT, and MPI-ESM1-2-LR—the most distant model for both temperature variables—is closest for SLP, consistent with the weaker and less symmetric SLP structure of fig. 4a. This weak result on SLP is expected: its atmospheric dynamics are faster and more complex than temperature, thus the DMDc model's poor fit propagates to the flag distance.

4 Conclusion

We introduced the *persistence-ordered invariant flag*: a fitted dynamical operator represented as a nested family of invariant subspaces, grouped and ordered by decay time. Comparing flags level by level with a chordal distance yields a dissimilarity invariant to the sign and phase ambiguities of an eigendecomposition, while preserving the timescale hierarchy that a single subspace discards. This distance has a Euclidean embedding, so an ensemble can be analyzed with clustering and kernel methods. Finally, it is stable under perturbations of the operator away from decay-time crossings. On synthetic data, the flag achieves the highest silhouette across all ablations and alone detects a reordering of decay times. On 216 CMIP6 and reanalysis realizations, it separates Earth System Models where a Euclidean distance is at chance, improves on a flat-subspace representation for all three variables, and ranks each model by its distance to 20CRv3.

Limitations and future work. Our experimental validation is limited to one real-world dataset, leaving further experimentation to future work. Three extensions are natural: adapting principal subspace analysis (Szwagier & Pennec, 2026b) to invariant subspaces, flag averaging (Mankovich & Birdal, 2023) of invariant flags, and generalizing our pipeline to the nonlinear case with invariant manifolds and nonlinear flags (Haller & Vizman, 2020).

AI use statement.

In this work, we used generative AI tools to generate synthetic data sets, help develop theoretical models or conceptual frameworks, formulate mathematical claims, assist in the writing of proofs, design or provide feedback on research methodology or experiments, and implement methods. Additionally, we used generative AI tools to create or modify scientific figures or images, suggest experimental parameters, create or edit software code, draft parts of a research paper, summarize or analyze existing literature, sourcing/searching for information, edit a research paper to improve readability, identify relevant literature, and propose a title or keywords for a research paper. We have not used generative AI tools to propose or refine hypotheses, assist with translation, clean and reformat datasets, support qualitative and thematic data analysis, nor interpret results.

We have reviewed all AI-assisted work. We have checked the code and verified it with test scripts. We have verified mathematical proofs by hand. We have reviewed and revised all AI-drafted text and stress-tested mathematical claims by running many examples. Furthermore, we have checked that the cited works actually exist and verified that the literature aligns with AI-generated summaries. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

Reproducibility statement

We provide our algorithm for computing a persistence-invariant flag from a dynamical operator in alg. 1. The synthetic data generation scheme is in sec. G. Code for alg. 1, synthetic experiments, and real-world experiments is available in the supplementary material and will be released in a public GitHub repository upon publication. We omit the raw CMIP6 and 20CRv3 data, along with the DMDc fits for SAT and SLP variables, due to file size; they follow the configuration of Mankovich et al. (2026) on publicly available data (MMLEA-v2, 20CRv3) and will be released upon publication.

References

Zhaojun Bai and James W Demmel. On swapping diagonal blocks in real Schur form. *Linear Algebra and its Applications*, 186:75–95, 1993.

Steven L. Brunton, Marko Budišić, Eurika Kaiser, and J. Nathan Kutz. Modern Koopman theory for dynamical systems. *SIAM Review*, 64(2):229–340, 2022.

Gustau Camps-Valls, Andreas Gerhardus, Urmi Ninad, Gherardo Varando, Georg Martius, Emili Balaguer-Ballester, Ricardo Vinuesa, Emiliano Diaz, Laure Zanna, and Jakob Runge. Discovering causal relations and equations from data. *Physics Reports*, 1044:1–68, 2023.

Matthew J Colbrook. The multiverse of dynamic mode decomposition algorithms. In *Handbook of numerical analysis*, volume 25, pp. 127–230. Elsevier, 2024.

Timothy DelSole and Michael K. Tippett. Average predictability time. part ii: Seamless diagnoses of predictability on multiple time scales. *Journal of the Atmospheric Sciences*, 66:1188–1204, 5 2009a. ISSN 0022-4928. doi: 10.1175/2008JAS2869.1.

Timothy DelSole and Michael K. Tippett. Average predictability time. part i: Theory. *Journal of the Atmospheric Sciences*, 66:1172–1187, 5 2009b. ISSN 0022-4928. doi: 10.1175/2008JAS2868.1.

Bruce Draper, Michael Kirby, Justin Marks, Tim Marrinan, and Chris Peterson. A flag representation for finite collections of subspaces of mixed dimensions. *Linear Algebra and its Applications*, 451: 15–32, 2014.

Zlatko Drmač. A LAPACK implementation of the dynamic mode decomposition. *ACM Transactions on Mathematical Software*, 50(1):1–32, 2024.

Zlatko Drmač and Igor Mezić. A data driven Koopman–Schur decomposition for computational analysis of nonlinear dynamics. *SIAM Journal on Scientific Computing*, 48(3):A1430–A1461, 2026.

Zlatko Drmac, Igor Mezic, and Ryan Mohr. Data driven modal decompositions: Analysis and enhancements. *SIAM Journal on Scientific Computing*, 40(4):A2253–A2285, 2018.

Alan Edelman, Tomás A Arias, and Steven T Smith. The geometry of algorithms with orthogonality constraints. *SIAM Journal on Matrix Analysis and Applications*, 20(2):303–353, 1998.

Veronika Eyring, Sandrine Bony, Gerald A. Meehl, Catherine A. Senior, Bjorn Stevens, et al. Overview of the coupled model intercomparison project phase 6 (cmip6) experimental design and organization. *Geoscientific Model Development*, 9(5):1937–1958, 2016.

A. Gavrilov, S. Kravtsov, and D. Mukhin. Analysis of 20th century surface air temperature using linear dynamical modes. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 30:123110, 12 2020. ISSN 10897682. doi: 10.1063/5.0028246.

Andrey Gavrilov, Sergey Kravtsov, Maria Buyanova, Dmitry Mukhin, Evgeny Loskutov, and Alexander Feigin. Forced response and internal variability in ensembles of climate simulations: identification and analysis using linear dynamical mode decomposition. *Climate Dynamics*, 62:1783–1810, 2024.

Thibaut Germain, Rémi Flamary, Vladimir Kostic, and Karim Lounici. A spectral-Grassmann Wasserstein metric for operator representations of dynamical systems. In *International Conference on Learning Representations*, volume 2026, pp. 119652–119688, 2026.

Gene H. Golub and Charles F. Van Loan. *Matrix Computations*. Johns Hopkins University Press, 4th edition, 2013.

Stefan Haller and Cornelia Vizman. Nonlinear flag manifolds as coadjoint orbits. *Annals of global analysis and geometry*, 58(4):385–413, 2020.

Harold Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of educational psychology*, 24(6):417, 1933.

Sadeep Jayasumana, Richard Hartley, Mathieu Salzmann, Hongdong Li, and Mehrtash Harandi. Kernel methods on Riemannian manifolds with gaussian RBF kernels. *IEEE transactions on pattern analysis and machine intelligence*, 37(12):2464–2477, 2015.

Tosio Kato. *Perturbation Theory for Linear Operators*. Classics in Mathematics. Springer-Verlag, 1995. Reprint of the 1980 edition.

Anatole Katok, AB Katok, and Boris Hasselblatt. *Introduction to the modern theory of dynamical systems*. Number 54. Cambridge university press, 1995.

Paula Lorenzo-Sánchez, Matthew Newman, John Albers, Aneesh C Subramanian, and Antonio Navarra. Koopman theory for enhanced tropical Pacific SST forecasting. *Artificial Intelligence for the Earth Systems*, 4(4):e240088, 2025.

Nicola Maher, Adam S Phillips, Clara Deser, Robert C Jnglin Wills, Flavio Lehner, John Fasullo, Julie M Caron, Lukas Brunner, Urs Beyerle, and Jemma Jeffree. The updated Multi-Model Large Ensemble Archive and the Climate Variability Diagnostics Package: new tools for the study of climate variability and change. *Geoscientific Model Development*, 18(18):6341–6365, 2025.

Nathan Mankovich and Tolga Birdal. Chordal averaging on flag manifolds and its applications. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023.

Nathan Mankovich, Emily J. King, Chris Peterson, and Michael Kirby. The flag median and flagirls. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.

Nathan Mankovich, Gustau Camps-Valls, and Tolga Birdal. Fun with flags: Robust principal directions via flag manifolds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 330–340, 2024.

Nathan Mankovich, Shahine Bouabid, Peer Nowack, Deborah Bassotto, and Gustau Camps-Valls. Analyzing climate scenarios using dynamic mode decomposition with control. *Environmental Data Science*, 2025a.

Nathan Mankovich, Ignacio Santamaría, Gustau Camps-Valls, and Tolga Birdal. A flag decomposition for hierarchical datasets. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 18738–18748, 2025b.

Nathan Mankovich, Andrei Gavrilov, and Gustau Camps-Valls. Disentangling forced and internal climate variability in single realizations using dynamic mode decomposition with control. *arXiv preprint arXiv:2607.18298*, 2026.

Thorsten Mauritsen, Jürgen Bader, Tobias Becker, Jörg Behrens, Matthias Bittner, Renate Brokopf, Victor Brovkin, Martin Claussen, Traute Crueger, Monika Esch, et al. Developments in the MPI-M Earth System Model version 1.2 (MPI-ESM1.2) and its response to increasing CO_2. *Journal of Advances in Modeling Earth Systems*, 11(4):998–1038, 2019.

Igor Mezić. Spectral properties of dynamical systems, model reduction and decompositions. *Nonlinear Dynamics*, 41(1):309–325, 2005.

Antonio Navarra, Joe Tribbia, and Stefan Klus. Estimation of Koopman transfer operators for the equatorial Pacific SST. *Journal of the Atmospheric Sciences*, 78(4):1227–1244, 2021.

Yasunori Nishimori, Shotaro Akaho, and Mark D. Plumbley. Riemannian optimization method on the flag manifold for independent subspace analysis. In *Independent Component Analysis and Blind Signal Separation (ICA 2006)*, volume 3889 of *Lecture Notes in Computer Science*, pp. 295–302, 2006.

Gerald R North, Thomas L Bell, Robert F Cahalan, and Fanthune J Moeng. Sampling errors in the estimation of empirical orthogonal functions. *Monthly weather review*, 110(7):699–706, 1982.

Renaud-Alexandre Pitaval and Olav Tirkkonen. Flag orbit codes and their expansion to Stiefel codes. In *2013 IEEE Information Theory Workshop (ITW)*, pp. 1–5. IEEE, 2013.

Joshua L. Proctor, Steven L. Brunton, and J. Nathan Kutz. Dynamic mode decomposition with control. *SIAM Journal on Applied Dynamical Systems*, 15(1):142–161, 2016.

Peter J Rousseeuw. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20:53–65, 1987.

Jakob Runge, Andreas Gerhardus, Gherardo Varando, Veronika Eyring, and Gustau Camps-Valls. Causal inference for time series. *Nature Reviews Earth & Environment*, 4(7):487–505, 2023.

Ignacio Santamaría, David Ramirez, and Louis L. Scharf. Subspace averaging and order determination for source enumeration. In *IEEE Transactions on Signal Processing*, 2019.

Peter J. Schmid. Dynamic mode decomposition of numerical and experimental data. *Journal of Fluid Mechanics*, 656:5–28, 2010.

Bernhard Schölkopf, Alexander Smola, and Klaus-Robert Müller. Kernel principal component analysis. In *International conference on artificial neural networks*, pp. 583–588. Springer, 1997.

Laura C. Slivinski et al. Towards a more reliable historical reanalysis: Improvements for version 3 of the twentieth century reanalysis system. *Quarterly Journal of the Royal Meteorological Society*, 145(724):2876–2908, 2019.

Neil C Swart, Jason NS Cole, Viatcheslav V Kharin, Mike Lazare, John F Scinocca, Nathan P Gillett, James Anstey, Vivek Arora, James R Christian, Sarah Hanna, et al. The Canadian Earth System Model version 5 (Canesm5.0.3). *Geoscientific Model Development*, 12(11):4823–4873, 2019.

Tom Szwagier and Xavier Pennec. Rethinking the riemannian logarithm on flag manifolds as an orthogonal alignment problem. In *Geometric Science of Information (GSI 2023)*, 2023.

Tom Szwagier and Xavier Pennec. Nested subspace learning with flags. *Journal of Machine Learning Research*, 27(106):1–48, 2026a. URL http://jmlr.org/papers/v27/25-0807.html.

Tom Szwagier and Xavier Pennec. The curse of isotropy: from principal components to principal subspaces. *Statistical Science*, 2026b. arXiv:2307.15348.

Naoya Takeishi, Yoshinobu Kawahara, and Takehisa Yairi. Learning koopman invariant subspaces for dynamic mode decomposition. *Advances in neural information processing systems*, 30, 2017.

Hiroaki Tatebe, Tomoo Ogura, Tomoko Nitta, Yoshiki Komuro, Koji Ogochi, Toshihiko Takemura, Kengo Sudo, Miho Sekiguchi, Manabu Abe, Fuyuki Saito, et al. Description and basic evaluation of simulated mean state, internal variability, and climate sensitivity in MIROC6. *Geoscientific Model Development*, 12(7):2727–2765, 2019.

Lloyd N Trefethen. Spectra and pseudospectra: The behaviour of non-normal matrices and operators. In *The graduate student's guide to numerical analysis' 98: Lecture notes from the VIII EPSRC Summer School in Numerical Analysis*, pp. 217–250. Springer, 2005.

Ke Ye and Lek-Heng Lim. Schubert varieties and distances between subspaces of different dimensions. *SIAM Journal on Matrix Analysis and Applications*, 37(3):1176–1197, 2016.

Ke Ye, Ken Sze-Wai Wong, and Lek-Heng Lim. Optimization on flag manifolds. *Mathematical Programming*, 194(1):621–660, 2022.

A NOTATION

Table 1: Notation used throughout the paper.

Symbol	Meaning
Dynamics and operators	
d	Ambient dimension; state dimension and dictionary size, $\boldsymbol{A} \in \mathbb{R}^{d\times d}$
$\mathcal{X}$	State space, $\mathcal{X} \subseteq \mathbb{R}^d$
F	Nonlinear dynamics, $\boldsymbol{x}_{t+1} = F(\boldsymbol{x}_t)$
$\boldsymbol{x}_t,\ \boldsymbol{y}_t$	State and exogenous control signal at time t
$\mathcal{K}$	Koopman (transfer) operator, $(\mathcal{K}g)(\boldsymbol{x}) = g\big(F(\boldsymbol{x})\big)$
$g,\ \psi_i$	Observable; dictionary function
$\mathcal{D}$	Finite dictionary, $\operatorname{span}\{\psi_1, \dots, \psi_d\}$
$\varphi_j,\ \lambda_j,\ \boldsymbol{v}_j$	jth Koopman eigenfunction, eigenvalue, and mode
$\boldsymbol{A},\ \boldsymbol{B}$	Fitted operator and control matrix, $\boldsymbol{x}_{t+1} \approx \boldsymbol{A}\boldsymbol{x}_t + \boldsymbol{B}\boldsymbol{y}_t$; $\boldsymbol{A} \in \mathbb{R}^{d\times d}$
$\boldsymbol{A}'$	Second operator (realization) compared to $\boldsymbol{A}$; primes mark all quantities derived from it ($\mathcal{V}'_\ell$, $\boldsymbol{V}'_\ell$, n'_ℓ, $\boldsymbol{\Pi}'_\ell$)
$\sigma(\boldsymbol{A})$	Spectrum of $\boldsymbol{A}$
$\boldsymbol{I}$	Identity matrix
N	Number of operators (realizations) in an ensemble
$\boldsymbol{\xi}_t,\ \boldsymbol{\epsilon}_t$	Process and observation noise (synthetic experiments)
Spectrum and grouping	
$\lambda,\ \mu$	Eigenvalues of $\boldsymbol{A}$; $\lambda = a + bi$ with $a = \Re\lambda$, $b = \Im\lambda$
$\bar{\lambda}$	Complex conjugate of λ
$\lvert\lambda\rvert$	Modulus of an eigenvalue (persistence); $\lambda = \lvert\lambda\rvert e^{i\omega}$, $\omega = \arg\lambda$
$\boldsymbol{w},\ \bar{\boldsymbol{w}}$	Eigenvector of a conjugate pair and its conjugate, $\boldsymbol{A}\boldsymbol{w} = \lambda\boldsymbol{w}$
$\rho_1 \geq \cdots \geq \rho_d$	Eigenvalue magnitudes of $\boldsymbol{A}$, with algebraic multiplicity, sorted decreasing (remark 2)
$\mathcal{E}_\lambda$	Generalized eigenspace $\ker(\boldsymbol{A} - \lambda\boldsymbol{I})^{d_\lambda} \subset \mathbb{C}^d$, $d_\lambda = \text{alg. mult}(\lambda)$
Λ_ℓ	Eigenvalues of the top ℓ magnitude groups; written $\Lambda_1(n)$ in sec. C for the set selected at cut n
L	Number of flag levels (hyperparameter); $L = 3$ in the real-data experiments
ℓ	Level index, $\ell \in \{1, \dots, L\}$
Subspaces, projectors, flags	
$\mathcal{V}_\ell$	ℓth flag subspace, the invariant subspace of the top ℓ magnitude groups
$\mathcal{V}_1 \subset \cdots \subset \mathcal{V}_L$	The invariant flag of $\boldsymbol{A}$ (Dfn. 2)
n_ℓ	$\dim \mathcal{V}_\ell$: cumulative algebraic multiplicity of Λ_ℓ
$\boldsymbol{n}$	Flag signature, $\boldsymbol{n} = (n_1, \dots, n_L)$
$\operatorname{sig}_L(\boldsymbol{A})$	Signature of $\boldsymbol{A}$ at level L (Dfn. 3)
m	Number of strict spectral gaps of $\boldsymbol{A}$; sig_L requires $L \leq m$
$U_{\boldsymbol{n}},\ U_L$	Signature region $\{\boldsymbol{A} : \operatorname{sig}_L(\boldsymbol{A}) = \boldsymbol{n}\}$; its union over signatures of length L (Dfn. 4)
$\mathcal{F}_{L,d}$	Set of all L-level flags in $\mathbb{R}^d$, of any signature
fl	Operator-to-flag map, $\text{fl} : U_L \to \mathcal{F}_{L,d}$
$\boldsymbol{\Pi}_\ell,\ \boldsymbol{\Pi}_\mathcal{V}$	Orthogonal projector onto $\mathcal{V}_\ell$ (resp. onto $\mathcal{V}$)
$\boldsymbol{V}_\ell$	Real orthonormal Schur basis of $\mathcal{V}_\ell$; $\boldsymbol{V}_\ell \in \mathbb{R}^{d\times n_\ell}$, $\boldsymbol{\Pi}_\ell = \boldsymbol{V}_\ell\boldsymbol{V}_\ell^\top$
$\boldsymbol{Z}, \boldsymbol{T}$	Real Schur factors, $\boldsymbol{A} = \boldsymbol{Z}\boldsymbol{T}\boldsymbol{Z}^\top$; $\boldsymbol{T}$ block upper triangular, $\boldsymbol{T}_{11}$ its leading block
$\boldsymbol{Z}_\ell$	Orthogonal factor of the Schur form reordered at level ℓ; $\boldsymbol{V}_\ell = \boldsymbol{Z}_\ell[:, 1{:}n_\ell]$
π_ℓ	Selection mask for the diagonal blocks of levels $1, \dots, \ell$ (alg. 1)
$\boldsymbol{M}$	Real form $\left[\begin{smallmatrix} a & b \\ -b & a \end{smallmatrix}\right]$ of a conjugate pair
$\boldsymbol{C}$	Unitary $\frac{1}{\sqrt{2}}\left[\begin{smallmatrix} 1 & -i \\ 1 & i \end{smallmatrix}\right]$ realifying a conjugate pair (sec. B)
$\boldsymbol{Q}$	Orthonormal basis rotation, $\boldsymbol{Q} \in O(n_\ell)$
θ_i	Principal angles between two subspaces

continued on next page

Table 1 (continued)

Symbol	Meaning
Distances, embeddings, kernels	
$\mathrm{Gr}(n,d)$	Grassmannian of n-planes in $\mathbb{R}^d$
$\mathrm{Gr}(d)$	All subspaces of $\mathbb{R}^d$, $\bigcup_{k=0}^{d}\mathrm{Gr}(k,d)$
$\mathrm{Sym}(d)$	Real symmetric $d\times d$ matrices
ι	Subspace embedding, $\iota(\mathcal{V})=\frac{1}{\sqrt{2}}\boldsymbol{\Pi}_{\mathcal{V}}\in\mathrm{Sym}(d)$
$d_c(\mathcal{V},\mathcal{V}')$	Chordal (projection) distance, equation 3
$\boldsymbol{\Psi}$	Levelwise embedding of a flag, $\boldsymbol{\Psi}(\mathcal{F})=\big(\iota(\mathcal{V}_1),\dots,\iota(\mathcal{V}_L)\big)$
$\boldsymbol{\Phi}$	Flag embedding of an operator, $\boldsymbol{\Phi}=\boldsymbol{\Psi}\circ\mathrm{fl}$, $\boldsymbol{\Phi}(\boldsymbol{A})=\frac{1}{\sqrt{2}}(\boldsymbol{\Pi}_1,\dots,\boldsymbol{\Pi}_L)\in\mathrm{Sym}(d)^L$, equation 2
$d_{\mathcal{FL}}$	Chordal flag distance, $\|\boldsymbol{\Phi}(\boldsymbol{A})-\boldsymbol{\Phi}(\boldsymbol{A}')\|_F$, equation 4
γ	Kernel bandwidth; kPCA kernel $\exp(-\gamma\, d_{\mathcal{FL}}^2)$, $\gamma>0$
$\|\cdot\|_F$	Frobenius norm
Gauge freedom	
α	Eigenvector gauge, $\alpha=e^{i\phi}$ (or general $\alpha\in\mathbb{C}^*$); $\boldsymbol{w}\mapsto\alpha\boldsymbol{w}$
$\boldsymbol{R}(\phi)$	Planar rotation in $SO(2)$; gauge action $[\,\Re\boldsymbol{w}\;\;\Im\boldsymbol{w}\,]\mapsto[\,\Re\boldsymbol{w}\;\;\Im\boldsymbol{w}\,]\boldsymbol{R}(\phi)$

B Real Form of a Complex-Conjugate Pair

We record the real representation of a complex-conjugate eigenpair used in Example 2.1, and the gauge freedom it carries. Throughout, $\boldsymbol{A}\in\mathbb{R}^{d\times d}$ and $\lambda=a+bi\in\sigma(\boldsymbol{A})$ with $b\neq 0$.

Real form.. Since $\boldsymbol{A}$ is real, $\bar\lambda\in\sigma(\boldsymbol{A})$, and if $\boldsymbol{A}\boldsymbol{w}=\lambda\boldsymbol{w}$ then $\boldsymbol{A}\bar{\boldsymbol{w}}=\bar\lambda\bar{\boldsymbol{w}}$. Conjugating the diagonal pair by the unitary

$$\boldsymbol{C}=\tfrac{1}{\sqrt{2}}\begin{bmatrix}1 & -i\\ 1 & i\end{bmatrix}\qquad\text{gives}\qquad\boldsymbol{M}:=\boldsymbol{C}^*\,\mathrm{diag}(\lambda,\bar\lambda)\,\boldsymbol{C}=\begin{bmatrix}a & b\\ -b & a\end{bmatrix},$$

a real matrix with $\det(\boldsymbol{M})=a^2+b^2=|\lambda|^2$, so $\sqrt{\det(\boldsymbol{M})}=|\lambda|$. This is the identity used in Remark 2 to read eigenvalue magnitudes off the 2×2 diagonal blocks of $\boldsymbol{T}$, which are similar to $\boldsymbol{M}$ by a rotation (Golub & Van Loan, 2013).

The same conjugation maps the eigenvector pair to a real basis,

$$[\,\boldsymbol{w}\;\;\bar{\boldsymbol{w}}\,]\,\boldsymbol{C}=\tfrac{1}{\sqrt{2}}\big[\,\boldsymbol{w}+\bar{\boldsymbol{w}}\quad -i(\boldsymbol{w}-\bar{\boldsymbol{w}})\,\big]=\sqrt{2}\,[\,\Re(\boldsymbol{w})\;\;\Im(\boldsymbol{w})\,],$$

so the pair spans the real two-dimensional subspace $\mathcal{V}=\mathrm{span}_{\mathbb{R}}\{\Re(\boldsymbol{w}),\Im(\boldsymbol{w})\}$. Writing $\lambda=\rho e^{i\omega}$, the matrix $\boldsymbol{M}$ scales by $\rho=|\lambda|$ and rotates by ω. Since $\{\Re(\boldsymbol{w}),\Im(\boldsymbol{w})\}$ need not be orthonormal, the action of $\boldsymbol{A}$ on $\mathcal{V}$ is only *similar* to this, and physical amplitudes can grow transiently before decaying at the asymptotic rate ρ.

Gauge freedom.. An eigenvector is determined only up to a nonzero complex scalar. Taking $\alpha=e^{i\phi}$ of unit modulus and replacing $\boldsymbol{w}\mapsto\alpha\boldsymbol{w}$ (hence $\bar{\boldsymbol{w}}\mapsto\bar\alpha\bar{\boldsymbol{w}}$) gives

$$\Re(\alpha\boldsymbol{w})=\cos\phi\,\Re(\boldsymbol{w})-\sin\phi\,\Im(\boldsymbol{w}),\qquad\Im(\alpha\boldsymbol{w})=\sin\phi\,\Re(\boldsymbol{w})+\cos\phi\,\Im(\boldsymbol{w}),$$

that is, $[\,\Re(\alpha\boldsymbol{w})\;\;\Im(\alpha\boldsymbol{w})\,]=[\,\Re(\boldsymbol{w})\;\;\Im(\boldsymbol{w})\,]\,\boldsymbol{R}(\phi)$ for the planar rotation $\boldsymbol{R}(\phi)\in SO(2)$; a general $\alpha\in\mathbb{C}^*$ additionally scales both columns by $|\alpha|$. In either case the column span is unchanged:

$$\mathrm{span}_{\mathbb{R}}\{\Re(\alpha\boldsymbol{w}),\Im(\alpha\boldsymbol{w})\}=\mathrm{span}_{\mathbb{R}}\{\Re(\boldsymbol{w}),\Im(\boldsymbol{w})\}=\mathcal{V}.$$

The individual real vectors $\Re(\boldsymbol{w}),\Im(\boldsymbol{w})$ are therefore not intrinsic to $\boldsymbol{A}$, while the plane $\mathcal{V}$ is, and $\mathcal{V}$ is $\boldsymbol{A}$-invariant since it is the real part of the $\boldsymbol{A}$-invariant complex plane $\mathrm{span}_{\mathbb{C}}\{\boldsymbol{w},\bar{\boldsymbol{w}}\}$. This is the gauge freedom that Equation (6) quotients out: the projector $\Pi_{\mathcal{V}}=\boldsymbol{V}\boldsymbol{V}^\top$ is unchanged under $\boldsymbol{V}\mapsto\boldsymbol{V}\boldsymbol{Q}$ for $\boldsymbol{Q}\in O(2)$.

Relation to the Schur form.. The real Schur decomposition $\boldsymbol{A}=\boldsymbol{Z}\boldsymbol{T}\boldsymbol{Z}^\top$ of Dfn. 6 exists for every real $\boldsymbol{A}$ and represents each conjugate pair by a 2×2 diagonal block (Golub & Van Loan, 2013). It is

not unique. When $\boldsymbol{A}$ is diagonalizable over $\mathbb{C}$, the span of a leading collection of Schur vectors that does not split a diagonal block coincides with the span of the corresponding eigenvectors, with each conjugate pair replaced by $\Re(\boldsymbol{w}), \Im(\boldsymbol{w})$ as above. That this span depends only on $\boldsymbol{A}$, and not on the Schur form computed, is proved in Sec. C.

C A Flag of Invariant Subspaces is Well Defined

We now formalize remark 5: for every choice it makes, alg. 1 returns the same projectors, and these projectors form a flag. The key observation is that the selected Schur subspace at each level is the kernel of a real polynomial in $\boldsymbol{A}$ that depends only on $\boldsymbol{A}$ and the level, so it cannot depend on the Schur form. However, in this appendix section, *we omit the eigenvalue grouping tolerance choice in remark 3 and assume exact eigenvalue groupings.*

Notation. For a square matrix Fix $\boldsymbol{A} \in U_L$ and let $\chi_{\boldsymbol{A}}(z) = \det(z\boldsymbol{I}_k - \boldsymbol{A})$ denote its characteristic polynomial. Each $\lambda \in \sigma(\boldsymbol{A})$ is a root of $\chi_{\boldsymbol{A}}$ of multiplicity d_λ (its algebraic multiplicity), so $\chi_{\boldsymbol{A}}(z) = \prod_{\lambda\in\sigma(\boldsymbol{A})}(z-\lambda)^{d_\lambda}$. Let $\rho_1 \geq \cdots \geq \rho_d$ be the eigenvalue magnitudes of $\boldsymbol{A}$, listed with algebraic multiplicity in decreasing order. Since $m(\boldsymbol{A}) \geq L$ (Dfn. 4), the gap dimensions $n_1 < \cdots < n_L$ of Dfn. 3 exist, and $\rho_{n_\ell} > \rho_{n_\ell+1}$ for each $\ell \leq L$. We set

$$\Lambda_\ell = \{\lambda \in \sigma(\boldsymbol{A}) : |\lambda| \geq \rho_{n_\ell}\}, \qquad q_\ell(z) = \prod_{\lambda\in\Lambda_\ell}(z-\lambda)^{d_\lambda}.$$

In words, q_ℓ is the factor of $\chi_{\boldsymbol{A}}$ that carries the n_ℓ largest-magnitude eigenvalues. Indeed, every magnitude past position n_ℓ is at most $\rho_{n_\ell+1} < \rho_{n_\ell}$, so $\deg q_\ell = \sum_{\lambda\in\Lambda_\ell} d_\lambda = n_\ell$. Notice that q_ℓ is determined by $\boldsymbol{A}$ and ℓ alone. Since $|\bar{\lambda}| = |\lambda|$, the set Λ_ℓ is cut out by a condition on modulus alone and is closed under conjugation; as $\boldsymbol{A}$ is real, $d_{\bar{\lambda}} = d_\lambda$, so q_ℓ has real coefficients.

Proposition 3 ($\boldsymbol{\Phi}$ is well defined)**.** *Let $\mathrm{Sym}(d)^L$ denote the set of L-tuples of real symmetric $d \times d$ matrices, and let $\boldsymbol{A} \in U_L$. For every $\ell \leq L$, every real Schur decomposition and every reordering used by alg. 1 produce a basis $\boldsymbol{V}_\ell$ with orthonormal columns and $\mathrm{range}(\boldsymbol{V}_\ell) = \ker q_\ell(\boldsymbol{A})$. Consequently the map $\boldsymbol{\Phi} : U_L \to \mathrm{Sym}(d)^L$ of Dfn. 2 is well defined.*

Proof of Prop. 3. Fix $\ell \leq L$ and write $n = n_\ell$. Let $\boldsymbol{A} = \boldsymbol{Z}\boldsymbol{T}\boldsymbol{Z}^\top$ be any real Schur decomposition with $\boldsymbol{T}$ in standard form, and let π_ℓ be the selection mask of alg. 1 which chooses the n diagonal positions of $\boldsymbol{T}$ with the ℓ largest eigenvalue magnitudes.

Step 1: the mask selects whole blocks. By remark 2, a 2×2 diagonal block of $\boldsymbol{T}$ contributes one magnitude ρ twice. If π_ℓ split such a block, we would have $\rho_n \leq \rho \leq \rho_{n+1}$, contradicting the gap. Hence π_ℓ selects exactly the diagonal blocks of magnitude at least ρ_n, which together carry n eigenvalues counted with multiplicity.

Step 2: the reordered form exists. By Step 1 and the reordering for the real Schur form (Bai & Demmel, 1993), there is an orthogonal $\boldsymbol{Z}_\ell$ with

$$\boldsymbol{A} = \boldsymbol{Z}_\ell\,\widetilde{\boldsymbol{T}}\,\boldsymbol{Z}_\ell^\top, \qquad \widetilde{\boldsymbol{T}} = \begin{bmatrix}\boldsymbol{T}_{11} & \boldsymbol{T}_{12}\\ 0 & \boldsymbol{T}_{22}\end{bmatrix}, \qquad \boldsymbol{T}_{11} \in \mathbb{R}^{n\times n}, \tag{7}$$

whose leading diagonal blocks in $\boldsymbol{T}_{11}$ are those selected by π_ℓ. In particular, the procedure returns $\boldsymbol{V}_\ell = \boldsymbol{Z}_\ell[:, 1{:}n]$, which has orthonormal columns. Reordering preserves the eigenvalues of each block, so all eigenvalues of Λ_ℓ lie in $\boldsymbol{T}_{11}$. Since the characteristic polynomial is invariant under similarity, and the determinant of a block upper-triangular matrix is the product of the determinants of its diagonal blocks, $\chi_{\boldsymbol{A}} = \chi_{\widetilde{\boldsymbol{T}}} = \chi_{\boldsymbol{T}_{11}}\chi_{\boldsymbol{T}_{22}}$. The roots of $\chi_{\boldsymbol{T}_{11}}$ lie in Λ_ℓ and those of $\chi_{\boldsymbol{T}_{22}}$ do not, so comparing root multiplicities gives $\chi_{\boldsymbol{T}_{11}} = q_\ell$.

Step 3: the subspace is $\ker q_\ell(\boldsymbol{A})$. By equation 7 and compatibility with similarity, $q_\ell(\boldsymbol{A}) = \boldsymbol{Z}_\ell\, q_\ell(\widetilde{\boldsymbol{T}})\, \boldsymbol{Z}_\ell^\top$. Powers of $\widetilde{\boldsymbol{T}}$ are block upper triangular with diagonal blocks $\boldsymbol{T}_{11}^k$ and $\boldsymbol{T}_{22}^k$, so

$$q_\ell(\boldsymbol{A}) = \boldsymbol{Z}_\ell\,\boldsymbol{B}_\ell\,\boldsymbol{Z}_\ell^\top, \qquad \boldsymbol{B}_\ell = \begin{bmatrix} q_\ell(\boldsymbol{T}_{11}) & \boldsymbol{X}\\ 0 & q_\ell(\boldsymbol{T}_{22})\end{bmatrix},$$

for some $\boldsymbol{X} \in \mathbb{R}^{n\times(d-n)}$. By Cayley–Hamilton, $q_\ell(\boldsymbol{T}_{11}) = \chi_{\boldsymbol{T}_{11}}(\boldsymbol{T}_{11}) = 0$. For $\lambda \in \Lambda_\ell$, $\det(\lambda \boldsymbol{I} - \boldsymbol{T}_{22}) = \chi_{\boldsymbol{T}_{22}}(\lambda)$ is a product of factors $\lambda - \mu$ over the eigenvalues μ of $\boldsymbol{T}_{22}$, and each satisfies $|\lambda - \mu| \geq |\lambda| - |\mu| \geq \rho_n - \rho_{n+1} > 0$. Hence each $\boldsymbol{T}_{22} - \lambda \boldsymbol{I}$ is invertible, and so is $q_\ell(\boldsymbol{T}_{22}) = \prod_{\lambda\in\Lambda_\ell}(\boldsymbol{T}_{22} - \lambda \boldsymbol{I})^{d_\lambda}$. For $(\boldsymbol{u}, \boldsymbol{w}) \in \mathbb{R}^n \times \mathbb{R}^{d-n}$ we therefore have

$$\boldsymbol{B}_\ell \begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{w} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}\boldsymbol{w} \\ q_\ell(\boldsymbol{T}_{22})\boldsymbol{w} \end{bmatrix},$$

which vanishes if and only if $\boldsymbol{w} = 0$; that is, $\ker \boldsymbol{B}_\ell = \mathbb{R}^n \times \{0\}$. Since $\boldsymbol{Z}_\ell$ is orthogonal, $\ker q_\ell(\boldsymbol{A}) = \boldsymbol{Z}_\ell \ker \boldsymbol{B}_\ell$, which gives

$$\ker q_\ell(\boldsymbol{A}) = \boldsymbol{Z}_\ell\big(\mathbb{R}^n \times \{0\}\big) = \operatorname{range}(\boldsymbol{V}_\ell). \tag{8}$$

Conclusion. The left-hand side of equation 8 depends only on $\boldsymbol{A}$ and ℓ, so every Schur decomposition and every reordering produce the same subspace $\mathcal{V}_\ell := \ker q_\ell(\boldsymbol{A})$, of dimension n. The projector does not depend on the basis either: for any orthonormal basis $\boldsymbol{V}_\ell$ of $\mathcal{V}_\ell$, $\boldsymbol{V}_\ell \boldsymbol{V}_\ell^\top$ is the unique orthogonal projector onto $\mathcal{V}_\ell$. Finally, each $\boldsymbol{\Pi}_\ell = \boldsymbol{V}_\ell \boldsymbol{V}_\ell^\top$ is symmetric, so $\boldsymbol{\Phi}(\boldsymbol{A}) = \frac{1}{\sqrt{2}}(\boldsymbol{\Pi}_1, \ldots, \boldsymbol{\Pi}_L) \in \operatorname{Sym}(d)^L$. □

D Lifting the flag into shared coordinates

Remark 6 compares flags only after lifting them into the shared ambient space. We fix the coordinates that make this precise and show that the lift removes the gauge freedom of the reduced basis.

Coordinates.. Let $a \in \mathbb{R}^p_{>0}$ be the areas of the p grid cells and $W = \operatorname{diag}(a)$. Area-weighted PCA works in the inner product $\langle \boldsymbol{x}, \boldsymbol{y}\rangle_W = \boldsymbol{x}^\top W \boldsymbol{y}$, which discretizes the L^2 inner product on the sphere, so its basis satisfies $\boldsymbol{U}_{\text{phys}}^\top W \boldsymbol{U}_{\text{phys}} = \boldsymbol{I}_d$ and projectors that are orthogonal in this inner product are not symmetric matrices. We therefore work in $\sqrt{a}$-weighted coordinates: $\boldsymbol{x} \mapsto W^{1/2}\boldsymbol{x}$ is an isometry from $(\mathbb{R}^p, \langle\cdot,\cdot\rangle_W)$ onto Euclidean $\mathbb{R}^p$, and $\boldsymbol{U} = W^{1/2}\boldsymbol{U}_{\text{phys}}$ satisfies $\boldsymbol{U}^\top \boldsymbol{U} = \boldsymbol{I}_d$. The lifted bases $\boldsymbol{U}\boldsymbol{V}_\ell$ then have orthonormal columns, the lifted projectors $\boldsymbol{U}\boldsymbol{\Pi}_\ell \boldsymbol{U}^\top$ are orthogonal projectors and hence lie in $\operatorname{Sym}(p)$, and Prop. 1 applies with p in place of d. We compare flags from different realizations by evaluating equation 6 on these lifted bases, which requires a common ambient grid (we regrid before fitting; see sec. H).

Gauge freedom of the reduced basis.. Principal components are determined only up to sign, and components of equal variance only up to a rotation of their span. This gauge freedom differs from the eigenvector phase in sec. B, and the lift removes it. Replacing $\boldsymbol{U}$ by $\boldsymbol{U}\boldsymbol{Q}$ with $\boldsymbol{Q} \in O(d)$ replaces the fitted operator by $\boldsymbol{Q}^\top \boldsymbol{A}\boldsymbol{Q}$, which has the same spectrum and hence the same q_ℓ. Since $q_\ell(\boldsymbol{Q}^\top \boldsymbol{A}\boldsymbol{Q}) = \boldsymbol{Q}^\top q_\ell(\boldsymbol{A})\boldsymbol{Q}$, Prop. 3 gives the new level-ℓ subspace $\operatorname{range}(\boldsymbol{Q}^\top \boldsymbol{V}_\ell)$, and $(\boldsymbol{U}\boldsymbol{Q})(\boldsymbol{Q}^\top \boldsymbol{V}_\ell) = \boldsymbol{U}\boldsymbol{V}_\ell$ leaves the lifted flag and its projectors unchanged. The lift thus quotients out the choice of reduced basis, just as the projector representation of Dfn. 2 quotients out the eigenvector phase.

E Metric and valid kernel

Proof of Prop. 1. Write $\mathcal{F}_{L,d}$ for the set of flags in $\mathbb{R}^d$ with any signature of length L. The map $\mathrm{fl} : U_L \to \mathcal{F}_{L,d}$ sending an operator to its invariant flag is well defined by Prop. 3 and **??**.

For a subspace $\mathcal{V} \subseteq \mathbb{R}^d$ let

$$\iota(\mathcal{V}) = \tfrac{1}{\sqrt{2}}\boldsymbol{\Pi}_{\mathcal{V}}; \qquad (\mathcal{V}, \mathcal{W}) = \|\iota(\mathcal{V}) - \iota(\mathcal{W})\|_F.$$

Define $\boldsymbol{\Psi} : \mathcal{F}_{L,d} \to \operatorname{Sym}(d)^L$ as application of ι at each level of a flag; then $\boldsymbol{\Phi} = \boldsymbol{\Psi} \circ \mathrm{fl}$ by Dfn. 2.

We equip $\operatorname{Sym}(d)^L$ with the inner product $\langle \boldsymbol{X}, \boldsymbol{Y}\rangle = \sum_{\ell=1}^L \langle \boldsymbol{X}_\ell, \boldsymbol{Y}_\ell\rangle_F$, which makes it a Euclidean space, and write $\|\cdot\|_F$ for the induced norm (the Frobenius norm of the stacked matrices). Now set

$$d_{\mathcal{FL}}(\mathcal{F}, \mathcal{F}') = \|\boldsymbol{\Psi}(\mathcal{F}) - \boldsymbol{\Psi}(\mathcal{F}')\|_F, \qquad d_{\mathcal{FL}}(\boldsymbol{A}, \boldsymbol{A}') = \|\boldsymbol{\Phi}(\boldsymbol{A}) - \boldsymbol{\Phi}(\boldsymbol{A}')\|_F = d_{\mathcal{FL}}\big(\mathrm{fl}(\boldsymbol{A}), \mathrm{fl}(\boldsymbol{A}')\big).$$

Notice that $d^2_{\mathcal{FL}}(\mathcal{F}, \mathcal{F}') = \sum_{\ell=1}^{L} d_c^2(\mathcal{V}_\ell, \mathcal{V}'_\ell)$.

Step 1: $\boldsymbol{\Psi}$ is injective. If $\boldsymbol{\Psi}(\mathcal{F}) = \boldsymbol{\Psi}(\mathcal{F}')$, then $\boldsymbol{\Pi}_{\mathcal{V}_\ell} = \boldsymbol{\Pi}_{\mathcal{V}'_\ell}$ for every ℓ, and taking ranges gives $\mathcal{V}_\ell = \mathcal{V}'_\ell$, hence $\mathcal{F} = \mathcal{F}'$.

Step 2: metric and pseudometric. The pullback $\rho(f(\cdot), f(\cdot))$ of a metric ρ along any map f is a pseudometric, since nonnegativity, symmetry, and the triangle inequality hold pointwise, and it is a metric exactly when f is injective, since $\rho(f(s), f(t)) = 0$ if and only if $f(s) = f(t)$.

With $\rho = \|\cdot\|_F$, taking $f = \boldsymbol{\Psi}$ shows by Step 1 that $d_{\mathcal{FL}}$ is a metric on $\mathcal{F}_{L,d}$. Taking $f = \boldsymbol{\Phi}$ shows that it is a pseudometric on U_L. More precisely, Step 1 gives $d_{\mathcal{FL}}(\boldsymbol{A}, \boldsymbol{A}') = 0$ if and only if $\mathrm{fl}(\boldsymbol{A}) = \mathrm{fl}(\boldsymbol{A}')$, and example E.1 exhibits $\boldsymbol{A} \neq \boldsymbol{A}'$ with this property, so $d_{\mathcal{FL}}$ is not a metric on U_L.

Step 3: the Gaussian kernel. Let f be $\boldsymbol{\Psi}$ on $\mathcal{F}_{L,d}$ or $\boldsymbol{\Phi}$ on U_L, so that $d^2_{\mathcal{FL}}(s,t) = \|f(s) - f(t)\|_F^2$ in both cases. This kernel is symmetric, and $d^2_{\mathcal{FL}}$ is negative definite in the sense of Jayasumana et al. (2015, Def. 5.1) (cf. their Lemma 5.5). Hence $\exp(-\gamma\, d^2_{\mathcal{FL}})$ is positive definite for every $\gamma > 0$ by Schoenberg's theorem (Jayasumana et al., 2015, Thm. 5.2). The computation uses only that f maps into an inner product space, not that $d_{\mathcal{FL}}$ is a metric, so it covers U_L as well as flags. □

Recording a subspace as its projector trades the curved geometry of the Grassmannian for the flat geometry of $\mathrm{Sym}(d)$. That single move buys three properties at once: a metric on flags (even across levels of different dimension), an isometric Euclidean embedding of flags, and a positive-definite Gaussian kernel. On operators, $d_{\mathcal{FL}}$ is only a pseudometric, but the kernel stays positive definite, and a positive semidefinite Gram matrix is all the downstream kPCA step requires. The kernel is where flatness matters most: on a complete Riemannian manifold, a Gaussian kernel built on geodesic distance is positive definite for every $\gamma > 0$ only if the manifold is isometric to a Euclidean space, which rules out compact manifolds such as the Grassmannian (Jayasumana et al., 2015, Thm. 6.2). For $L = 1$ and a fixed dimension, our kernel is exactly their projection Gaussian kernel (Jayasumana et al., 2015, Cor. 6.4); Prop. 1 extends it to flags, whose levels have different dimensions.

Example E.1 (The flag forgets the spectrum)**.** *Work in $\mathbb{R}^2$ with truncation level $L = 1$, so that a flag is a single line. Take the diagonal operators $\boldsymbol{A} = \mathrm{diag}(2, 1)$ and $\boldsymbol{A}' = \mathrm{diag}(5, 1)$. Both have distinct eigenvalue magnitudes, so each has a well-defined invariant flag, and in both cases the leading eigenvector is $\boldsymbol{e}_1$. Hence $\mathrm{fl}(\boldsymbol{A}) = \mathrm{fl}(\boldsymbol{A}') = \mathrm{span}\{\boldsymbol{e}_1\}$ and $d_{\mathcal{FL}}(\boldsymbol{A}, \boldsymbol{A}') = 0$, even though $\|\boldsymbol{A} - \boldsymbol{A}'\|_F = 3$. Nothing about this is subtle: we widened the spectral gap and the flag did not notice, because a flag is a statement about* directions*, not about* magnitudes*. The same collapse occurs for $\boldsymbol{A}' = c\,\boldsymbol{A}$ with $c > 0$.*

F Stability bounds

We prove Prop. 2. We first record the conversion from spectral to orthogonal projectors.

Lemma 1 (From oblique to orthogonal projectors)**.** *Let $\boldsymbol{P}, \boldsymbol{P}'$ be real idempotent matrices with orthogonal range projectors $\boldsymbol{\Pi}, \boldsymbol{\Pi}'$. Then $\|\boldsymbol{\Pi} - \boldsymbol{\Pi}'\|_F \le \sqrt{2}\,\|\boldsymbol{P} - \boldsymbol{P}'\|_F$. The ranks need not be equal.*

Proof. Since $\boldsymbol{P}\boldsymbol{\Pi} = \boldsymbol{\Pi}$ and $(\boldsymbol{I} - \boldsymbol{\Pi}')\boldsymbol{P}' = \boldsymbol{0}$, we have $(\boldsymbol{I} - \boldsymbol{\Pi}')\boldsymbol{\Pi} = (\boldsymbol{I} - \boldsymbol{\Pi}')(\boldsymbol{P} - \boldsymbol{P}')\boldsymbol{\Pi}$, so $\|(\boldsymbol{I} - \boldsymbol{\Pi}')\boldsymbol{\Pi}\|_F \le \|\boldsymbol{P} - \boldsymbol{P}'\|_F$; symmetrically, $\|(\boldsymbol{I} - \boldsymbol{\Pi})\boldsymbol{\Pi}'\|_F \le \|\boldsymbol{P} - \boldsymbol{P}'\|_F$. The two terms of $\boldsymbol{\Pi} - \boldsymbol{\Pi}' = (\boldsymbol{I} - \boldsymbol{\Pi}')\boldsymbol{\Pi} - \boldsymbol{\Pi}'(\boldsymbol{I} - \boldsymbol{\Pi})$ are Frobenius-orthogonal, because $\mathrm{tr}\big(\boldsymbol{\Pi}(\boldsymbol{I} - \boldsymbol{\Pi}')\boldsymbol{\Pi}'(\boldsymbol{I} - \boldsymbol{\Pi})\big) = 0$. Hence $\|\boldsymbol{\Pi} - \boldsymbol{\Pi}'\|_F^2 \le 2\|\boldsymbol{P} - \boldsymbol{P}'\|_F^2$. □

Proof of Prop. 2. Fix ℓ, abbreviate $g = g_\ell$ and $r = r_\ell$, and orient $\Gamma = \{z \in \mathbb{C} : |z| = r\}$ counterclockwise.

Resolvent bound. For $\boldsymbol{R}(z) = (z\boldsymbol{I} - \boldsymbol{A})^{-1}$, diagonalization gives

$$\|\boldsymbol{R}(z)\|_2 \le \frac{\kappa}{\min_{\lambda \in \sigma(\boldsymbol{A})} |z - \lambda|} \le \frac{2\kappa}{g} \quad (z \in \Gamma),$$

because $|z - \lambda| \ge \big||\lambda| - r\big| \ge g/2$. This is the resolvent estimate in the proof of Trefethen (2005, §2, Theorem 2.3).

Perturbed resolvent. For $0 \le t \le 1$ let $\boldsymbol{A}_t = \boldsymbol{A} + t\boldsymbol{E}$, so that $\boldsymbol{A}_0 = \boldsymbol{A}$ and $\boldsymbol{A}_1 = \boldsymbol{A}'$. Since $(z\boldsymbol{I} - \boldsymbol{A})\boldsymbol{R}(z) = \boldsymbol{I}$,

$$(z\boldsymbol{I} - \boldsymbol{A})\big(\boldsymbol{I} - t\boldsymbol{R}(z)\boldsymbol{E}\big) = (z\boldsymbol{I} - \boldsymbol{A}) - t\,(z\boldsymbol{I} - \boldsymbol{A})\boldsymbol{R}(z)\boldsymbol{E} = z\boldsymbol{I} - \boldsymbol{A}_t.$$

Because $0 \le t \le 1$, $\|\boldsymbol{E}\|_2 \le \|\boldsymbol{E}\|_F$, and $\|\boldsymbol{E}\|_F \le \delta/(4\kappa)$ by hypothesis,

$$\|t\boldsymbol{R}(z)\boldsymbol{E}\|_2 \le \|\boldsymbol{R}(z)\|_2\|\boldsymbol{E}\|_2 \le \frac{2\kappa}{g}\|\boldsymbol{E}\|_F \le \frac{2\kappa}{g}\cdot\frac{\delta}{4\kappa} = \frac{\delta}{2g} \le \frac{1}{2}.$$

If $\|\boldsymbol{X}\|_2 < 1$, then $\boldsymbol{I}-\boldsymbol{X}$ is invertible, $(\boldsymbol{I}-\boldsymbol{X})^{-1} = \sum_{k\ge 0}\boldsymbol{X}^k$, and $\|(\boldsymbol{I}-\boldsymbol{X})^{-1}\|_2 \le (1-\|\boldsymbol{X}\|_2)^{-1}$ (Kato, 1995, Ch. I, §4.4). Applied to $\boldsymbol{X} = t\boldsymbol{R}(z)\boldsymbol{E}$, this makes $\boldsymbol{I} - t\boldsymbol{R}(z)\boldsymbol{E}$ invertible with $\|(\boldsymbol{I} - t\boldsymbol{R}(z)\boldsymbol{E})^{-1}\|_2 \le 2$. Hence $z\boldsymbol{I} - \boldsymbol{A}_t$, a product of two invertible matrices, is invertible for every $z \in \Gamma$, so no eigenvalue of $\boldsymbol{A}_t$ lies on Γ. Inverting a product reverses the order of its factors, so

$$\boldsymbol{R}_t(z) := (z\boldsymbol{I} - \boldsymbol{A}_t)^{-1} = \big(\boldsymbol{I} - t\boldsymbol{R}(z)\boldsymbol{E}\big)^{-1}\boldsymbol{R}(z), \qquad \|\boldsymbol{R}_t(z)\|_2 \le \frac{4\kappa}{g}, \tag{9}$$

where the bound is $\|\boldsymbol{R}_t(z)\|_2 \le \|(\boldsymbol{I} - t\boldsymbol{R}(z)\boldsymbol{E})^{-1}\|_2\,\|\boldsymbol{R}(z)\|_2 \le 2\cdot 2\kappa/g$.

Resolvent differences. Let $0 \le s, t \le 1$ and $z \in \Gamma$. By the definition of $\boldsymbol{R}_t(z)$ in equation 9,

$$\boldsymbol{R}_t(z) - \boldsymbol{R}_s(z) = \boldsymbol{R}_t(z)\big[(z\boldsymbol{I} - \boldsymbol{A}_s) - (z\boldsymbol{I} - \boldsymbol{A}_t)\big]\boldsymbol{R}_s(z) = (t-s)\,\boldsymbol{R}_t(z)\boldsymbol{E}\boldsymbol{R}_s(z).$$

The first equality follows by multiplying out, because $(z\boldsymbol{I}-\boldsymbol{A}_s)\boldsymbol{R}_s(z) = \boldsymbol{I}$ and $\boldsymbol{R}_t(z)(z\boldsymbol{I}-\boldsymbol{A}_t) = \boldsymbol{I}$. The second holds because $(z\boldsymbol{I} - \boldsymbol{A}_s) - (z\boldsymbol{I} - \boldsymbol{A}_t) = \boldsymbol{A}_t - \boldsymbol{A}_s = (t-s)\boldsymbol{E}$, by the definition of $\boldsymbol{A}_t$. Since $\|\boldsymbol{R}_t(z)\boldsymbol{E}\boldsymbol{R}_s(z)\|_F \le \|\boldsymbol{R}_t(z)\|_2\|\boldsymbol{E}\|_F\|\boldsymbol{R}_s(z)\|_2$,

$$\|\boldsymbol{R}_t(z) - \boldsymbol{R}_s(z)\|_F \le |t-s|\,\|\boldsymbol{R}_t(z)\|_2\,\|\boldsymbol{R}_s(z)\|_2\,\|\boldsymbol{E}\|_F. \tag{10}$$

Spectral projectors. Let

$$\boldsymbol{P}_{\mathrm{in}}(t) = \frac{1}{2\pi i}\oint_\Gamma \boldsymbol{R}_t(z)\,dz$$

be the Riesz projector of $\boldsymbol{A}_t$ for its eigenvalues inside Γ. It is idempotent; its range is the sum of the generalized eigenspaces of $\boldsymbol{A}_t$ for the eigenvalues inside Γ, and its kernel is the corresponding sum for the eigenvalues outside Γ (Kato, 1995, Ch. I, §5.3–5.4; Ch. II, §1.4) [3] Its complement $\boldsymbol{P}(t) = \boldsymbol{I} - \boldsymbol{P}_{\mathrm{in}}(t)$ is also a projector, and $\operatorname{range}\boldsymbol{P}(t) = \ker\boldsymbol{P}_{\mathrm{in}}(t)$ is the sum of the generalized eigenspaces of $\boldsymbol{A}_t$ for the eigenvalues outside Γ. Both projectors are real. Because $\boldsymbol{A}_t$ is real, complex conjugation maps the generalized eigenspace of $\boldsymbol{A}_t$ for λ onto the one for $\bar\lambda$, and $|\bar\lambda| = |\lambda|$. Hence $\overline{\boldsymbol{P}_{\mathrm{in}}(t)}$ is a projector with the same range and kernel as $\boldsymbol{P}_{\mathrm{in}}(t)$, and a projector is determined by its range and kernel.

Projector differences. For $0 \le s, t \le 1$, $\boldsymbol{P}(t) - \boldsymbol{P}(s) = -\big(\boldsymbol{P}_{\mathrm{in}}(t) - \boldsymbol{P}_{\mathrm{in}}(s)\big)$. The length $2\pi r$ of Γ cancels the prefactor $1/(2\pi)$, so equation 10 gives

$$\begin{aligned}\|\boldsymbol{P}(t) - \boldsymbol{P}(s)\|_F &= \Big\|\frac{1}{2\pi i}\oint_\Gamma \big(\boldsymbol{R}_t(z) - \boldsymbol{R}_s(z)\big)\,dz\Big\|_F \le \frac{2\pi r}{2\pi}\sup_{z\in\Gamma}\|\boldsymbol{R}_t(z) - \boldsymbol{R}_s(z)\|_F \\ &\le r\,|t-s|\,\|\boldsymbol{E}\|_F\sup_{z\in\Gamma}\|\boldsymbol{R}_t(z)\|_2\,\|\boldsymbol{R}_s(z)\|_2.\end{aligned} \tag{11}$$

Counting eigenvalues. The range of $\boldsymbol{P}(t)$ is the direct sum of the generalized eigenspaces of $\boldsymbol{A}_t$ for its eigenvalues outside Γ, and each generalized eigenspace has dimension equal to the algebraic multiplicity of its eigenvalue (Kato, 1995, Ch. I, §5.4). Hence $\operatorname{rank}\boldsymbol{P}(t)$ is the number of eigenvalues of $\boldsymbol{A}_t$ outside Γ, counted with algebraic multiplicity. For a projector $\boldsymbol{P}$, every $\boldsymbol{v} \in \mathbb{C}^d$ splits as $\boldsymbol{v} = \boldsymbol{P}\boldsymbol{v} + (\boldsymbol{I} - \boldsymbol{P})\boldsymbol{v}$ with $\boldsymbol{P}\boldsymbol{v} \in \operatorname{range}\boldsymbol{P}$ and $(\boldsymbol{I} - \boldsymbol{P})\boldsymbol{v} \in \ker\boldsymbol{P}$, and the two subspaces meet only in $\boldsymbol{0}$ because $\boldsymbol{P}\boldsymbol{v} = \boldsymbol{v}$ on the range. In a basis adapted to $\mathbb{C}^d = \operatorname{range}\boldsymbol{P} \oplus \ker\boldsymbol{P}$, therefore, $\boldsymbol{P} = \operatorname{diag}(\boldsymbol{I}_k, \boldsymbol{0})$, so $\operatorname{tr}\boldsymbol{P} = k = \operatorname{rank}\boldsymbol{P}$. By equation 11 and equation 9, $\boldsymbol{P}(t)$ is Lipschitz in t, so $t \mapsto \operatorname{tr}\boldsymbol{P}(t) = \operatorname{rank}\boldsymbol{P}(t)$ is a continuous integer-valued function on $[0,1]$ and hence constant.

At $t = 0$ this count is n_ℓ: $\boldsymbol{A}$ has exactly n_ℓ eigenvalues of modulus at least $\rho_{n_\ell} > r$, and the others have modulus at most $\rho_{n_\ell+1} < r$. Evaluating at $t = 1$, where $\boldsymbol{A}_1 = \boldsymbol{A}'$, shows that $\boldsymbol{A}'$ has exactly n_ℓ

[3] Kato writes the resolvent as $(T-\zeta)^{-1}$, the negative of our $(\zeta\boldsymbol{I}-T)^{-1}$. His projector $-\frac{1}{2\pi i}\oint_\Gamma (T-\zeta)^{-1}\,d\zeta$ is therefore our $\boldsymbol{P}_{\mathrm{in}}(t)$ when $T = \boldsymbol{A}_t$.

eigenvalues outside Γ. Specifically, none of these lie on Γ, by the invertibility of $z\boldsymbol{I} - \boldsymbol{A}_1$ established above. Hence $\rho'_{n_\ell} > r > \rho'_{n_\ell+1}$, and the eigenvalues of $\boldsymbol{A}'$ outside Γ are its n_ℓ eigenvalues of largest modulus. Since $\operatorname{sig}_L(\boldsymbol{A}') = \boldsymbol{n}$, these form the top ℓ magnitude groups of $\boldsymbol{A}'$, so the real projector $\boldsymbol{P}(1)$ has range $\mathcal{V}'_\ell$ (Dfn. 2; equivalently $\ker q'_\ell(\boldsymbol{A}')$ by Prop. 3). Likewise $\operatorname{range} \boldsymbol{P}(0) = \mathcal{V}_\ell$.

Conclusion. At $(t, s) = (1, 0)$, equation 9 gives $\|R_1(z)\|_2 \le 4\kappa/g$, and $\|R_0(z)\|_2 = \|\boldsymbol{R}(z)\|_2 \le 2\kappa/g$ by the resolvent bound, so equation 11 yields

$$\|\boldsymbol{P}(1) - \boldsymbol{P}(0)\|_F \le r \cdot \frac{4\kappa}{g} \cdot \frac{2\kappa}{g} \|\boldsymbol{E}\|_F = \frac{8r\kappa^2}{g^2} \|\boldsymbol{E}\|_F.$$

As $\boldsymbol{P}(0)$ and $\boldsymbol{P}(1)$ are real idempotents with ranges $\mathcal{V}_\ell$ and $\mathcal{V}'_\ell$, lemma 1 and equation 3 imply

$$d_c(\mathcal{V}_\ell, \mathcal{V}'_\ell) \le \frac{8r_\ell\kappa^2}{g_\ell^2} \|\boldsymbol{E}\|_F.$$

Summing the squared estimates over ℓ in equation 4 gives the first inequality in equation 5; $0 < r_\ell < 1$ and $g_\ell \ge \delta$ give the second. □

G SYNTHETIC DATA GENERATION

We first define the systems, then the dynamics and observation model, and finally the individual experiments. The notation in this section is consistent and self-contained, but does not align with the Tab. 1 because it is impossible to find enough variable names without overlapping with previously defined variables.

Systems. Every system is a linear map $\boldsymbol{A} = \boldsymbol{Q}\,\boldsymbol{\Lambda}(\boldsymbol{\theta})\,\boldsymbol{Q}^\top \in \mathbb{R}^{n\times n}$ with $n = 14$ and control dimension $m = 1$. For $\boldsymbol{\theta} = (\lambda_s, \rho, \varphi, \lambda_f)$, the spectral template is

$$\boldsymbol{\Lambda}(\boldsymbol{\theta}) = \operatorname{blockdiag}\big(\lambda_s,\ \rho R(\varphi),\ \lambda_f,\ \nu_1, \ldots, \nu_{10}\big), \qquad \boldsymbol{R}(\varphi) = \begin{pmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{pmatrix},$$

i.e., a slow real mode, an oscillatory pair $\rho e^{\pm i\varphi}$, a faster real mode, and ten nuisance modes whose eigenvalues $\nu_k = 0.05 + 0.2\,(k-1)/9$ are identical across all systems. We use

$$\boldsymbol{\theta}_1 = (0.95,\ 0.80,\ 25^\circ,\ 0.60).$$

This satisfies $\lambda_s > \rho > \lambda_f > \max_k \nu_k = 0.25$, so sorting eigenvalues by modulus (equivalently, by decay time $-1/\ln|\lambda|$) always yields the same hierarchy. The rotation $\boldsymbol{Q}$ is the orthogonal factor of the QR decomposition of a 14×14 matrix with i.i.d. $\mathcal{N}(0, 1)$ entries.

Writing $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_{14}$ for the columns of $\boldsymbol{Q}$, the true flag of $\boldsymbol{A}$ truncated to three levels is

$$\operatorname{span}(\boldsymbol{q}_1) \subset \operatorname{span}(\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{q}_3) \subset \operatorname{span}(\boldsymbol{q}_1, \ldots, \boldsymbol{q}_4) \subset \mathbb{R}^{14},$$

of signature $\boldsymbol{n} = (1, 3, 4)$. Notice that two systems with the same $\boldsymbol{\Lambda}$ but different $\boldsymbol{Q}$ share every eigenvalue and differ through these subspaces (and through their nuisance subspaces, which are the orthogonal complements).

Dynamics and observations. Given $\boldsymbol{A}$, we simulate

$$\boldsymbol{x}_{t+1} = \boldsymbol{A}\boldsymbol{x}_t + \boldsymbol{b}u_t + \boldsymbol{w}_t \quad (t = 0, \ldots, T-1), \qquad \boldsymbol{y}_t = \boldsymbol{x}_t + \boldsymbol{v}_t \quad (t = 0, \ldots, T),$$

with $\boldsymbol{x}_0 \sim \mathcal{N}(0, \boldsymbol{I})$, $u_t \sim \mathcal{N}(0, 1)$, $\boldsymbol{w}_t \sim \mathcal{N}(0, \sigma^2\boldsymbol{I})$, and $\boldsymbol{v}_t \sim \mathcal{N}(0, \eta^2\boldsymbol{I})$, all independent. $\eta = 0$ except in the observation-noise ablation. The control vector $\boldsymbol{b} \in \mathbb{R}^{14\times 1}$, $b_j \sim \mathcal{N}(0, 1)$, is drawn once per experiment and shared by all systems. Within a trial, the same $K = 20$ control sequences $u^{(1)}, \ldots, u^{(K)}$ drive every system, so an experiment with C systems yields $C \times K$ labeled trajectories $\boldsymbol{Y} \in \mathbb{R}^{14\times(T+1)}$ that differ across systems only through $\boldsymbol{A}$, $\boldsymbol{x}_0$, and the noise.

We fit each trajectory independently by DMDc least squares with a timelag of 1. Every T in the table exceeds the 15 unknowns, so the least squares solution is unique.

Ablations. The ablations isolate differences in eigenvectors: three clusters $\boldsymbol{A}_c = \boldsymbol{Q}_c\boldsymbol{\Lambda}\boldsymbol{Q}_c^\top$, $c = 1, 2, 3$, share the spectrum $\boldsymbol{\Lambda} = \boldsymbol{\Lambda}(\boldsymbol{\theta}_1)$ (unless stated otherwise), and we sweep one factor at a time

Ablation	Swept	Grid	Fixed
Process noise	σ	$\{0.1, 0.2, 0.35, 0.5, 0.75, 1, 1.5, 2, 3, 5, 7.5, 10\}$	$T = 30$
Observation noise	η	$\{1, 1.5, 2, 2.5, 3, 3.5, 4, 5, 6, 7, 8, 10\}$	$T = 200$, $\sigma = 0.2$
Trajectory length	T	$\{17, 20, 22, 25, 30, 35, 40, 50, 60, 75, 100\}$	$\sigma = 0.2$
Eigenvector blend	α	$\{0.7, 0.725, \ldots, 1.0\}$	$T = 200$, $\sigma = 0.2$
Eigenvalue perturbation	ε	$\{0, 0.015, 0.03, 0.045, 0.06, 0.065, 0.07, 0.075, 0.08, 0.085, 0.09, 0.105, 0.12, 0.15\}$	$T = 200$, $\sigma = 0.2$

Table 2: Ablation sweeps. All use three clusters sharing one spectrum ($\mathbf{\Lambda}(\boldsymbol{\theta}_1)$, or $\mathbf{\Lambda}(\varepsilon)$ for the eigenvalue perturbation), $K = 20$ trajectories per cluster, 20 trials per sweep point, and $\eta = 0$ unless swept.

while fixing the others (Tab. 2). The rotations $\boldsymbol{Q}_c$ and the control vector $\boldsymbol{b}$ are drawn once and held fixed across all sweep points and trials; only the controls, initial conditions, and noise are redrawn.

Two ablations modify the construction itself. First, the *eigenvector blend* moves $\boldsymbol{Q}_c \in SO(14)$ along geodesics of $SO(14)$ toward a shared target $\boldsymbol{Q}_\star$ (Edelman et al., 1998), using the principal matrix logarithm,

$$\boldsymbol{Q}_c(\alpha) = \exp\big(\alpha \log(\boldsymbol{Q}_\star \boldsymbol{Q}_c^\top)\big)\, \boldsymbol{Q}_c, \qquad \boldsymbol{A}_c(\alpha) = \boldsymbol{Q}_c(\alpha)\, \Lambda\, \boldsymbol{Q}_c(\alpha)^\top,$$

so the clusters are unrelated at $\alpha = 0$ and identical at $\alpha = 1$. The target $\boldsymbol{Q}_\star \in SO(14)$, drawn in the same way as, and immediately after, the $\boldsymbol{Q}_c$.

Second, the *eigenvalue perturbation* uses $\mathbf{\Lambda}(\varepsilon) = \mathbf{\Lambda}(0.95-\varepsilon,\ 0.80+\varepsilon,\ 25^\circ,\ 0.60)$ with the rotations fixed, pushing the slow mode and the pair toward each other. Their moduli cross at $\varepsilon^\star = 0.075$: at $\varepsilon^\star$ the first two levels merge (so the three-level truncation of the true flag, of signature $(3, 4, 5)$, reaches the largest nuisance mode), and for $\varepsilon > \varepsilon^\star$ the true flag becomes $\mathrm{span}(\boldsymbol{q}_2, \boldsymbol{q}_3) \subset \mathrm{span}(\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{q}_3) \subset \mathrm{span}(\boldsymbol{q}_1, \ldots, \boldsymbol{q}_4)$, of signature $(2, 3, 4)$.

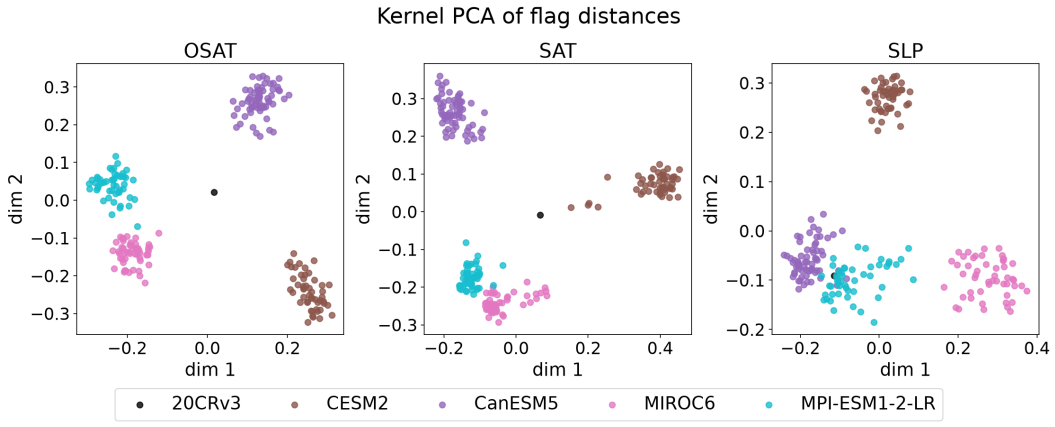
H Extended CMIP6 and reanalysis results

Figure 5 plots kPCA embeddings and distributions of distances to reanalysis for all 3 climate variables. We see the worst separation between ESMs for SLP.

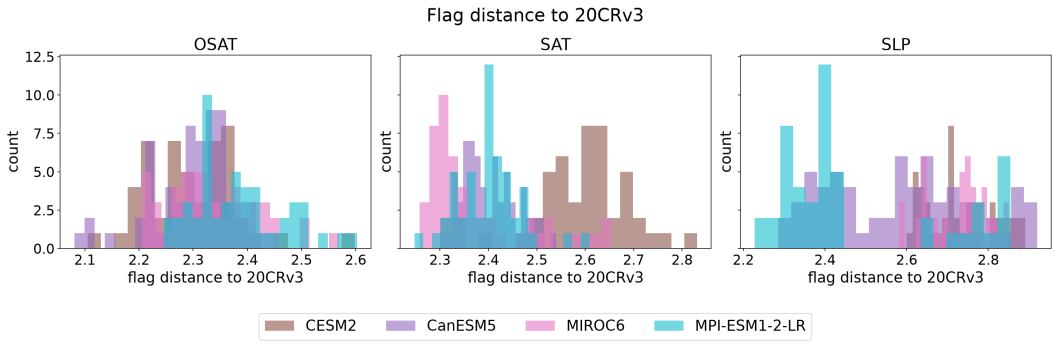
Figure 7 decomposes the distances by origin and shows the clearest structure in these data: for OSAT and SAT, within-ESM distances (mode near 2) are much smaller than both cross-ESM distances and distances to the reanalysis (both near 4), so in flag geometry the members of one model are far more like each other than like any other model or the reanalysis. For SLP the within- and cross-ESM distributions overlap, yet every ESM still sits farther from 20CRv3 than from itself.

Figure 6 ranks the ESMs based on how close they are to reanalysis data under our flag distance. We find that CESM2 is closest for OSAT, MIROC6 for SAT, and MPI-ESM1-2-LR for SLP.

Figure 8 shows the distribution of flag signatures detected by alg. 1. For OSAT, no ESM realization matches the reanalysis signature (R R C). For SAT, only 8 of 215 realizations match (7 CESM2, 1 MIROC6), and most CanESM5, MIROC6, and MPI-ESM1-2-LR realizations are all-real. For SLP, matches come only from CanESM5 (14) and MPI-ESM1-2-LR (19).



(a) kPCA embeddings of flag dissimilarity matrices.



(b) Flag dissimilarity between ensemble members and reanalysis.

Figure 5: A comprehensive comparison between ESM ensemble members and reanalysis for OSAT, SAT, and SLP variables.

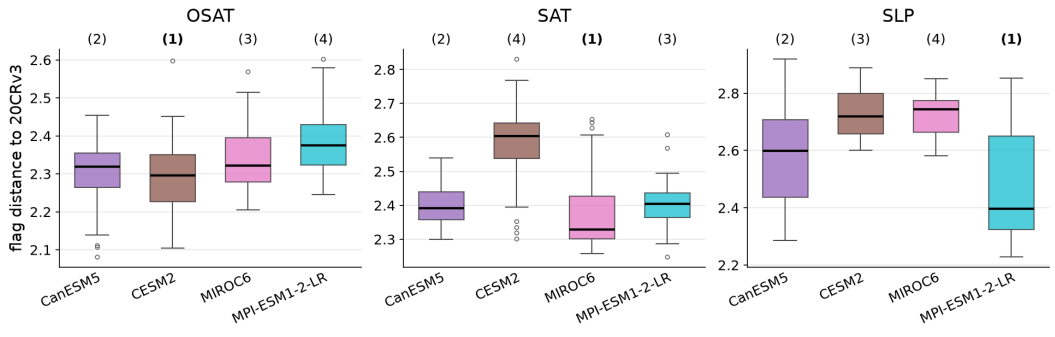


Figure 6: ESMs compared by flag distance to the 20CRv3 reanalysis for each climate variable (flag truncated to 3 levels). For each ESM, the marker shows the median over its realizations and the whiskers show the interquartile range; lower is closer to the reanalysis. Distances are computed between subspaces mapped onto each variable's physical grid. Because the variables differ in field and EOF truncation, distances are not comparable across panels, so models should be compared only within a panel. The closest model differs by variable: CESM2 for OSAT, MIROC6 for SAT, and MPI-ESM1-2-LR for SLP.

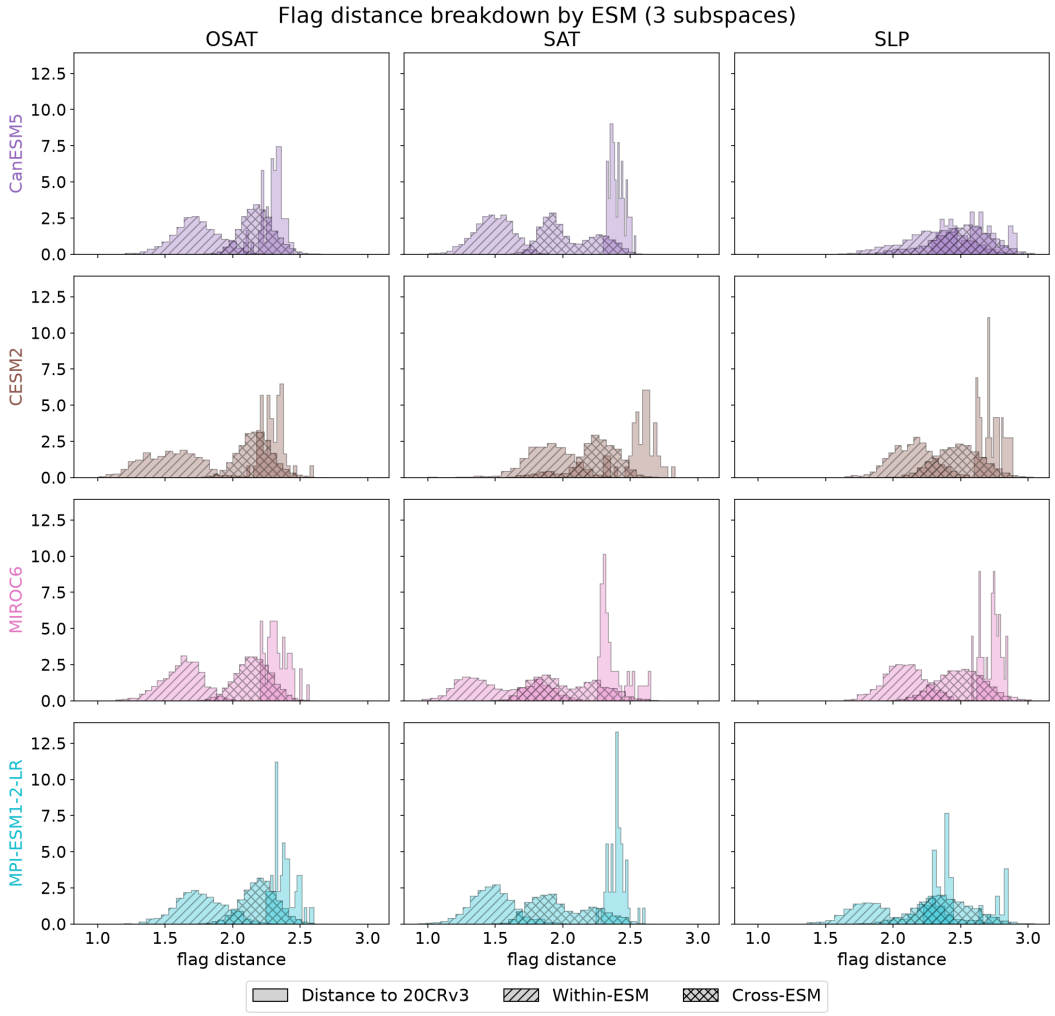


Figure 7: Histograms of flag dissimilarities by variable and ESM, split by origin: to the 20CRv3 reanalysis, within one ESM, and across ESMs.

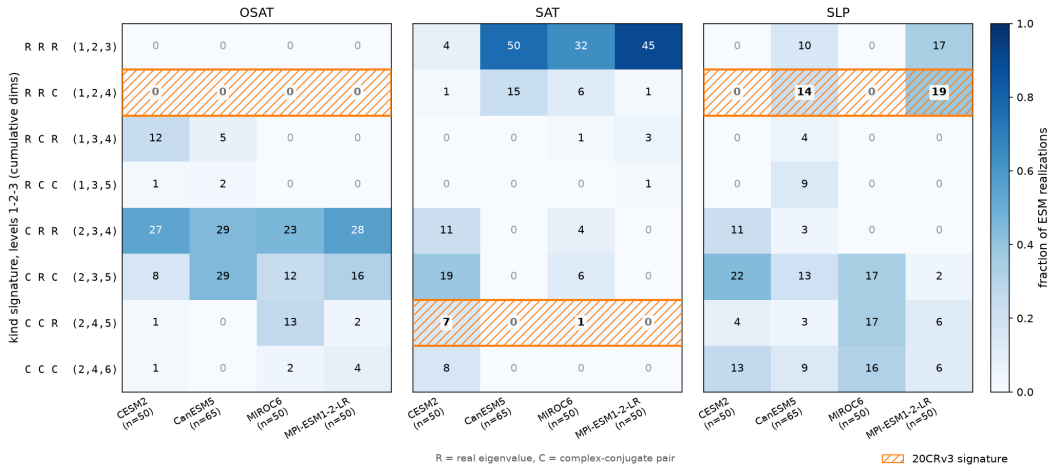


Figure 8: Distribution of flag kind signatures over the first three levels, per ESM and climate variable. Each flag level adds either a real eigenvector (R, one dimension) or the real subspace of a complex-conjugate eigenpair (C, two dimensions; see example 2.1), so a signature fixes the cumulative dimensions of the nested subspaces, shown in parentheses. Each cell gives the number of that ESM's realizations with the signature, and its color gives the corresponding fraction of that ESM's realizations. Hatched orange rows mark the signature of the 20CRv3 reanalysis.