

# Physics-Informed Representation Learning for Emergent Organization in Complex Dynamical Systems

Adam Rupe,<sup>1,2,\*</sup> Karthik Kashinath,<sup>3,4</sup> Nalini Kumar,<sup>5</sup> and James P. Crutchfield<sup>6,†</sup>

<sup>1</sup>*Pacific Northwest National Laboratory, Richland WA, USA.*

<sup>2</sup>*Center For Nonlinear Studies, Los Alamos National Laboratory, Los Alamos NM, USA.*

<sup>3</sup>*NVIDIA Corporation, Santa Clara CA, USA.*

<sup>4</sup>*NERSC, Lawrence Berkeley National Laboratory, Berkeley CA, USA.*

<sup>5</sup>*Intel Corporation, Santa Clara CA, USA.*

<sup>6</sup>*Complexity Sciences Center, Department of Physics and Astronomy,  
University of California Davis, Davis CA 95616, USA.*

(Dated: April 26, 2023)

Nonlinearly interacting system components often introduce instabilities that generate phenomena with new properties and at different space-time scales than the components. This is known as spontaneous self-organization and is ubiquitous in systems far from thermodynamic equilibrium. We introduce a theoretically-grounded framework for emergent organization that, via data-driven algorithms, is constructive in practice. Its building blocks are spacetime lightcones that capture how information propagates across a system through local interactions. We show that predictive equivalence classes of lightcones—local causal states—capture organized behaviors and coherent structures in complex spatiotemporal systems. Using our unsupervised physics-informed machine learning algorithm and a high-performance computing implementation, we demonstrate the applicability of the local causal states for real-world domain science problems. We show that the local causal states capture vortices and their power-law decay behavior in two-dimensional turbulence. We then show that known (hurricanes and atmospheric rivers) and novel extreme weather events can be identified on a pixel-level basis and tracked through time in high-resolution climate data.

Emergent phenomena are often of primary interest in the study of complex systems, but disentangling the web of nonlinear interactions that give rise to them presents enormous challenges to traditional paradigms of scientific inquiry. Although the broad use of numerical models has significantly improved our ability to study nonlinear systems, difficulties remain. The scale and complexity of model outputs can be such that they are now essentially as difficult to understand as the natural phenomena they approximate. That is, the challenge of uncovering the physical and causal mechanisms of emergent phenomena in a large-scale numerical model can be comparable to doing so in the natural system itself.

The Earth system is a quintessential complex system [1, 2], and the emergent behaviors associated with climate change present one of the most pressing issues of our time [3]. Many of the direct impacts of climate change on society are felt through emergent organization in the form of *extreme weather events* [4], such as hurricanes and atmospheric rivers (AR). As the Earth warms, the character of localized extreme weather events (EWE) has and will continue to change [5, 6].

Large-scale climate models are the primary tool used by scientists to study EWEs and climate change more generally [7, 8]. While high-resolution models can produce EWEs under a variety of warming scenarios, it is not clear what underlying mechanisms drive specific shifts in intensity, duration, and spatial dynamics of individual events. Moreover, statistical changes in EWE behaviors in large-scale surveys of various models and warming scenarios [9] are difficult to assess due to the sheer quantity

of data produced. For example, CMIP 6 is estimated to contain 15-30 PB of total climate data [10]. There is an immediate need in the climate community for automated discovery and tracking of EWEs to better understand and predict how they are changing in a warming world and to help mitigate their deleterious effects. Discovering EWEs in a principled, robust, and scalable manner remains an outstanding challenge.

Data-driven and machine learning methods are increasingly called upon as an additional layer of analysis to parse complex phenomena, whether produced by natural systems or complicated numerical models. At first glance, EWE detection and tracking seems an ideal use case for deep learning algorithms that have brought dramatic advances in computer vision. In machine learning terminology, we seek a *segmentation* analysis of extreme weather events—identifying events at the single-pixel or model-grid-cell scale. Indeed, there is considerable effort devoted to extreme weather segmentation using deep learning [11–15].

Deep learning’s main weakness is that neural networks have been most successful at *supervised* computer vision tasks. In this, the networks are trained on *ground truth* labeled examples—data exemplars typically labeled by human experts. As with general fluid vortices [16], no ground truth is available for extreme weather events; see, e.g., [17]. Automated heuristics [18] are used to generate “ground-truth” training examples, but the trained neural networks’ high evaluation scores show that they merely learn to reproduce the output of these heuristics.

Rather than rely on supervised deep learning methods

from computer vision, here we solve the EWE segmentation problem using an unsupervised method derived from the physics of self-organization. Extreme weather events are discovered as specific cases of *coherent structures* [19] that occur ubiquitously in far-from-equilibrium systems [20]. As we demonstrate, unsupervised methods carry the potential for novel structure discovery. While hurricanes, ARs, fronts, and the like are known structures in the atmosphere associated with precipitation extremes, there may be other structures closely linked with extremes that have yet to be identified or pinned down. Moreover, as the climate system changes, the emergent structures and their influence will shift in novel ways that cannot be assumed a priori. These circumstances call for the unsupervised discovery of novelty.

To tackle these challenges head on, we bring together current trends in big data analytics, physics-informed machine learning, and complex systems theory. The *local causal state* method is used to create an extreme weather tracker from an unsupervised spacetime segmentation of high-resolution General Circulation Model (GCM) data. In addition to known events like hurricanes and atmospheric rivers, we show that the local causal states extract novel structures associated with precipitation extremes. Quantitative verification of unsupervised methods is challenging without ground truth for comparison. We address this with a novel verification: the local causal states can identify the nonstationary power-law relaxation of vortex dynamics in two-dimensional free-decay turbulence.

### Local Causal States

We employ an unsupervised structure discovery method using physics-informed geometric representation learning. We stress, though, that it does not involve neural networks. Rather it relies on principles of state-space organization delineated by dynamical systems theory [21].

At present, the canonical representation learning method for spatiotemporal systems is the proper orthogonal decomposition (POD) [22]. In this, a fixed set of spatial modes are found that provide an optimal reconstruction of the spatiotemporal system through linear superposition using time-varying coefficients.

POD modes are the learned representations from which we may extract potential coherent structures. However, restricting to linear modes of full spatial fields is structurally very limiting. Moreover, POD captures dominant energetic modes [23], which generally do not correspond to the organized coherent structures we are interested in. These limitations are one reason Lagrangian methods are favored for coherent structures that dominate material transport in fluid flows [24]. That said, for generality and computational efficiency, we wish to avoid direct reliance

on advection, as needed for most Lagrangian methods.

Our *local causal state* approach leads to a representation learning method similar to POD in that it learns a finite set of “template” representations. However, rather than learning full spatial field templates (modes), the templates of our model, known as *local causal states*, are localized—assigned at each point in space and time. This greatly increases the representational capacity of the finite set of templates, as they can be arranged arbitrarily in space and time.

True to their name, the local causal states are defined using a weak notion of causality. Not only does the future follow from the past, but in spatiotemporal systems that evolve through local interactions, there is a limit on how fast causal influence can propagate. This limit defines *lightcones* in the system that are essential features used in constructing local causal states. The *past lightcone* ( $L^-$ ) of a point in spacetime is the collection of all points at previous times that could possibly have influenced the spacetime point through the local interactions. Similarly, the *future lightcone* ( $L^+$ ) of a spacetime point is the collection of all points at later times that the spacetime point could influence through local interactions.

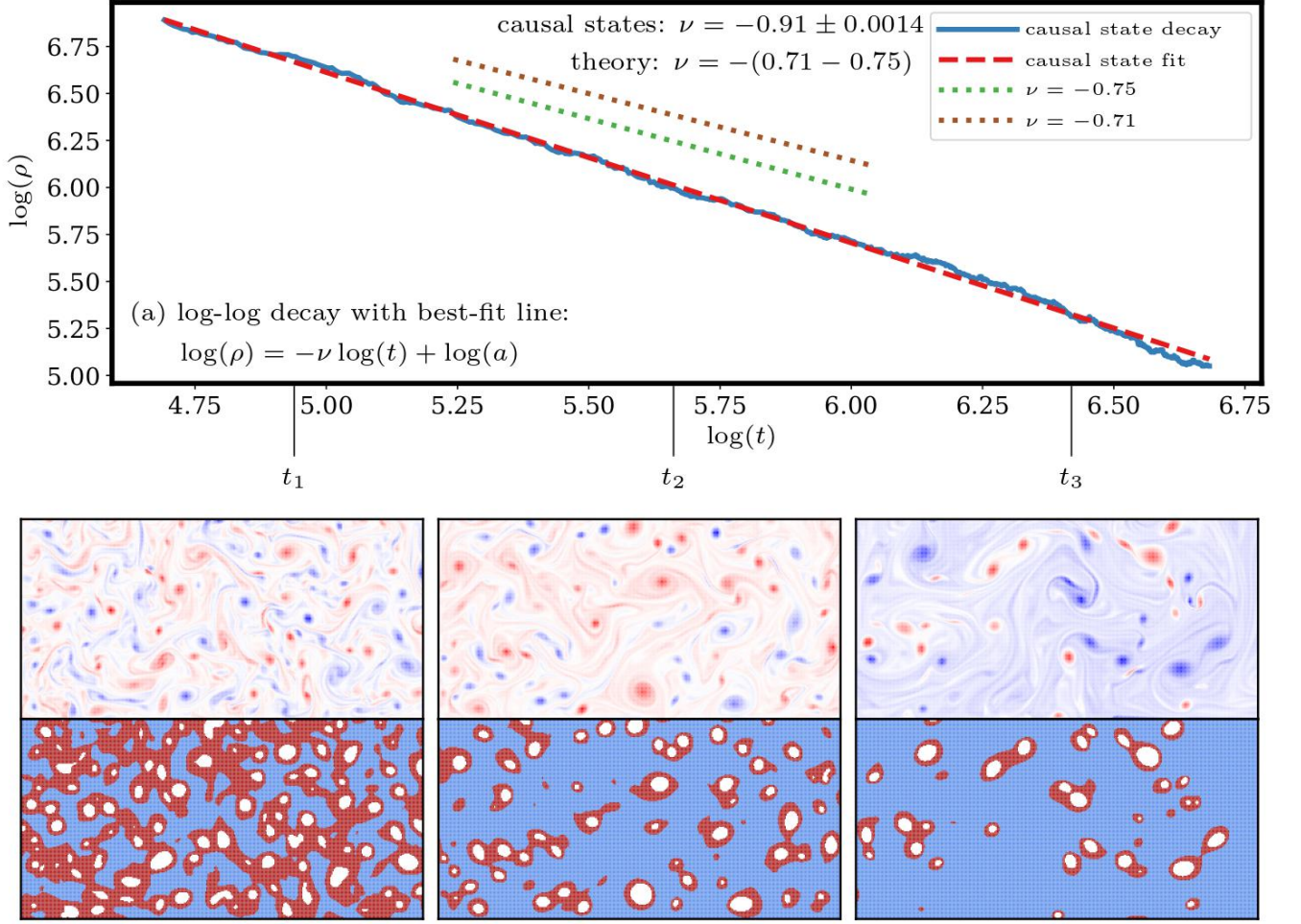
Local causal states are then defined through the *local causal equivalence relation*: two past lightcone configurations, denoted  $\ell^-$ , are considered causally equivalent, denoted  $\sim_\epsilon$ , if they have the same distribution over co-occurring future lightcones (see Figure S1 in the Supplementary Information):

$$\ell_i^- \sim_\epsilon \ell_j^- \iff \Pr(L^+ | L^- = \ell_i^-) = \Pr(L^+ | L^- = \ell_j^-).$$

The local causal states are the equivalence classes of the local causal equivalence relation. A given local causal state is a set of past lightcone configurations that have the same distribution over co-occurring future lightcones [25]. This is expressed functionally by the map  $\epsilon(\cdot)$  from past lightcone configurations to a local causal state:  $\epsilon(\ell_i^-) = \epsilon(\ell_j^-)$  if and only if  $\ell_i^-$  and  $\ell_j^-$  belong to the same local causal state.

Crucially, the  $\epsilon$ -function provides a pointwise mapping from points in spacetime to corresponding local causal states, through their past lightcones. When applied to all points in a spacetime field  $X(\mathbf{r}, t)$  this produces a corresponding *local causal state field*, denoted  $S(\mathbf{r}, t) = \epsilon(X(\mathbf{r}, t))$ . Since  $\epsilon(\ell^-)$  is a pointwise mapping, applying at each point in spacetime, the resulting local causal state field  $S(\mathbf{r}, t)$  shares the same coordinate geometry as its associated spacetime field  $X(\mathbf{r}, t)$ .

This means the local causal states provide a *spacetime segmentation*, as desired. That is, the finite set  $\mathcal{S} = \{S_i\}$  of local causal state templates can be thought of as a set of **class labels**  $S_i$  and each point in spacetime is assigned one of these local causal state labels through the  $\epsilon$ -function. Since the  $\epsilon$ -function is a local mapping defined using lightcones, it is *equivariant* under spacetime translations, rotations, and reflections. This guarantees



(b) Snapshots of vorticity (top) and local causal state (bottom) fields over time;  
 vortex core local causal states in white, transition states in red, and background flow in blue

FIG. 1. Local causal state vortex cores over time and their power law decay.

that the local causal state representations do not depend on spacetime location or orientation of structures they extract.

Data-driven approximation of the  $\epsilon$ -function is achieved using two stages of clustering [26]. The first clustering stage  $\gamma$  is a distance-based partitioning of the space of finite-depth past lightcones, which we implement using the K-Means algorithm [27]. The  $\gamma$ -function maps past lightcones to their associated distance-based cluster. Performing distance-based clustering on past and future lightcones creates sets of cluster labels so that the distributions  $\Pr(\gamma(L^+)|\gamma(\ell^-))$  can be empirically approximated through simple counting. During the second clustering stage  $\psi$ , elements of the  $\gamma$ -partition are clustered together if they have approximately the same empirical distribution  $\Pr(\gamma(L^+)|\gamma(\ell^-))$ . Elements of the  $\gamma$ -partition are mapped to their resulting clusters under the  $\psi$ -function. The approximated  $\epsilon$ -function used for

inference is thus given as:

$$\epsilon(\ell^-) \approx \psi(\gamma(\ell^-)) .$$

Systems under study often have multiple physical fields of interest, such as temperature, pressure, wind speeds, and water vapor. In these multivariate cases, the spacetime fields and lightcones extracted from them are tensors, with each point in spacetime having a vector of values over the multiple physical fields. Multivariate local causal state analysis is performed using a tensor lightcone metric that computes distances for the  $\gamma$ -partition using values from multiple physical fields. See the Supplementary Materials for more details on the local causal states and their data-driven reconstruction.

## Vortex Decay in Two-Dimensional Turbulence

Lacking ground truth, quantitatively evaluating unsupervised coherent structure detection methods is difficult. Typically, one falls back on qualitative visual evaluation and comparison [24]. In contrast, the ability of local causal states to capture nonstationary behavior, including coherent structures with finite lifespans, provides quantitative assessment and so stands in for ground truth.

We demonstrate this using the problem of vortex detection in two-dimensional free-decay turbulence [28]. Like-signed vortices—the coherent structures of interest—undergo pairwise annihilation. This results in a nonstationary decay behavior with a power-law decay rate:

$$\rho(t) = a t^{-\nu},$$

where  $\rho(t)$  is the number of vortices and  $\nu$  is the vortex decay rate. Empirically,  $\nu$  is observed to be  $\nu \sim 0.71 - 0.75$  [28, 29]. Extracting the decay rate provides an opportunity for a quantitative comparison with known physical behavior.

Local causal state segmentation distinguishes between vortices spinning in different directions, but estimating the decay rate requires tracking only the total number of vortices. Therefore, segmentation is performed on the absolute value of vorticity using three class labels, as Figure 1 (b) shows. These labels can be qualitatively interpreted as the background potential flow (blue), vortex cores (white), and transition regions (red) surrounding the vortex cores [28].

Quantitative comparison comes from algorithmically counting the number of vortex cores at each time using a union-find algorithm [30]. A log-log plot of the number of vortex cores over time is shown in Figure 1 (a). Since the decay rate is a power law, the log-log plot is linear, with the decay rate  $\nu$  given as the slope. A best-fit line shows the decay rate of local causal state vortex cores over time fits very well to a power law, with decay rate  $\nu \approx 0.91$ . Slopes representing the accepted theoretical range  $\nu = -(0.71 - 0.75)$  are shown for comparison. As seen in the full spacetime segmentation video [31], the higher-than-expected decay rate is due to a small number of vortex core states eventually being mislabeled as background, and so they vanish without pairwise merging. However, the overall behavior is dominated by the correct physical mechanism of pairwise vortex merging, giving a strong fit to the power law decay.

## Discovering and Tracking Extreme Weather Events

Employing data from the 0.25-degree CAM5.1 Global Circulation Model [32] the following analyzes extreme weather events in global climate data. The goal of our

analysis is twofold: (i) identify known structures, such as hurricanes and ARs, *without pre-specified labels or ad-hoc thresholding*, and (ii) explore as-of-yet undiscovered structures and their relevance for precipitation extremes. The physics of EWEs is incorporated through multivariate segmentation utilizing wind velocities and column-integrated water vapor. This approach is based on the Integrated Vapor Transport (IVT) field, commonly used by climate scientists to study EWEs [33].

Performing a univariate segmentation on the IVT field itself, we find a set of local causal states that visually coincide with hurricanes. That is, local causal states in this **hurricane** set (almost) only occur in spacetime locations where a hurricane is present. Using the shared coordinate geometry, we can then create an unsupervised hurricane tracker by simply highlighting spacetime points where these **hurricane** states occur. A snapshot of this tracker is shown in Figure 2, which depicts the integrated vapor field with spatial cells highlighted red if a **hurricane** state occurs at the same spacetime location in the local causal state field. The full spacetime video of this hurricane tracker can be seen in [34].

Atmospheric rivers are more difficult to isolate with a unique set of **AR** states from a similar IVT field segmentation. However, the more general extreme weather segmentation we now describe does encompass ARs, along with other EWE coherent structures.

In contrast to conforming the local causal state analysis with known EWE structures as above, we now turn to novel EWE structure discovery. A principled definition of coherent structures as local deviations from generalized spacetime symmetries can be expressed in terms of the local causal states [25]. For continuous fields, coherent structures appear against a uniform background given by a single local causal state, as with the blue background state for the turbulent vortices shown in Figure 1 (b). Such a background state is not recovered using a univariate segmentation of the IVT field nor the integrated vapor field; see Reference [26].

A uniform background state is recovered, however, using a multivariate segmentation with the tensor light-cone metric based on IVT, given by Equation (S12) in the Supplementary Information. A snapshot of the local causal state field from this multivariate segmentation is shown in Figure 3. Video of a general EWE tracker, similar to the hurricane tracker above, using these local causal states is seen in [35]. The uniform background local causal state is colored white, with all other colors then corresponding to coherent structures identified by the local causal state definition. Because these states encompass hurricanes and ARs, we thus label the nonbackground local causal states as **EWE** states. We see a unique subset of these states again occur with hurricanes—one in the Pacific and two in the Atlantic. Atmospheric rivers are also seen, predominantly in the southern ocean, and they appear as distinct features disconnected from the

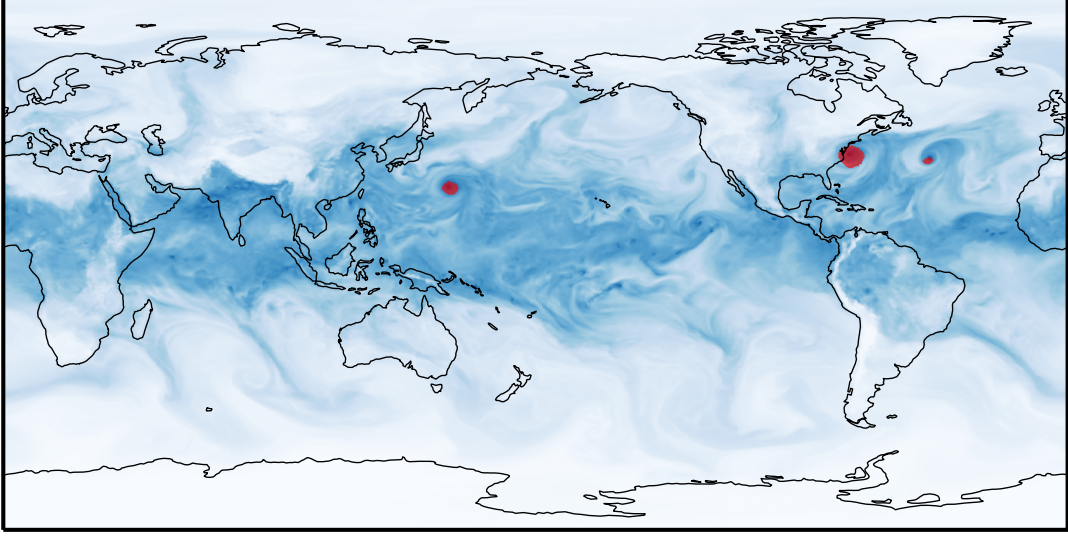


FIG. 2. Hurricane segmentation masks created from local causal state segmentation of the IVT field.

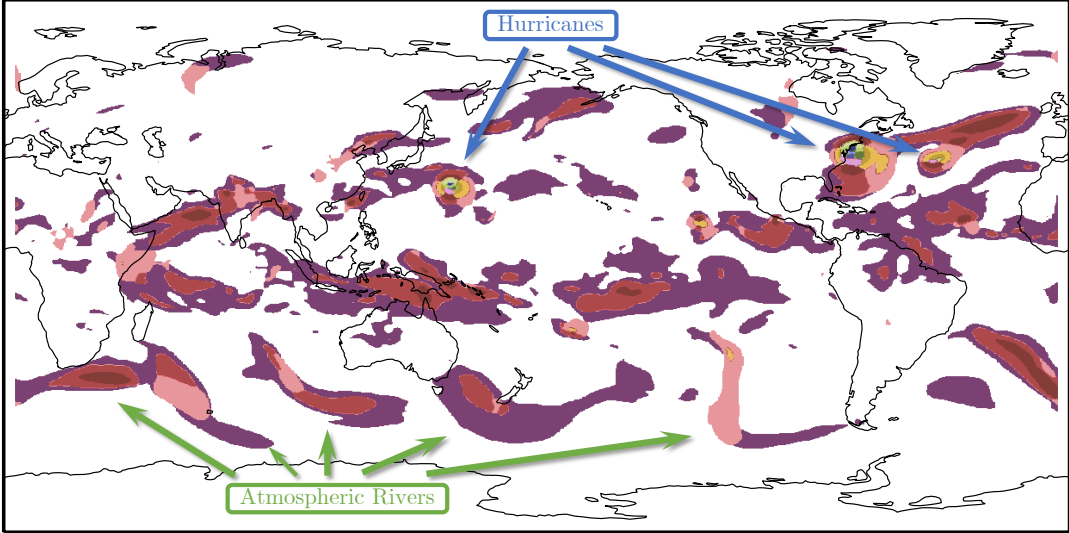


FIG. 3. Extreme weather event segmentation analysis using multivariate lightcone tensor metric.

tropics.

While these known structures appear in this general coherent structure segmentation, there are many additional signatures that show up that do not correspond to known structures. To test their significance, we investigate the co-occurrence of precipitation extremes [36] with these EWE local causal states. That is, how many precipitation extremes occur in a spacetime location for which that same location in the local causal state field is an EWE state, rather than a **background** state?

Despite occupying only 19.6% of spacetime volume, the EWE states capture 45% of extremes in 90<sup>th</sup> percentile

globally, 58% of extremes in the 99<sup>th</sup> percentile, and 75% in the 99.9<sup>th</sup>. Therefore, these newly discovered structures can be identified as extreme weather events, as they capture the majority of extreme precipitation events despite being created without employing the precipitation field. For some level of comparison, Reference [36] finds 51% of global extremes (in the 99<sup>th</sup> percentile) co-occur with fronts.



## Conclusions

Advances in dynamical systems and applied ergodic theory are forming the foundations of a new paradigm for data-driven science [37]. For example, it is now clear that data-driven and physics-based predictive models are intimately related [38]. Beyond predictive modeling, uncovering the underlying mechanisms behind emergent behaviors is a crucial component of scientific inquiry and remains an outstanding challenge. Physics-informed data-driven methods, like those introduced here, provide a promising new avenue in this [39, 40].

Motivated by local causality and the physics of organization encapsulated by causal equivalence, the local causal states provide a physics-informed geometric representation learning method for capturing emergent organization in complex spatiotemporal systems. Moreover, our approximation algorithm and high-performance computing implementation make the local causal states actionable in practice on large-scale cutting-edge domain science problems. Here, we produced a novel quantitative verification of coherent structure detection by automatically counting vortices in free-decay two-dimensional turbulence and recovering the known power-law decay over time using the local causal states and their ability to capture nonstationary behavior. We then employed the local causal states for automated, *fully-unsupervised* extreme weather discovery and demonstrated their ability to extract both known and novel structures in global climate data. With further refinement, the local causal states may be able to additionally provide a scale for strength and intensity of identified extreme weather events [41].

---

\* adam.rupe@pnnl.gov

† crutchfield@ucdavis.edu

- [1] M. Ghil and V. Lucarini, The physics of climate variability and climate change, *Rev. of Mod. Phys.* **92**, 035002 (2020).
- [2] A. R. Ravishankara, D. A. Randall, and J. W. Hurrell, Complex and yet predictable: The message of the 2021 Nobel Prize in physics, *Proc. Natl. Acad. Sci. USA* **119**, e2120669119 (2022).
- [3] H.-O. Pörtner, D. C. Roberts, H. Adams, C. Adler, P. Aldunce, E. Ali, R. A. Begum, R. Betts, R. B. Kerr, R. Biesbroek, *et al.*, *Climate change 2022: Impacts, adaptation and vulnerability* (IPCC Geneva, Switzerland, 2022).
- [4] M. K. Tippett, Extreme weather and climate, *npj Climate and Atmospheric Science* **1**, 45 (2018).
- [5] P. J. Webster, G. J. Holland, J. A. Curry, and H.-R. Chang, Changes in tropical cyclone number, duration, and intensity in a warming environment, *Science* **309**, 1844 (2005).
- [6] A. Robinson, J. Lehmann, D. Barriopedro, S. Rahmstorf, and D. Coumou, Increasing heat and rainfall extremes now far outside the historical climate, *npj Climate and Atmospheric Science* **4**, 45 (2021).
- [7] P. N. Edwards, *A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming* (The MIT Press, 2010).
- [8] G. Flato, J. Marotzke, B. Abiodun, P. Braconnot, S. Chou, W. Collins, P. Cox, F. Driouech, S. Emori, V. Eyring, C. Forest, P. Gleckler, E. Guilyardi, C. Jakob, V. Kattsov, C. Reason, and M. Rummukainen, Evaluation of climate models, in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by T. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. Midgley (Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013).
- [9] V. Eyring, S. Bony, G. A. Meehl, C. A. Senior, B. Stevens, R. J. Stouffer, and K. E. Taylor, Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, *Geoscientific Model Development* **9**, 1937 (2016).
- [10] M. Stockhause and M. Lautenschlager, CMIP6 data citation of evolving data, *Data Science Journal* **16** (2017).
- [11] M. Mudigonda, S. Kim, A. Mahesh, S. Kahou, K. Kashinath, D. Williams, V. Michalski, T. O'Brien, and Prabhat, Segmenting and tracking extreme climate events using neural networks, in *DLPS Workshop, NeurIPS* (2017).
- [12] C. Jiang, J. Huang, K. Kashinath, Prabhat, P. Marcus, and M. Niessner, Spherical CNNs on unstructured grids, in *International Conference on Learning Representations* (2019).
- [13] T. Cohen, M. Weiler, B. Kicanaoglu, and M. Welling, Gauge equivariant convolutional networks and the icosahedral CNN, in *Proceedings of the 36th International Conference on Machine Learning*, Proceedings of Machine Learning Research, Vol. 97, edited by K. Chaudhuri and R. Salakhutdinov (PMLR, Long Beach, California, USA, 2019) pp. 1321–1330.
- [14] T. Kurth, S. Treichler, J. Romero, M. Mudigonda, N. Luehr, E. Phillips, A. Mahesh, M. Matheson, J. Deslippe, M. Fatica, Prabhat, and M. Houston, Exascale deep learning for climate analytics, in *Proceedings of the International Conference for High Performance Computing, Networking, Storage, and Analysis* (IEEE Press, 2018) p. 51.
- [15] Prabhat, K. Kashinath, M. Mudigonda, S. Kim, L. Kapp-Schwoerer, A. Graubner, E. Karaismailoglu, L. Von Kleist, T. Kurth, A. Greiner, A. Mahesh, *et al.*, ClimateNet: an expert-labeled open dataset and deep learning architecture for enabling high-precision analyses of extreme weather, *Geoscientific Model Development* **14**, 107 (2021).
- [16] B. Epps, Review of vortex identification methods, in *55th AIAA Aerospace Sciences Meeting* (2017) p. 0989.
- [17] C. A. Shields, J. J. Rutz, L.-Y. Leung, F. M. Ralph, M. Wehner, B. Kawzenuk, J. M. Lora, E. McClenny, T. Osborne, A. E. Payne, P. Ullrich, A. Gershunov, N. Goldenson, B. Guan, Y. Qian, A. M. Ramos, C. Sarangi, S. Sellars, I. Gorodetskaya, K. Kashinath, V. Kurlin, K. Mahoney, G. Muszynski, R. Pierce, A. C. Subramanian, R. Tome, D. Waliser, D. Walton, G. Wick, A. Wilson, D. Lavers, Prabhat, A. Collow, H. Krishnan, G. Magnusdottir, and P. Nguyen, Atmospheric river tracking method intercomparison project (ART-

- MIP): project goals and experimental design, *Geoscientific Model Development* **11**, 2455 (2018).
- [18] Prabhat, O. Rübel, S. Byna, K. Wu, F. Li, M. Wehner, W. Bethel, *et al.*, TECA: A parallel toolkit for extreme climate analysis, in *Third Workshop on Data Mining in Earth System Science (DMESS)* (2012).
  - [19] G. Haller, Lagrangian coherent structures, *Ann. Rev. Fluid Mech.* **47**, 137 (2015).
  - [20] M. C. Cross and P. C. Hohenberg, Pattern formation outside of equilibrium, *Rev. Mod. Phys.* **65**, 851 (1993).
  - [21] S. Wiggins, *Introduction to Applied Nonlinear Dynamical Systems and Chaos* (Springer-Verlag, New York, 1990).
  - [22] P. Holmes, J. Lumley, G. Berkooz, and C. Rowley, *Turbulence, Coherent Structures, Dynamical Systems and Symmetry* (Cambridge University Press, Cambridge, United Kingdom, 2012).
  - [23] K. Taira, S. L. Brunton, S. T. Dawson, C. W. Rowley, T. Colonius, B. J. McKeon, O. T. Schmidt, S. Gordeyev, V. Theofilis, and L. S. Ukeiley, Modal analysis of fluid flows: An overview, *AIAA Journal* **55**, 4013 (2017).
  - [24] A. Hadjighasem, M. Farazmand, D. Blazeviski, G. Froyland, and G. Haller, A critical comparison of Lagrangian methods for coherent structure detection, *Chaos* **27**, 053104 (2017).
  - [25] A. Rupe and J. P. Crutchfield, Local causal states and discrete coherent structures, *Chaos* **28**, 1 (2018).
  - [26] A. Rupe, N. Kumar, V. Epifanov, K. Kashinath, O. Pavlyk, F. Schlimbach, M. Patwary, S. Maidanov, V. Lee, Prabhat, and J. P. Crutchfield, Disco: Physics-based unsupervised discovery of coherent structures in spatiotemporal systems, in *2019 IEEE/ACM Workshop on Machine Learning in High Performance Computing Environments (MLHPC)* (IEEE, 2019) pp. 75–87.
  - [27] D. Arthur and S. Vassilvitskii, K-Means++: the advantages of careful seeding, in *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms* (Society for Industrial and Applied Mathematics, 2007) pp. 1027–1035.
  - [28] J. C. McWilliams, The vortices of two-dimensional turbulence, *J. of Fluid Mech.* **219**, 361 (1990).
  - [29] G. F. Carnevale, J. C. McWilliams, Y. Pomeau, J. B. Weiss, and W. R. Young, Evolution of vortex statistics in two-dimensional turbulence, *Phys. Rev. Lett.* **66**, 2735 (1991).
  - [30] C. Fiorio and J. Gustedt, Two linear time union-find strategies for image processing, *Theoretical Computer Science* **154**, 165 (1996).
  - [31] A. Rupe, 2D turbulence segmentation video, [https://drive.google.com/file/d/1RoAh6J\\_gadhDcgWFPXa5z6g9q55TDM7G/view](https://drive.google.com/file/d/1RoAh6J_gadhDcgWFPXa5z6g9q55TDM7G/view) (2022).
  - [32] M. F. Wehner, K. Reed, F. Li, Prabhat, J. Bacmeister, C.-T. Chen, C. Paciorek, P. Gleckler, K. Sperber, W. D. Collins, A. Gettelman, and C. Jablonowski, The effect of horizontal resolution on simulation quality in the community atmospheric model, CAM5.1., *J. of Modeling the Earth System* **06**, 980 (2014).
  - [33] P. M. Sousa, A. M. Ramos, C. C. Raible, M. Messmer, R. Tomé, J. G. Pinto, and R. M. Trigo, North Atlantic integrated water vapor transport—from 850 to 2100 CE: Impacts on western European rainfall, *J. of Climate* **33**, 263 (2020).
  - [34] A. Rupe, Hurricane tracker video, [https://drive.google.com/file/d/11WxW0hNVL3eT1VHg0TcCBMdJQrbzrPLK/view?usp=share\\_link](https://drive.google.com/file/d/11WxW0hNVL3eT1VHg0TcCBMdJQrbzrPLK/view?usp=share_link) (2022).
  - [35] A. Rupe, General EWE tracker video, [https://drive.google.com/file/d/1mFnmHHLxK34IEVvMJA5axPNitVohMy9\\_/view?usp=share\\_link](https://drive.google.com/file/d/1mFnmHHLxK34IEVvMJA5axPNitVohMy9_/view?usp=share_link) (2022).
  - [36] J. L. Catto and S. Pfahl, The importance of fronts for extreme precipitation, *J. Geophys. Res.: Atmospheres* **118**, 10,791 (2013).
  - [37] T. Berry, D. Giannakis, and J. Harlim, Bridging data science and dynamical systems theory, *Notices of the American Mathematical Society* **67**, 1336 (2020).
  - [38] A. Rupe, V. V. Vesselinov, and J. P. Crutchfield, Nonequilibrium statistical mechanics and optimal prediction of partially-observed complex systems, *New Journal of Phys.* **24**, 103033 (2022).
  - [39] J. Runge, V. Petoukhov, J. F. Donges, J. Hlinka, N. Jajcay, M. Vejmelka, D. Hartman, N. Marwan, M. Paluš, and J. Kurths, Identifying causal gateways and mediators in complex spatio-temporal systems, *Nature Comms.* **6**, 1 (2015).
  - [40] S. Klus, B. E. Husic, M. Mollenhauer, and F. Noé, Kernel methods for detecting coherent structures in dynamical data, *Chaos* **29**, 123112 (2019).
  - [41] F. M. Ralph, J. J. Rutz, J. M. Cordeira, M. Dettinger, M. Anderson, D. Reynolds, L. J. Schick, and C. Smallcomb, A scale to characterize the strength and impacts of atmospheric rivers, *Bulletin of the American Meteorological Society* **100**, 269 (2019).
  - [42] E. R. Weeks, J. S. Urbach, and H. L. Swinney, Anomalous diffusion in asymmetric random walks with a quasi-geostrophic flow example, *Physica D* **97**, 291 (1996).
  - [43] V. Algazi and D. Sakrison, On the optimality of the Karhunen-Loève expansion, *IEEE Transactions on Information Theory* **15**, 319 (1969).
  - [44] C. W. Rowley, T. Colonius, and R. M. Murray, Model reduction for compressible flows using POD and Galerkin projection, *Physica D* **189**, 115 (2004).
  - [45] T. Peacock and G. Haller, Lagrangian coherent structures: The hidden skeleton of fluid flows, *Physics Today* **66**, 41 (2013).
  - [46] J. H. Tu, C. W. Rowley, D. M. Luchtenburg, S. L. Brunton, and J. N. Kutz, On dynamic mode decomposition: Theory and applications, *J. Comp. Dyn.* **1**, 391 (2014).
  - [47] C. Shalizi, Optimal nonlinear prediction of random fields on networks, *DMTCS Proceedings* **vol. AB**, 11 (2003).
  - [48] S. Caires and J. A. Ferreira, On the non-parametric prediction of conditionally stationary sequences, *Statistical Inference for Stochastic Processes* **8**, 151 (2005).
  - [49] M. M. Bronstein, J. Bruna, Y. LeCun, A. Szlam, and P. Vandergheynst, Geometric deep learning: going beyond Euclidean data, *IEEE Sig. Proc. Mag.* **34**, 18 (2017).
  - [50] M. M. Bronstein, J. Bruna, T. Cohen, and P. Veličković, Geometric deep learning: Grids, groups, graphs, geodesics, and gauges, *arXiv preprint arXiv:2104.13478* (2021).
  - [51] G. Goerg and C. Shalizi, LICORS: Light cone reconstruction of states for non-parametric forecasting of spatio-temporal systems, *arXiv:1206.2398*.
  - [52] H. Jänicke, A. Wiebel, G. Scheuermann, and W. Kollmann, Multifield visualization using local statistical complexity, *IEEE Trans. Vis. Comp. Graphics* **13**, 1384 (2007).
  - [53] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, *et al.*, A

- density-based algorithm for discovering clusters in large spatial databases with noise., in *KDD*, Vol. 96 (1996) pp. 226–231.
- [54] R. Balestrierio and R. G. Baraniuk, Mad Max: Affine spline insights into deep learning, *Proceedings of the IEEE* **109**, 704 (2021).
  - [55] N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw, Geometry from a time series, *Phys. Rev. Let.* **45**, 712 (1980).
  - [56] A. Rupe and J. P. Crutchfield, Algebraic theory of patterns as generalized symmetries, *Symmetry* **14**, 1636 (2022).
  - [57] H. Bénard, *Les Tourbillons Cellulaires dans une nappe Liquide Propageant de la Chaleur par Convection: en Régime Permanent* (Gauthier-Villars, 1901).
  - [58] L. Rayleigh, On convection currents in a horizontal layer of fluid, when the higher temperature is on the under side, *Phil. Mag. (Series 6)* **32**, 529 (1916).
  - [59] S. Chandrasekhar, *Hydrodynamic and Hydromagnetic Stability* (Oxford, Clarendon Press, 1968).
  - [60] F. H. Busse, Non-linear properties of thermal convection, *Reports on Progress in Physics* **41**, 1929 (1978).
  - [61] P. Fenstermacher, H. Swinney, and J. Gollub, Dynamical instabilities and the transition to chaotic Taylor vortex flow, *J. Fluid Mech.* **94**, 103 (1979).
  - [62] V. Steinberg, G. Ahlers, and D. S. Cannell, Pattern formation and wave-number selection by Rayleigh-Bénard convection in a cylindrical container, *Physica Scripta* **T9**, 97 (1985).
  - [63] W. Heisenberg, Nonlinear problems in physics, *Physics Today* **20**, 23 (1967).
  - [64] J. T. C. Liu, Contributions to the understanding of large-scale coherent structures in developing free turbulent shear flows, *Advances in applied mechanics* **26**, 183 (1988).
  - [65] H. W. Liepmann, Aspects of the turbulence problem, *Zeitschrift für angewandte Mathematik und Physik ZAMP* **3**, 407 (1952).
  - [66] A. A. Townsend, *The structure of turbulent shear flow* (Cambridge university press, 1956).
  - [67] J. L. Lumley, The structure of inhomogeneous turbulence, *Atmospheric turbulence and radio wave propagation*, 166 (1967).
  - [68] J. C. McWilliams, On the relevance of two-dimensional turbulence to geophysical fluid motions, *Journal de Mecanique Theorique et Appliquee Supplement*, 83 (1983).
  - [69] W. S. Parker, Reanalyses and observations: What’s the difference?, *Bulletin of the American Meteorological Society* **97**, 1565 (2016).
  - [70] K. Dagon, J. Truesdale, J. C. Biard, K. E. Kunkel, G. A. Meehl, and M. J. Molina, Machine learning-based detection of weather fronts and associated extreme precipitation in historical and future climates, *Journal of Geophysical Research: Atmospheres* **127**, e2022JD037038 (2022).
  - [71] NASA, Jupiter cloud sequence from cassini, <https://svs.gsfc.nasa.gov/cgi-bin/details.cgi?aid=3610> (Accessed: 2019-03-14).
  - [72] A. Hadjighasem and G. Haller, Geodesic transport barriers in Jupiter’s atmosphere: Video-based analysis, *SIAM Review* **58**, 69 (2016).



# Supplementary Information

## Methodology

### *Spatiotemporal Systems*

Local causal states are learned representations that extract emergent organization from spatially-extended dynamical systems. The *behaviors* of spatiotemporal systems are *spacetime fields*, elements of the extended phase space that includes time as a dimension. Mathematically, a spacetime field is a tensor  $X(\mathbf{r}, t) : \mathbb{S} \rightarrow \mathbb{R}^n$ , where  $\mathbb{S}$  is some spacetime coordinate system and  $n$  is the number of physical field quantities—e.g., temperature, pressure, velocity components, and so on. Each coordinate in spacetime maps to a length- $n$  vector of the values of the physical quantities at that point.

The systems presently analyzed live in a simple Euclidean spacetime, so that  $\mathbb{S} = \mathbb{R}^{d+1}$ , where  $d$  is the number of spatial dimensions. Note though that the spacetime data from the climate reanalysis and Jupiter’s clouds are Euclidean projections from a spherical spatial coordinate system.

Numerical data, with which any data-driven method must necessarily work, will be given on some finite grid  $\mathbb{G}$  of dimension  $(X, Y, T)$  for two spatial dimensions. The data used here is on a simple integer Euclidean grid (as opposed to a complex adaptive mesh, or spherical grid), so that for each field quantity,  $X(\mathbf{r}, t) : \mathbb{G} \rightarrow \mathbb{R}$ , where  $\mathbb{G} = \{1, \dots, X\} \times \{1, \dots, Y\} \times \{1, \dots, T\}$ . As standard, we denote this as  $X(\mathbf{r}, t) \in \mathbb{R}^{X \times Y \times T}$ .

Loosely speaking, a spatiotemporal system possesses some level of *organization* if many of its degrees of freedom evolve collectively. Coherent structures—the form of organization we are most interested in—implies an emergent higher-level degree of freedom in the system. Particles in a fluid vortex, for instance, will largely travel together through a complex fluid flow. Thus, the particles need not be tracked individually; one can simply track the single vortex [42].

### *Related Coherent Structure Methods*

This motivates the idea of physics-informed representation learning to extract coherent structures. If a spacetime field  $X(\mathbf{r}, t) \in \mathbb{R}^{X \times Y \times T}$  is compressed through some encoding, then the encoded representations may leverage the collective organization present in the system. Particularly when representations are learned through an encoder-decoder framework, the collective higher-order degrees of freedom—coherent structures—capture the majority of a system’s behavior and so provide a natural set of compressed representations.

The Proper Orthogonal Decomposition (POD) provides a fixed set of  $N$  spatial modes  $\{\phi_i(\mathbf{r})\}$  through an eigen-decomposition of a data covariance matrix. The modes provide an optimal reconstruction [43] of the spatiotemporal system  $X(\mathbf{r}, t)$  via linear superposition:

$$X(\mathbf{r}, t) \approx \sum_{i=1}^N a_i(t) \phi_i(\mathbf{r}) .$$

where  $a_i(t)$  are time-dependent coefficients. When applied to fluid flows, it has been shown that the POD modes capture the dominant energetic components of the flow [44]. POD then has become a standard choice for representation learning in fluid dynamics, as POD modes capture “coherent structures” in the energy landscape [22].

Dynamically-coherent structures, collective higher-order degrees of freedom that emerge in complex fluid flows, are not necessarily dominant contributors to system energetics. Rather, they provide a “hidden skeleton” that dominates material transport in the flow [45]. Thus, Lagrangian methods have become popular for identifying coherent structures in complex fluid flows. Inspired by the geometry of various types of invariant sets in dynamical systems theory, Lagrangian Coherent Structures methods [19, 24, 45] seek the strongest repelling, attracting, or shearing material surfaces in the flow over a finite time interval [19].

Tools from dynamical systems theory, particularly linear stability analysis, can be applied through the use of the Lagrangian flow map [19, Eq. (3)]:

$$\mathbf{y}(t; t_0, \mathbf{y}_0) := F_{t_0}^t(\mathbf{y}_0) , \tag{S1}$$

where  $\mathbf{y}(t)$  represents the spatial position of a fluid particle or tracer at time  $t$ . Candidate coherent structures are given as a material surface  $\mathcal{M}(t)$ —a smooth surface of fluid particles. A material surface is deformed by the flow

through the Lagrangian advection of all the points in  $\mathcal{M}(t)$ , which is denoted simply as [19, Eq. (5)]:

$$\mathcal{M}(t) := F_{t_0}^t(\mathcal{M}(t_0)) . \quad (\text{S2})$$

While conceptually appealing and qualitatively quite successful, the Lagrangian Coherent Structure approach poses difficulties for the more complicated applications considered here. First, most Lagrangian Coherent Structure methods capable of coherent structure *segmentation* (as opposed to simply producing a diagnostic scalar field, like Finite-Time Lyapunov Exponents [24]) require costly computation of the Cauchy-Green or Green-Lagrange strain tensors, which in turn require having the flow map (S2) in hand, to analyze the effect of material surfaces on nearby flow particles. This of course also excludes the incorporation of non-advection quantities for Lagrangian methods, like temperature or pressure fields.

Moreover, the Lagrangian approaches are typically applied to structures that exist in a finite time interval. Once a structure is identified through a given criteria, its time evolution is given simply through Lagrangian advection, as in Eq. (S2). For extreme weather segmentation, we want to identify all EWEs present in a given climate run. This involves the “birth” and “death” of individual events over the course of the run. Once an individual hurricane, say, is identified, the advection of that hurricane given by Eq. (S2) is applied to all points identified as a hurricane by Eq. (S1), which is well-defined at all times. That is, points that evolve coherently as a hurricane during its lifetime will scatter throughout the flow after they cease to be a hurricane. Thus, in the Lagrangian Coherent Structure framework, the status of a material surface as an EWE must be evaluated through a costly computation at all points in time.

### Local Causal States

Our local causal state approach is more closely related to general encoder-decoder representation learning methods like POD, rather than the more specific approaches of Lagrangian Coherent Structures. That said, we show shortly that the local causal states can be viewed as a Lagrangian method when applied to fluid flows. However, it does not rely on computing specific forms of material surface deformation, and it does not require the Lagrangian flow map.

A key distinction between linear modal decompositions [23]—such as POD or the related Dynamic Mode Decomposition [46]—and the local causal states is that modal decompositions provide a finite set of *spatial field templates* (modes). Each spatial mode  $\phi(\mathbf{r})$  provides a fixed template with the geometry of a spatial field  $X(\mathbf{r})$ . The local causal states, by contrast, provide a finite set of *localized templates* that are assigned at each individual point in spacetime.

The locality of local causal states is achieved through the use of *lightcones* in the spatiotemporal system. For systems that evolve according to local interactions, lightcones delineate the causal influence of a point in spacetime. Specifically, the *past lightcone*  $L^-(\mathbf{r}, t)$  of a spacetime point  $(\mathbf{r}, t) \in \mathbb{S}$  is the set of all spacetime points in the past of  $(\mathbf{r}, t)$  that could possibly influence  $(\mathbf{r}, t)$  through propagation of the local interactions:

$$L^-(\mathbf{r}, t) := \{(\mathbf{r}', t') : t' \leq t, \|\mathbf{r}' - \mathbf{r}\| \leq c(t' - t)\}, \quad (\text{S3})$$

where  $c$  is the speed at which local interactions propagate through the system. Similarly, the *future lightcone*  $L^+(\mathbf{r}, t)$  of a spacetime point  $(\mathbf{r}, t)$  is the set of all spacetime points at later times that  $(\mathbf{r}, t)$  itself can possibly influence through the local interactions, so that:

$$L^+(\mathbf{r}, t) := \{(\mathbf{r}', t') : t' > t, \|\mathbf{r}' - \mathbf{r}\| \leq c(t' - t)\}. \quad (\text{S4})$$

We include the present spacetime point  $(\mathbf{r}, t)$  as a member of its past lightcone, but not its future lightcone. See Figure S1.

We denote past and future lightcone random variables as  $L^-$  and  $L^+$  respectively, with realizations denoted as  $\ell^-$  and  $\ell^+$ . Lightcone realizations  $\ell^\pm(\mathbf{r}, t)$ , or *configurations*, are given by assignment of values from a particular spacetime field  $X(\mathbf{r}, t) \in \mathbb{R}^{X \times Y \times T}$  to spacetime points  $(\mathbf{r}, t)$  in  $L^\pm(\mathbf{r}, t)$ .

For systems defined in a discrete spacetime, such as cellular automata [25], the speed of light  $c$  is simply given by the radius of local interactions. For systems in a continuous spacetime, like fluid flows,  $c$  is given by the speed of sound in the system. As described above in the context of Lagrangian Coherent Structures, we are interested in organization that emerges at advective scales in the flow, which are typically much lower velocities than the speed of sound. In this case, we consider lightcones in the scope of Lagrangian advection, defined in Eq. (S1).

Consider a fluid particle  $\mathbf{y}$  initially at spacetime point  $(\mathbf{r}_0, t_0)$ . The trajectory of the particle  $\mathbf{y}(t)$ , given by the flow map in Eq. (S1), is the solution of the differential equation:

$$\dot{\mathbf{y}} = \mathbf{v}(\mathbf{r}, t) , \quad (\text{S5})$$

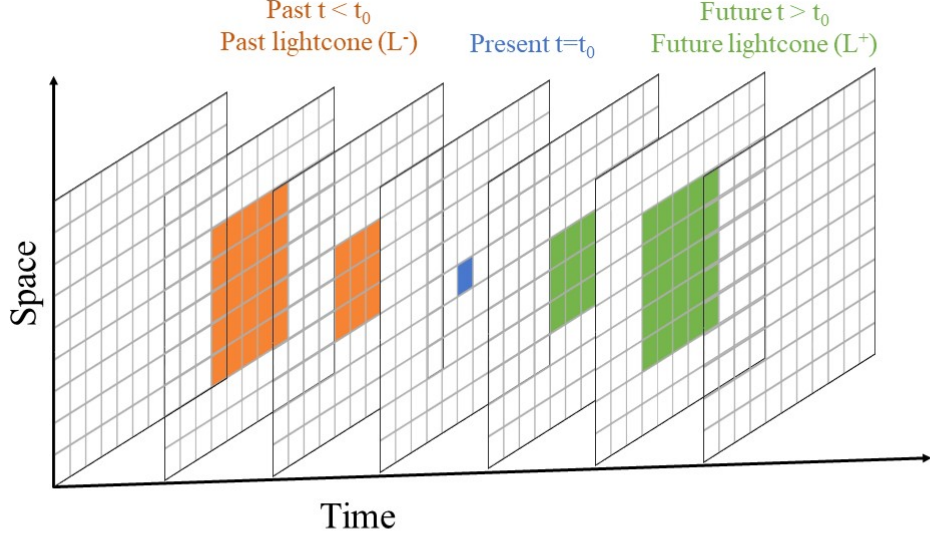


FIG. S1. Co-occurring past and future lightcones in discrete 2 + 1 dimensional spacetime with  $c = 1$ .

where  $v(\mathbf{r}, t)$  is a smooth velocity field. While the velocity field varies in space and time, we can assign the maximum velocity value  $v_{\max}$  so that  $v(\mathbf{r}, t) \leq v_{\max}$ .

The *Lagrangian lightcones* then are defined by setting  $c = v_{\max}$  in Eqs. (S3) and (S4). Again,  $v_{\max}$  is typically much smaller than the speed of sound. Lagrangian lightcones capture local causality derived from Lagrangian advection. A Lagrangian past lightcone delineates the possible reach of a particle  $\mathbf{y}(t; t_0, \mathbf{y}_0)$  initially at point  $(\mathbf{r}_0, t_0)$  through Lagrangian advection in reverse time. That is, all the spacetime points at earlier times from where the particle could possibly have come when evolving according to Lagrangian advection in Eq. (S1). In practice, the value of  $c$  used can be seen as a hyperparameter that controls the spatial scale of organization captured by the approximated local causal states. A larger  $c$  captures more coarse-grained structures.

In addition to delineating causal influence for points in spacetime, note that lightcones are defined solely in terms of distances in spacetime. This implies that (i) they are *equivariant* under spacetime isometries, such as translations and rotations, and (ii) they are well-defined for any spatial geometry with a distance metric, such as the surface of a sphere or an arbitrary spatial network [47]. Thus, they can be used in a wide variety of spatiotemporal systems, and they transform appropriately under translations, rotations, and reflections. We will return to this later point shortly. Note that planar projections of spherical data, like those we use here for climate data, can break global rotational symmetry, and therefore will not be robust across different projections. However, the local causal states appear to be robust to local rotations in a given projection. For example, the local causal state signature of hurricanes does not change as the hurricane rotates locally.

Having defined lightcones, we can now define local causal states through the *local causal equivalence relation*. Two past lightcone configurations are considered causally equivalent if they have the same conditional distribution over co-occurring future lightcones:

$$\ell_i^- \sim_{\epsilon} \ell_j^- \iff \Pr(\mathbf{L}^+ | \mathbf{L}^- = \ell_i^-) = \Pr(\mathbf{L}^+ | \mathbf{L}^- = \ell_j^-) . \quad (\text{S6})$$

The equivalence classes of the local causal equivalence relation Eq. (S6) are the *local causal states*. Individual local causal states are denoted  $\xi$  and the set of all local causal states for a given system is  $\Xi$ .

A local causal state is a set of past lightcone configurations that all have the same conditional distribution  $\Pr(\mathbf{L}^+ | \mathbf{L}^- = \ell^-)$ . It can also simply be thought of as the conditional distribution itself, since all past lightcone configurations in a local causal state by definition share that one distribution.

Here, we consider deterministic systems whose dynamics do not change over time; e.g., a fluid flow governed by the Navier-Stokes equations with time-independent parameters. Therefore, the conditional distributions over lightcones  $\Pr(\mathbf{L}^+ | \mathbf{L}^- = \ell^-)$  are also time-independent—a condition known as *conditional stationarity* [48]. For such systems, each past lightcone configuration  $\ell_i^-$  has a unique and well-defined distribution  $\Pr(\mathbf{L}^+ | \mathbf{L}^- = \ell_i^-)$  and, thus, is also associated with a unique local causal state defined by that distribution. Therefore, local causal states do not require the system

dynamics to be stationary to be well-defined. This allows us to apply local causal states to conditionally-stationary behaviors, like vortex decay in two-dimensional turbulence.

We define the  $\epsilon$ -function as the mapping from past lightcone configurations to their corresponding local causal state  $\epsilon : \ell^- \mapsto \xi$ . The functional form of local causal equivalence relation is given in terms of the  $\epsilon$ -function as:

$$\ell_i^- \sim_\epsilon \ell_j^- \iff \epsilon(\ell_i^-) = \epsilon(\ell_j^-) . \quad (\text{S7})$$

For a given spacetime field  $X(\mathbf{r}, t)$ , each spacetime point  $(\mathbf{r}, t) \in \mathbb{S}$  has a unique past lightcone configuration  $\ell^-(\mathbf{r}, t)$ . Applying the  $\epsilon$ -function then gives a unique local causal state at that point,  $\xi(\mathbf{r}, t) = \epsilon(\ell^-(\mathbf{r}, t))$ . Therefore, the  $\epsilon$ -function provides a local point-wise mapping from a spacetime field  $X(\mathbf{r}, t)$  to an associated *local causal state field*, denoted  $S(\mathbf{r}, t) = \epsilon(X(\mathbf{r}, t))$ . Crucially, the locality of the  $\epsilon$ -function ensures that a spacetime field  $X(\mathbf{r}, t)$  and its associated local causal state field  $S(\mathbf{r}, t) = \epsilon(X(\mathbf{r}, t))$  share the same spacetime coordinate geometry  $\mathbb{S}$ . For each spacetime point  $(\mathbf{r}, t) \in \mathbb{S}$ ,  $X(\mathbf{r}, t) : \mathbb{S} \rightarrow \mathbb{R}^n$  provides the values of the physical variables at that point and  $S(\mathbf{r}, t) : \mathbb{S} \rightarrow \Xi$  gives the local causal state  $\xi(\mathbf{r}, t) = \epsilon(\ell^-(\mathbf{r}, t))$  at the point.

And so, as promised, the local causal states are local representations assigned to each point  $(\mathbf{r}, t)$  in spacetime  $\mathbb{S}$ . In this way, the local causal state field  $S(\mathbf{r}, t) = \epsilon(X(\mathbf{r}, t))$  provides a *spacetime segmentation* that assigns a local causal state **class label** to each point in spacetime.

The  $\epsilon$ -function's locality, together with the equivariance of lightcones under spacetime isometries, implies that the local causal states are also equivariant under spacetime isometries. That is, if  $g$  is an isometric transformation (preserves spacetime distances) then it commutes with the  $\epsilon$ -function:  $g \circ \epsilon = \epsilon \circ g$ . In the context of coherent structure segmentation, this means that local causal states associated with a particular coherent structure, say hurricanes, do not depend on the hurricane's spacetime location or orientation. In the Lagrangian Coherent Structure literature, this property ensures local causal states are an *objective* method for coherent structure identification [19, 24]. Analogous to the interpretation of equivariances in neural networks as *geometric deep learning* [49, 50], we can interpret the local causal states as a form of *geometric representation learning* due to their spacetime equivariance and shared coordinate geometry.

Note, too, that locality allows local causal states to capture system structure across scales. All spacetime points associated to a coherent structure may be assigned to a set of local causal states uniquely identifying that structure, independent of the structure's size [25]. Larger structures simply involve a larger number of spacetime points assigned to those structures' states.

Similarly, there is no requirement for balanced statistics across occurrences of local causal states. For example, in the results shown here there is often a single local causal state associated with an ambient "background" that occurs most commonly in spacetime, with local causal states associated to coherent structures being relatively rarer.

Finally, we emphasize the local causal state approach's flexibility. While all examples examined here are on discrete Euclidean spacetimes, the local causal states are well-defined in any spacetime geometry with a distance metric. That said, analyzing large datasets with two spatial dimensions is already computationally taxing, even with large supercomputer resources. In principle, though, the local causal states can be approximated for systems with higher spatial dimensions.

### Approximations and Reconstruction

Consider a *data-generating process*  $\mathcal{P}$  that produces spacetime fields  $X \in \mathcal{P}$  and can either be a natural system or a numerical model. Here, we are interested in real-valued fields, with  $X(\mathbf{r}, t) \in \mathbb{R}^n$ . Therefore, the spaces of all past and future lightcones for  $\mathcal{P}$ , denoted  $\mathcal{L}^-$  and  $\mathcal{L}^+$  respectively, are uncountable. Each distribution  $\Pr(\mathcal{L}^+ | \mathcal{L}^- = \ell^-)$  is a continuous probability density. Moreover, they are densities conditioned on measure-zero events  $\mathcal{L}^- = \ell^-$ . There are an uncountable number of past lightcone configurations  $\ell^-$  and, for each of them, there is a conditional density  $\Pr(\mathcal{L}^+ | \mathcal{L}^- = \ell^-)$  over an uncountable number of future lightcone configurations.

Thus, in practice approximation is necessary to empirically reconstruct the conditional distributions  $\Pr(\mathcal{L}^+ | \mathcal{L}^- = \ell^-)$ . Most importantly, we replace the measure-zero past lightcone configurations with something of finite measure. In particular, we partition the space of past lightcones  $\mathcal{L}^-$  into finite-measure events  $\mathbf{p}^- \in \Sigma_{\mathcal{L}^-}$  for our empirical probability estimates. This allows the following empirical densities to be properly sampled from finite data:

$$\Pr(\mathcal{L}^+ | \mathfrak{P}^- = \mathbf{p}^-) = \int_{\ell^- \in \mathbf{p}^-} \Pr(\mathcal{L}^+ | \mathcal{L}^- = \ell^-) d\mu , \quad (\text{S8})$$

where  $\mu$  is the data-generating distribution over past lightcones. The finite-measure events partition  $\mathcal{L}^-$ , so that  $\bigcup_i \mathbf{p}_i^- = \mathcal{L}^-$ ,  $\mathbf{p}_i^- \cap \mathbf{p}_j^- = \emptyset$ , and  $\mu(\mathbf{p}_i^-) \neq 0$  for all  $\mathbf{p}_i^-$ .

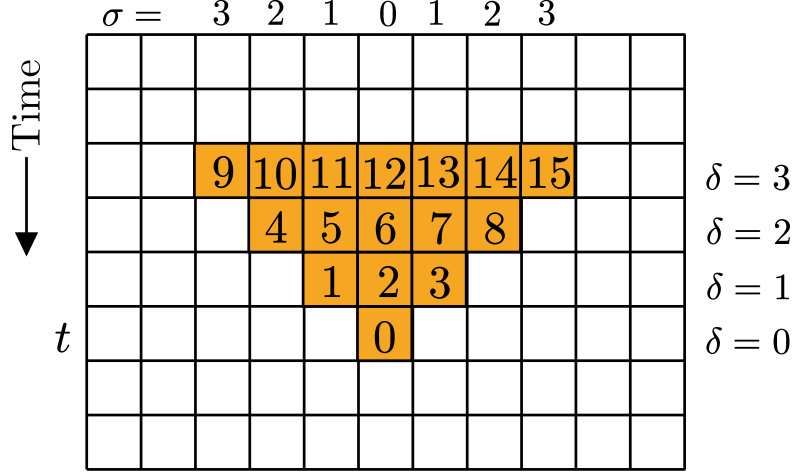


FIG. S2. Finite vector for the past lightcone configuration  $\ell^-(\mathbf{r}, t)$  in discrete 1+1 dimensional spacetime with  $c = 1$  and finite temporal horizon  $h^- = 3$ . Overlaid integers indicate the index of that site in the lightcone vector.  $\sigma$  is the space-only internal distance between the base of the lightcone,  $i = 0$ , and other sites in each vertical column. Similarly,  $\delta$  is the time-only internal distance for each site in the horizontal row.

To achieve such a finite partitioning, we rely on the *continuous histories assumption* [51, Assumption 3.1] which states that if two past lightcone configurations are similar, they should have similar conditional distributions over future lightcones. Formally, this says that the  $\epsilon$ -function is continuous over the space of past lightcones, and so  $\Pr(L^+|L^- = \ell_j^-) \rightarrow \Pr(L^+|L^- = \ell_i^-)$  as  $\ell_j^- \rightarrow \ell_i^-$ .

The continuous histories assumption leads us to employ distance-based clustering to create the finite partitioning of the past-lightcone space  $\mathcal{L}^-$ . Let  $\gamma : \ell^- \mapsto \mathbf{p}^-$  be the mapping from past lightcone configurations to their distance-based cluster element. We refer to individual cluster elements  $\mathbf{p}^-$  as *pasts*. The  $\gamma$ -function induces an equivalence relation over past lightcone configurations, similar to the  $\epsilon$ -function:

$$\begin{aligned} \ell_i^- \sim_\gamma \ell_j^- &\iff \ell_i^- \in \mathbf{p}_a^- \text{ and } \ell_j^- \in \mathbf{p}_a^- \\ \ell_i^- \sim_\gamma \ell_j^- &\iff \gamma(\ell_i^-) = \gamma(\ell_j^-) . \end{aligned}$$

Two past lightcone configurations are  $\gamma$ -equivalent if they are assigned to the same distance-based cluster  $\mathbf{p}_a^-$ .

With distance-based  $\gamma$ -equivalence, we state the actionable version of continuous histories our reconstruction algorithm uses as:

$$\gamma(\ell_i^-) = \gamma(\ell_j^-) \implies \epsilon(\ell_i^-) = \epsilon(\ell_j^-) . \quad (\text{S9})$$

That is, if two past lightcone configurations are assigned to the same cluster element from a distance-based clustering, we assume them to be “similar” and so assume that they have the same distribution over future lightcones.

To perform distance-based clustering over lightcones, we need a distance metric on the space of lightcones. First, note that in practice lightcones of finite temporal depth must be used. Individual lightcone configurations are collected into a vector, using a canonical ordering of their elements, up to a finite depth *horizon* cutoff  $h^\pm$  in time. Figure S2 shows a past lightcone vector in 1+1 dimensional spacetime with  $c = 1$  and  $h^- = 3$ . Integers overlaid on the lightcone sites indicate the indices of the lightcone vector in the canonical ordering we use.

Prior work [51, 52] used a Euclidean distance over lightcones. In practice with a finite horizon cutoff, this means that values in the lightcone below the cutoff are given a uniform weighting when computing distances, while lightcone values beyond the cutoff are given zero weight. To smooth this step discontinuity, our algorithm uses an exponentially-decaying lightcone distance  $D_{lc}$  given as

$$D_{lc}(\mathbf{a}, \mathbf{b}) \equiv \sqrt{(a_0 - b_0)^2 + \dots + e^{-\tau d(l)}(a_l - b_l)^2} , \quad (\text{S10})$$

where  $\mathbf{a}$  and  $\mathbf{b}$  are length- $l$  finite lightcone vectors, as depicted in Figure S2, and  $\tau$  is the decay rate.

The decay is applied relative to an internal distance  $d(i)$  between the base of the lightcone vector, with index  $i = 0$ , and other sites in the lightcone with indices  $0 < i \leq l$ . Figure S2 shows a space-only internal distance as  $d(i) = \sigma(i)$

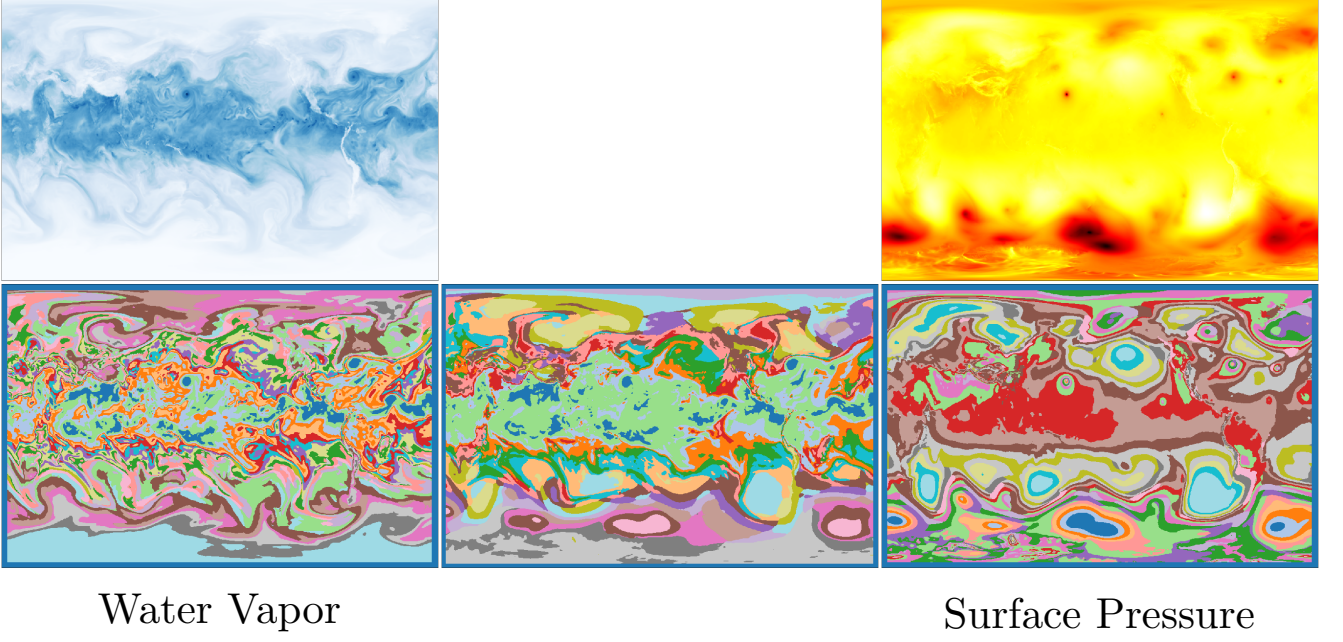


FIG. S3. Local causal state segmentations of TMQ and PSL, with a linear multivariate combined segmentation in between.

and a time-only internal distance as  $d(i) = \delta(i)$ . An internal spacetime distance is given as  $d(i) = \sqrt{\sigma(i)^2 + \delta(i)^2}$ . Results here use the latter spacetime internal distance. This family of exponentially-decaying lightcone distances smooth the step-discontinuity of a simple Euclidean distance by giving less significance to sites in the lightcone that are further away from the present (the base of lightcones).

In addition, the exponential decay has a practical benefit of reducing the effective dimensionality of the lightcone vectors during clustering, as discussed in more detail in Ref. [26]. We use K-Means [27] for our distance-based clustering of lightcones. Comparison with DBSCAN [53], along with the distributed High-Performance Computing (HPC) implementations of these algorithms, is also given in Ref. [26].

After performing a  $\gamma$ -partitioning of  $\mathcal{L}^-$  using K-Means clustering, the densities  $\Pr(\mathbf{L}^+|\mathfrak{P}^- = \mathbf{p}^-) = \int_{\ell \in \mathbf{p}^-} \Pr(\mathbf{L}^+|\mathbf{L}^- = \ell^-)d\mu$  can be empirically sampled from finite data. To simplify further for our HPC implementation, we also perform a (separate) K-Means clustering over future lightcones to similarly produce finite-measure *futures*  $\mathbf{p}_i^+$ . This gives a discrete approximation  $\Pr(\mathfrak{P}^+|\mathfrak{P}^- = \mathbf{p}_i^-)$  that can be sampled by simple counting. That is, each co-occurring set of pasts  $\mathbf{p}_a^- = \gamma^-(\ell^-(\mathbf{r}, t))$  and futures  $\mathbf{p}_b^+ = \gamma^+(\ell^+(\mathbf{r}, t))$  are collected in a matrix  $\mathfrak{P}$ , where  $\mathfrak{P}_{ab}$  is the number of co-occurrences of  $\mathbf{p}_a^-$  and  $\mathbf{p}_b^+$ .

With the empirical distributions  $\Pr(\mathfrak{P}^+|\mathfrak{P}^- = \mathbf{p}^-)$  we use an empirical approximation to causal equivalence over pasts, rather than over past lightcone configurations. (In principle, the densities  $\Pr(\mathbf{L}^+|\mathfrak{P}^- = \mathbf{p}^-)$  can also be used.) If two empirical future distributions are close, according to some empirical test, then their pasts are  $\psi$ -equivalent:

$$\begin{aligned} \mathbf{p}_i^- \sim_{\psi} \mathbf{p}_j^- &\iff \Pr(\mathfrak{P}^+|\mathbf{p}_i^-) \approx \Pr(\mathfrak{P}^+|\mathbf{p}_j^-) \\ \mathbf{p}_i^- \sim_{\psi} \mathbf{p}_j^- &\iff \psi(\mathbf{p}_i^-) = \psi(\mathbf{p}_j^-) . \end{aligned}$$

We use hierarchical agglomerative clustering with a chi-squared similarity test ( $p = 0.05$ ) for  $\psi$ -equivalence.

Thus, the approximation of the  $\epsilon$ -function our reconstruction algorithm uses is:

$$\epsilon(\ell^-) \approx \psi(\gamma(\ell^-)) . \quad (\text{S11})$$

Note that because past lightcones of depth 0 reduce to the lightcone base—simply the field value at that point in spacetime—our  $\gamma$  clustering over depth 0 past lightcones reduces to a standard K-Means clustering over the spacetime field. Recent insights into deep learning utilizing maximum affine spline operators [54] shows that trained deep learning models partition their input space, just as under distance-based clustering. This implies “that a D(eep)N(eural network) constructs a set of signal-dependent, class-specific templates” [54], formally similar to local causal state



representations. Whereas a neural network uses “ground truth” labels during the supervised learning process to build its templates, the local causal states must be learned in an unsupervised fashion. Our use of lightcones is motivated by weak causality and a local spacetime generalization of delay-coordinate embeddings and their associated intrinsic geometry [55]. This, combined with predictive equivalence, provides the necessary physics of organization [56] with which to extract coherent structures [25].

### Multivariate Interpolation

To reconstruct approximate local causal states for multivariate systems with  $m$  physical fields, we aggregate lightcone distances over all the fields. The simplest way to do this is to use the following tensor lightcone metric:

$$D_\ell(\mathbf{a}, \mathbf{b}) \equiv \sqrt{w_0 \left( (a_0^0 - b_0^0)^2 + \dots + e^{-\tau d(n)} (a_n^0 - b_n^0)^2 \right) + \dots + w_m \left( (a_0^m - b_0^m)^2 + \dots + e^{-\tau d(n)} (a_n^m - b_n^m)^2 \right)}.$$

A practical advantage of this particular metric is that it can be achieved by simply concatenating the scalar lightcone vectors taken from each field. Empirically, we observe that this metric yields a local causal state field that is a weighted interpolation of the component fields, with relative weighting given by the coefficients  $w_i$ . We use the water vapor and surface pressure fields of the CAM5.1 model to demonstrate this.

From visual inspection, signatures of hurricanes can be seen in the column-integrated water vapor field (TMQ) and surface pressure field (PSL), among others. Running local causal state segmentation on these fields alone is not sufficient for producing a unique segmentation class corresponding to hurricanes. The segmentation class of the TMQ field corresponding to hurricanes also shows up in the tropics in regions of high water concentration, which are not hurricanes. Similarly, the PSL segmentation class corresponding to hurricanes shows up in similar, but larger scale, pressure patterns in the extra tropics (particularly in the southern hemisphere).

Figure S3 gives example segmentations of these fields. The middle of Figure S3 shows a multivariate segmentation incorporating both fields, with the PSL field given  $\frac{1}{4}$  the weight of TMQ in the lightcone tensor metric described above. Qualitatively, one sees how clustering with this metric produces a weighted interpolation between segmentations of the two fields separately. In this case, the tropics and northern hemisphere are more reflective of the TMQ field, while the southern hemisphere is more reflective of the PSL field. Unfortunately, interpolating features from both fields together does not result in the desired unique hurricane segmentation class.

## Application Problems and Data Sets

### Turbulence

The formal study of emergent organization had its genesis with Bénard’s work [57] on the spontaneous formation of fluid convection cells [58–62] at the turn of the 20<sup>th</sup> century. Around the same time, the dawn of quantum mechanics took prominence in fundamental physics. Even up to the present, however, fluid turbulence has remained lingering as “the last mystery in classical physics”. Arguably, the primary “mystery” is that of emergent organization [20], with turbulence being the flagship instantiation [63].

The importance of secondary large-scale coherent structures in general theories of turbulence was recognized in the latter half of the 20<sup>th</sup> century, as documented in Ref. [64]. The first exploration of turbulent coherent structures were statistical in nature. It was observed that turbulence has a statistical separation of scales in fluctuations about the mean flow. There are smaller, fine-scaled “random” fluctuations on top of larger “not random” fluctuations, the latter being identified as coherent structures [64]. This statistical organization was identified and corroborated through characteristic signatures in the correlation function [65, 66].

Structural approaches, rather than statistical, appeared later, particularly with the introduction of the Proper Orthogonal Decomposition in the study of turbulence [67]. The POD approach follows an Eulerian paradigm, viewing the full spatial fields as the state of a spatially-extended dynamical systems. The spatial fields are then decomposed into POD modes, which have an optimality property that the leading modes capture the dominant contributions to kinetic energy. A (relatively) low-dimensional dynamical system analysis of the evolution of the POD modes provides an analysis of the flow in terms of its dominant *energetic* coherent structures [22].

Also building on ideas and tools from low-dimensional dynamical systems, the geometric theory of coherent structures follows the Lagrangian paradigm. Rather than seeking to capture energetically-dominant structures, the La-

grangian approach seeks coherent structures that organize advective transport in the flow. Lagrangian coherent structures are thus found as the most attracting, repelling, or shearing material surface in the flow [19].

The local causal states are a general representation learning method like POD. However, when applied to fluid flows using Lagrangian lightcones, as described above, they are more closely related to Lagrangian coherent structure approaches.

To analyze the ability of local causal states to capture fluid coherent structures we examine vortex dynamics in two-dimensional turbulence. This particular flow is ideal for several reasons. First, it is two-dimensional, making it easier to analyze with computationally expensive algorithms. However, unlike simple two-dimensional flows like von Kármán vortex streets, the two-dimensional turbulent flow supports behaviors characteristic of anisotropic turbulence in three dimensions, particularly in geophysical flows [68].

Second, the coherent structures that emerge in two-dimensional turbulence—coherent vortices—are paradigmatic fluid coherent structures. Moreover, while a general theory of the emergent vortex dynamics remains elusive, the dynamics are known empirically and are qualitatively simple. Two like-signed vortices, under the right conditions, will pairwise merge when they are close enough. Third, this pairwise merging behavior results in a power-law decay of the total number of vortices in the flow over time [28, 29]. Identification of the power-law decay provides a quantitative metric to test potential vortex coherent structure detection methods. Note that the identification of the “theoretical” range for the decay rate  $\nu = -(0.71 - 0.75)$  is identified using a specialized vortex identification algorithm based on vorticity thresholding along with geometric considerations [28].

We use data generated from the Fluid2D solver publicly available on GitHub ([https://github.com/pvthinker/Fluid2d/tree/master/experiments/TwoDim\\_turbulence](https://github.com/pvthinker/Fluid2d/tree/master/experiments/TwoDim_turbulence)). The local causal state segmentation used to count vortex cores, shown in Figure 1 (b), uses the following parameters: past lightcone horizon 14, future lightcone horizon 2, propagation speed  $c = 1$ ,  $K = 3$  for past lightcone K-Means,  $K = 25$  for future lightcone K-Means, and a spacetime decay with decay rate  $\tau = 0.25$  for the lightcone distance metric.

### Climate

While spacecraft allow for direct video observations of portions of the atmosphere, similar to that of Jupiter shown below, direct observations of the relevant physical fields of the full atmosphere are not possible. Numerical models are used to aggregate inhomogeneous data sources and create *data images* of the physical fields most consistent with the observations [7]. This process is known as *reanalysis* and represents the closest one can come to “true” observational climate data [69].

In contrast to reanalysis data, general circulation models (GCM) simulate the dynamics of the atmosphere to make short-term predictions (e.g., numerical weather prediction) and long-term climate forecasts. Long-term forecasts are the key tool for analyzing the potential future impacts of global warming [7, 9]. Tools such as the local causal states provide an enhanced level of analysis to answer detailed questions, such as how the intensity and dynamics of extreme weather events will change under different warming scenarios.

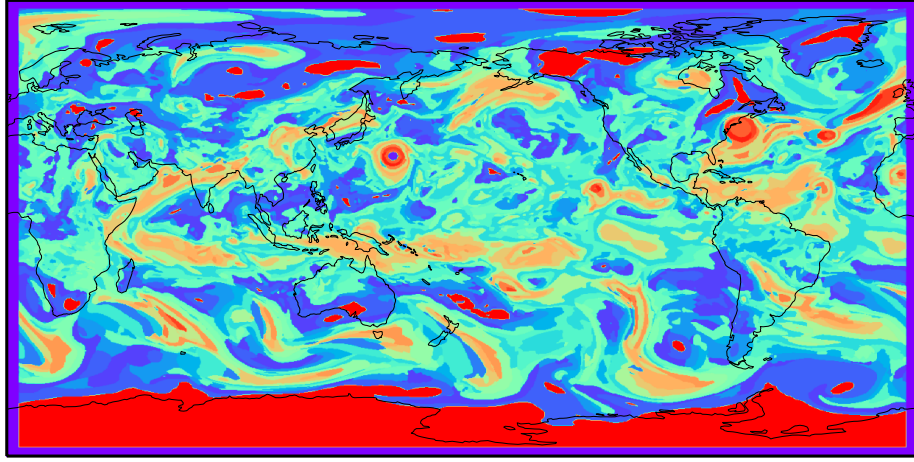
In this study, we use data from the historical calibration run of the high resolution  $0.25^\circ$  CAM5.1 GCM [32]. Single-precision climate variables are stored on an  $1152 \times 768$  spatial grid, with a temporal resolution of 3 hours. The segmentation results shown here use data from the autumn of 2015 of the CAM5.1 historical run 1. The physical fields used for the various local causal state segmentation results are the column-integrated water vapor (TMQ), the near-surface velocity components (U850) and (V850), as well as the integrated vapor transport (IVT) field, defined as

$$\left[ \left( \frac{1}{g} \int_{1000hPa}^{600hPa} qu \, dp \right)^2 + \left( \frac{1}{g} \int_{1000hPa}^{600hPa} qv \, dp \right)^2 \right]^{1/2},$$

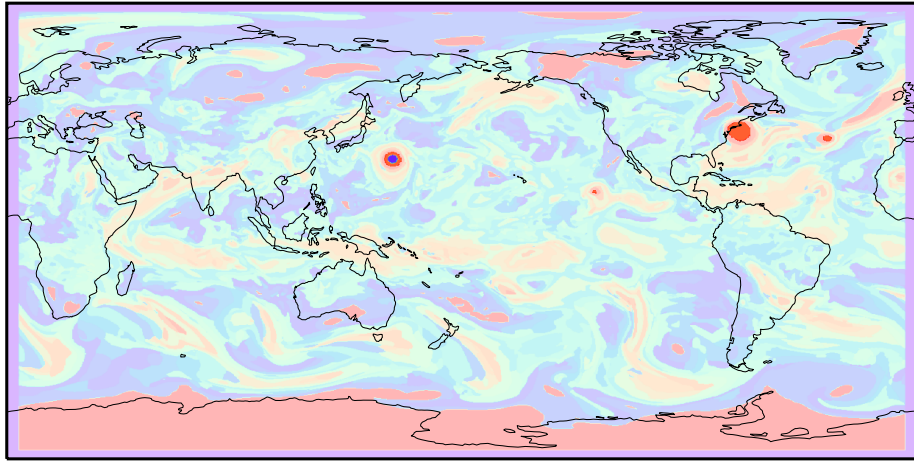
where here  $q$  is water vapor,  $u$  and  $v$  are wind velocity components,  $p$  is pressure.

### Hurricane Tracker

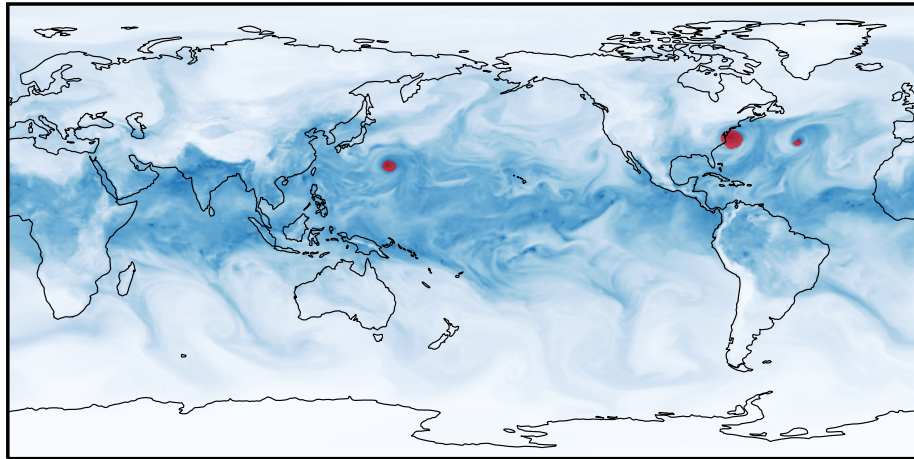
We create a hurricane tracker using a univariate local causal state segmentation of the IVT field, as shown in Figure S4. A snapshot of the local causal state field  $S(\mathbf{r}, t)$  from this segmentation is shown in (a). From visual inspection, we find a particular set of **hurricane** states that seem to co-occur with hurricanes. That is, it appears that (almost) all spacetime points  $(\mathbf{r}, t)$  such that  $S(\mathbf{r}, t)$  is one of the **hurricane** states are points where a hurricane



(a) Local causal state segmentation of IVT field



(b) hurricane local causal states highlighted



(c) hurricane states overlaid on top of water vapor field

FIG. S4. Creation of hurricane tracker using local causal state segmentation of the IVT field.

is present (as seen in the TMQ field). The **hurricane** states are highlighted in the local causal state segmentation of Figure S4 (b). To aid the visual identification of the **hurricane** states with actual hurricanes, a snapshot of the water vapor field (TMQ) is shown in (c) with spacetime points  $(\mathbf{r}, t)$ , such that  $S(\mathbf{r}, t)$  is one of the **hurricane** states, given a red mask overlaid on top of the vapor field. We emphasize again that this is possible due to the shared coordinate geometry of the physical fields and the local causal state field.

The local causal state segmentation of the IVT field, shown in Figure S4, was created using the following parameters: past lightcone horizon 16, future lightcone horizon 3, propagation speed  $c = 1$  (one spatial grid cell per single 3-hour time step),  $K = 24$  for past lightcone K-Means,  $K = 50$  for future lightcone K-Means, and decay rate  $\tau = 1.0$  for the lightcone distance metric.

### General Extreme Weather Events

The more general extreme weather local causal states shown in Figure 3 are produced using a multivariate segmentation that combines the TMQ, U850, and V850 vapor and velocity fields using a tensor lightcone metric based on IVT,

$$D_{lc}(\mathbf{a}, \mathbf{b}) \equiv \sqrt{\left( (a_0^q a_0^u - b_0^q b_0^u)^2 + \dots + e^{-\tau d(n)} (a_n^q a_n^u - b_n^q b_n^u)^2 \right) + \left( (a_0^q a_0^v - b_0^q b_0^v)^2 + \dots + e^{-\tau d(n)} (a_n^q a_n^v - b_n^q b_n^v)^2 \right)}. \quad (\text{S12})$$

Here,  $\mathbf{a}$  and  $\mathbf{b}$  are lightcone tensors such that lower indices are the lightcone locations (as shown in Figure S2) and upper indices are the physical fields;  $q$  is water vapor, and  $u$  and  $v$  are the near-surface velocity components. Using this metric, the local causal state segmentation shown in Figure 3 is reconstructed using the following parameters: past lightcone horizon 6, future lightcone horizon 3, propagation speed  $c = 2$ ,  $K = 14$  for past lightcone K-Means,  $K = 20$  for future lightcone K-Means, and spatiotemporal decay rate  $\tau = 0.5$ .

As described above, the non-white states in Figure 3 represent general coherent structures, according to the local causal state definition [25], that we hypothesize correspond to extreme weather events (EWEs). These include known structures like hurricanes and atmospheric rivers, as well as additional as-of-yet unknown structures. To test the relevance of these structures (particularly the unknown ones) to weather extremes, we examine their relation to precipitation extremes using the PRECT 3 hour cumulative precipitation field of the CAM5.1 model. Note that PRECT is not used to reconstruct the EWE local causal states.

We follow a similar procedure as Reference [70], which uses supervised deep learning to track fronts and associate precipitation extremes with the identified fronts. Here, we count the number of global precipitation extremes that co-occur with the EWE local causal states (the non-white states in Figure 3). That is, we find the spacetime locations  $(\mathbf{r}, t)$  of all extreme precipitation events of a given percentile and check to see if  $S(\mathbf{r}, t)$  is one of the EWE states. If so, the extreme precipitation event co-occurs with the EWE local causal states. As given above, we find that co-occurrence of 45% of global precipitation extremes in the 90<sup>th</sup> percentile, 58% co-occurrence with extremes in the 99<sup>th</sup> percentile, and 75% co-occurrence in the 99.9 percentile.

Following Reference [36], we also breakdown precipitation extremes zonally, rather than globally. Examining just the tropics, we find 61.7% of precipitation extremes in the 90<sup>th</sup> percentile co-occur with EWE local causal states, 69.5% co-occurrence of extremes in the 99<sup>th</sup> percentile, and 76.7% co-occurrence with extremes in the 99.9 percentile. For the northern temperate mid latitudes we find co-occurrence of 39.5% of extremes in the 90<sup>th</sup> percentile, 54.8% of extremes in the 99<sup>th</sup>, and 79.2% in the 99.9 percentile. For the southern temperate mid latitudes, there is 42.1% co-occurrence of extremes in the 90<sup>th</sup> percentile, 60.8% in the 99<sup>th</sup>, and 76.4% in the 99.9 percentile. Reference [36] finds 51% of global extremes (in the 99<sup>th</sup> percentile) co-occur with fronts, with 75% co-occurrence in the mid-latitudes and 31% in the tropics. They examine a single (known) form of extreme weather event, whereas our local causal state analysis produces much more general forms of EWEs to associate with precipitation extremes.

### Complexity and Symmetry Breaking: Clouds of Jupiter

We close by examining several local causal state segmentations of the clouds of Jupiter. These examples compare the effects of various inference parameters on the resulting segmentation and serve to illustrate the idea of increasing complexity through symmetry breaking.

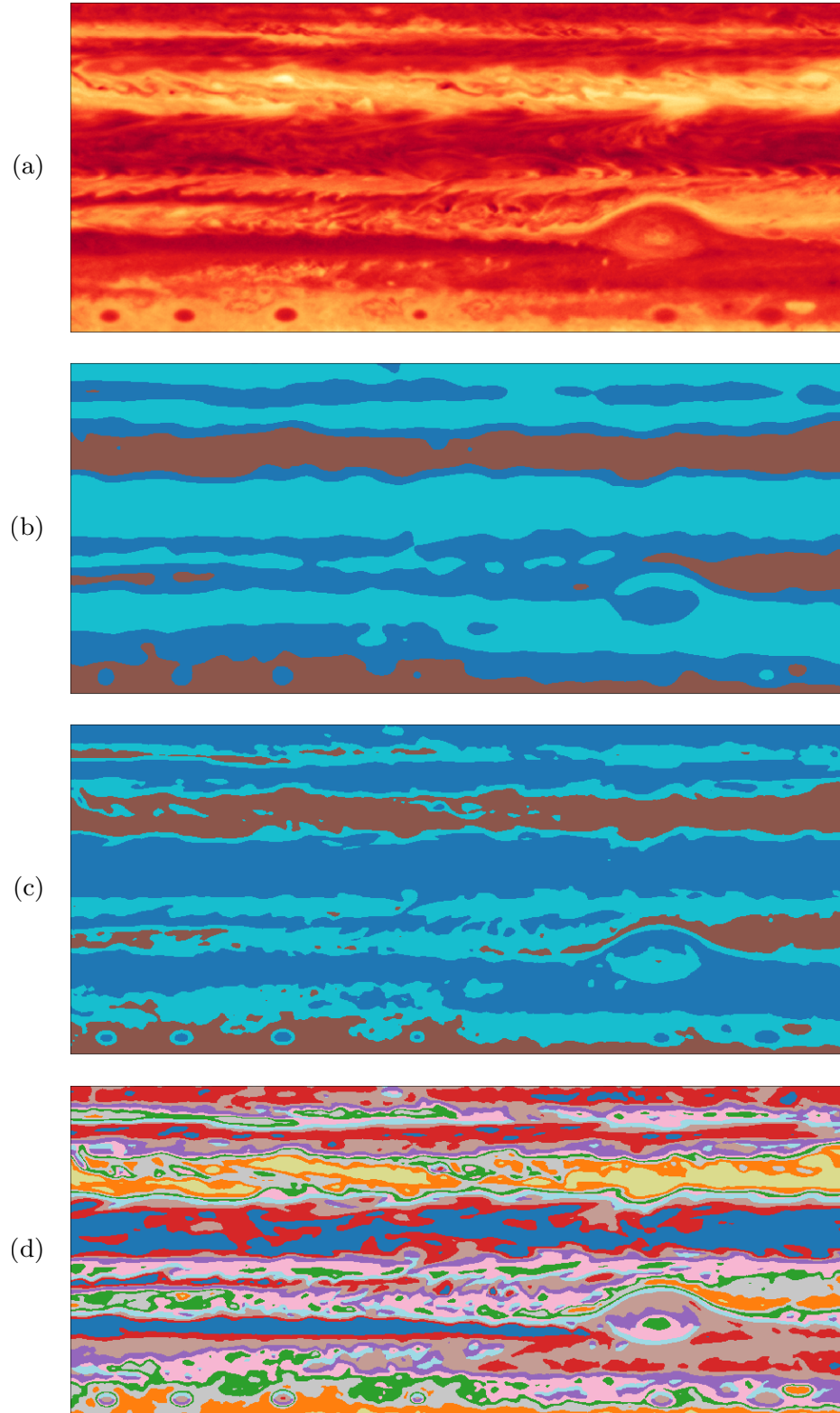


FIG. S5. Suite of local causal state segmentations (b)-(d) of Jupiter's clouds (a).

Figure S5 shows the results with the observable grayscale image of Jupiter’s clouds, taken from the NASA Cassini spacecraft [71], shown in (a). The largest scale structures that appear are the east-west zonal bands separated by strong jet streams [72]. To capture this large scale structure, we first perform a segmentation using a (relatively) large speed of light for the lightcones  $c = 4$  (memory requirements of the algorithm scale worst with  $c$ ). This effectively provides a larger spatial convolution kernel with each lightcone and thus averages over the smaller scale degrees of freedom within that kernel. The result, shown in (b), provides segmentation classes that follow the large scale structure and zonal bands. In addition, the boundaries between classes corresponding to the zonal bands identify the east-west jet streams that separate the bands.

To capture more smaller-scale detail, particularly at the boundaries of zonal bands, the segmentation in (c) uses the exact same parameters as (b), but includes a nonzero spatial decay in its lightcone metric. This decay effectively reduces the size of the spatial convolution kernels and so captures smaller scale detail. Most noticeable is the appearance of the “string of pearls” in the southern hemisphere that are much more clearly outlined in (c) than in (b). Note that the different spatial scales found between (b) and (c) could not be recovered using a standard K-Means segmentation;  $K$  is set to 3 for both cases, only the spatial decay rate differs.

While decreasing the (effective) size of the spatial kernel reveals smaller scale details at zonal band boundaries, we can see from (a) that there is additional turbulent structure within the bands themselves. The segmentation in (d) reveals more of this internal structure by decreasing the size of the spatial kernel further using  $c = 1$  and increasing the number of segmentation classes found; i.e., increasing  $K$  in K-Means. The increased complexity corresponds to system’s broken symmetries, reminiscent of the bifurcation theory of pattern formation. That is, points within a zonal band belong to the same segmentation class in (b), signaling they are part of the same large scale structure. However, they may correspond to different classes in (d), identifying the points belong to different structures at smaller scales. The separation of scales then represents an elevated level of complexity in the system’s spatial structure—a coexistence of organization and chaos (turbulence).

Note that in all segmentation analyses the local causal states vary more in the vertical direction of the images than the horizontal. As mentioned, the east-west horizontal bands are the most prominent structures in Jupiter’s atmosphere, created by strong jet streams that act as transport barriers. We again emphasize that Lagrangian Coherent Structures and our local causal state approach using Lagrangian lightcones discover coherent structures associated with material transport. It is not surprising then that local causal states are extended in the east-west direction following the zonal banding. Particularly evident in the coarse-grained segmentation shown in (b), the north-south boundaries between two local causal state bands then indicate the presence of a jet that acts as a transport boundary between the bands.

## CODE AVAILABILITY

Supporting Python source code, SLURM run scripts, parameter logs, and Jupyter notebooks displaying results and figures can be found at <https://github.com/adamrupe/Emergent-Organization>.

## AUTHOR CONTRIBUTIONS

AR and JPC developed the theoretical framework; KK, AR, and JPC conceptualized the problem applications; AR and KK performed the application experiments and analysis; AR wrote the prototype code; NK and AR wrote the distributed HPC code; AR and JPC wrote the manuscript.

## ACKNOWLEDGMENTS

The authors thank Vladislav Epifanov, Oleksandr Pavlyk, Frank Schlimbach, Mostofa Patwary, Sergey Maidanov, and Victor Lee for their help in developing the HPC implementation of local causal state reconstruction. We thank Nicolas Brodu, Jian Lu, Anastasiya Salova, and Mikhael Semaan for helpful comments and feedback. We also thank Wahid Bhimji for help with NERSC resources, and Prabhat initiating and leading the collaboration Project DisCo that ultimately led to the results in this work.

Part of this research was performed while AR was visiting the Institute for Pure and Applied Mathematics, which is supported by the National Science Foundation grant DMS-1440415. AR acknowledges the support of the U.S. Department of Energy through the LANL/LDRD Program and the Center for Nonlinear Studies. JPC would like



to acknowledge Intel® for supporting the IPCC at UC Davis. KK was supported by the Intel® Big Data Center. This research is based upon work supported by, or in part by, the U. S. Army Research Laboratory and the U. S. Army Research Office under contracts W911NF-13-1-0390 and W911NF-18-1-0028, and the U.S Department of Energy (DOE), Office of Science, Office of Biological and Environmental Research, **Earth and Environmental Systems Modeling program**. This work used resources of the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. The Pacific Northwest National Laboratory (PNNL) is operated for DOE by Battelle Memorial Institute under contract DE-AC05-76RLO1830.