Data-driven stochastic climate modeling and prediction

Dmitri Kondrashov
Department of Atmospheric and Oceanic Sciences, UCLA

In collaboration with Niraj Agarwal, Pavel Berloff, Mark Cane, Mickael Chekroun, Peter Dueben, Michael Ghil, James McWilliams, Sergey Kravtsov, Andy Robertson, Evgeniy Ryzhov

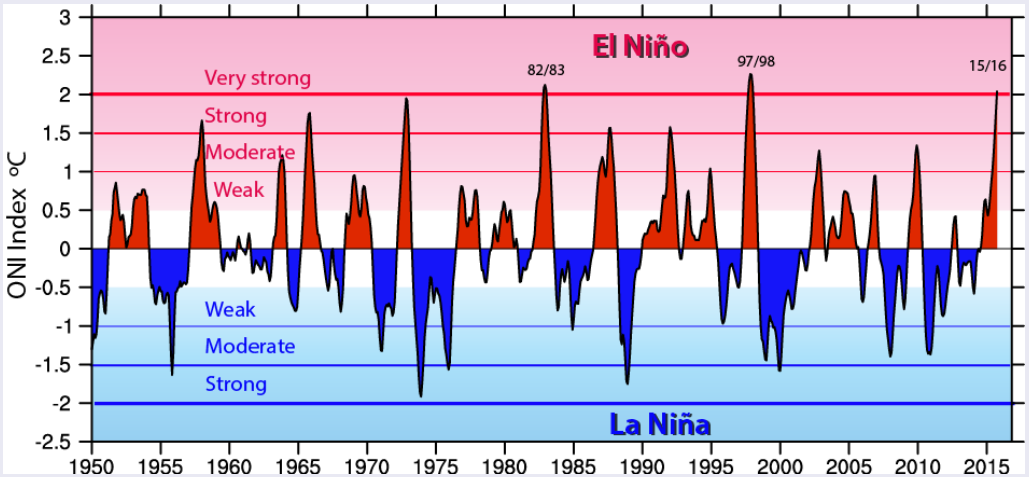
KITP, November 8, 2021

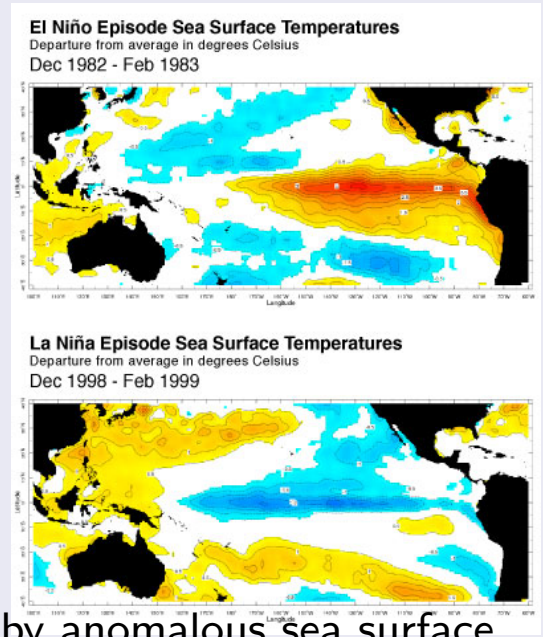
Outline

- State-of-the-art, highly resolved GCMs, while able to simulate detailed interactions within the climate system over a wide range of scales, generate detailed climate variability that is as complex as currently available observational datasets, and is hence no less challenging to interpret.
- Dynamical analysis of climatic phenomena typically involves a set of multiple GCM simulations that are designed to isolate physical processes governing the simulated, and by inference, observed climate variability.
- These simulations are computationally expensive, however, and their interpretation is hindered by the presence of model biases due to incomplete or imperfect parameterizations of the unresolved physical processes.
- While these GCMs represent a broad range of time and space scales, important aspects of observed climate variability can be represented by a substantially smaller number of degrees of freedom associated with large-scale (space) and low-frequency (time) modes of climate variability – (LFV).

- Thus it motivates development of **reduced-order approximations** of either the full governing equations or the phenomenon itself by inverse modeling approaches, (iii) **data-adaptive analysis** for identifying and predictive modeling of LFV modes by using output GCMs as well as observations, (iv) development of **purely data-driven ("Machine Learning") dynamical models** as statistical benchmarks.
- Data-driven dynamical climate models should account for
  - interactions between the LFV modes
  - estimation of the dynamical contribution of "fast-small" scales.
  - the interactions between the few LFV modes and the much large number of fast-small scale modes
- Depending on climate application, LFV can be on intraseasonal, interannual and decadal time scales.
- Intraseasonal-to-seasonal prediction of ENSO and Arctic Sea Ice.

Global mode of terrestrial climate





- Ocean-atmospheric mode of climate variability manifested by anomalous sea surface temperature (SST) in Equatorial Pacific.
- El Nino/ La Nina events start develop in Apr-Jun, maximum in Dec-Feb.
- Strong ENSO events cause extremes in weather and precipitation over the globe.
- Events tend to recur every 2 to 7 years but fairly irregularly.
- Despite good physical understanding, ENSO remains to be very challenging for seasonal prediction, i.e. from spring/early summer to upcoming winter ("Spring barrier").

Theoretical guidance: Mori-Zwanzig Formalism

- Given $x(t)$ – partial d.o.f of climate model, e.g. SST but not winds,...
- A closure problem [Kondrashov et al. 2015, *Physica D*]: How to find a low-order system of "closed" equations that describe evolution of partial observations $x(t)$?
- **Theoretical guidance** from the Mori-Zwanzig formalism of statistical mechanics – generalized Langevin equation [Palmer, 2019; Ghil and Lucarini, 2020]

$$\frac{dx}{dt} = F(x) + \int_0^t G(t, s, x(s))ds + \eta_t \qquad \text{(GLE)}$$

- F(x)– **self-interactions** among the observed d.o.f: Markovian contribution.
- Integral term – (non-linear) **cross-interactions** between the observed and unobserved d.o.f; it involves the past history of the observed d.o.f and brings non-Markovian contribution or memory effects.
- η_t – **spatio-temporal correlated noise** by unobserved d.o.f.
- In practice GLE solution ($F(x)$, $G(x)$) is computationally difficult to obtain except in very limited cases (time-scale separation).
- The goal is to find optimal $F(x)$, $G(x)$, η_t by inverse modeling techniques.
- Linear Inverse Modeling (LIM, Penland 1989): $F = Ax$, $G \equiv 0$, η_t is white noise.
- Extending LIM, i.e. by adding memory, nonlinearity and more complicated noise.

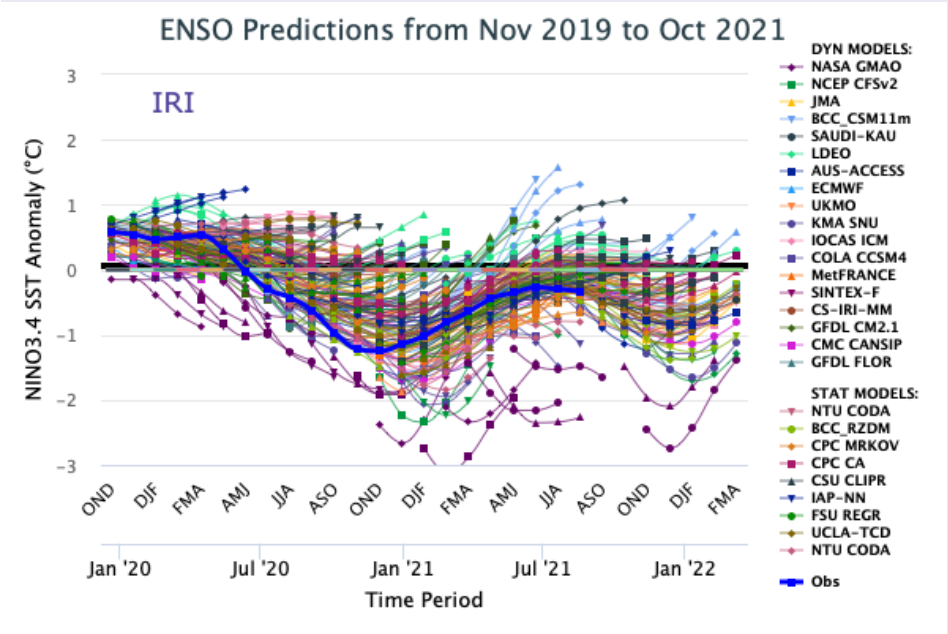
Multilayer Stochastic Modeling

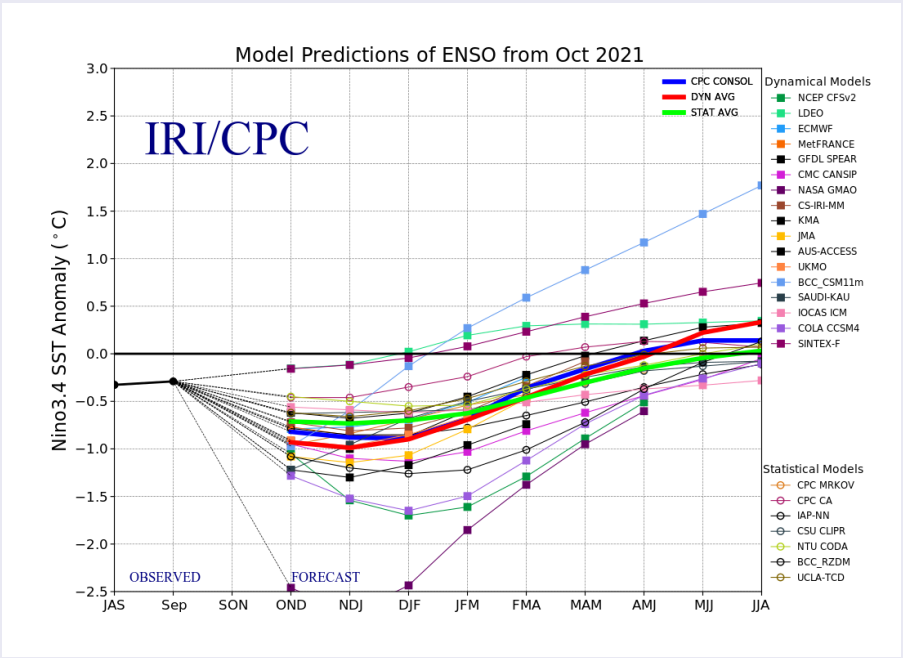
- Kondrashov et al. (*Physica D.*, 2015) have shown that **multilevel regression models,** such as Empirical Reduction Model (EMR, Kravtsov, Kondrashov and Ghil, *Cambridge UP*, 2009) can provide successful "Markovian" approximation of *GLE*:

$$\begin{aligned} \mathbf{x_{k+1}} - \mathbf{x_k} &= \Big[\mathbf{F} + \mathbf{A}\mathbf{x_k} + \mathbf{B}(\mathbf{x_k}, \mathbf{x_k})\Big]\delta t + \mathbf{r}_k^0 \delta t, \\ \mathbf{r}_{k+1}^{m-1} - \mathbf{r}_k^{m-1} &= \mathbf{L}^m \Big[\mathbf{x_k}, \mathbf{r}_k^0, ..., \mathbf{r}_k^{m-1}\Big]\delta t + \mathbf{r}_k^m \delta t,\ 1 \le m \le M, \end{aligned}$$

- $\mathbf{x} = (x_1, ..., x_d)$ are time series of d leading modes from data-adaptive decomposition, such as Principal Component Analysis (PCA).
- Model coefficients $\mathbf{F}, \mathbf{A}, \mathbf{B}, \mathbf{L}^m$ are estimated by consecutive *top-down* regressions and can include seasonal dependence; $\mathbf{r}_t^m$ are regression residuals.
- $\mathbf{r}_t^m$ represent hidden (unobserved) scales in a stack of "matrioshka" layers, each supplementary layer representing faster scales until $\mathbf{r}_t^M$ can be approximated by white noise.
- Multi-layer structure conveys memory effects, i.e. temporally correlated noise.
- Linear constraints on $\mathbf{B}$ for numerical stability, i.e. $<B(x,x), x> = 0$.
- Prediction is obtained by integrating model forward from given i.c. and forcing by ensemble of random noise realizations.

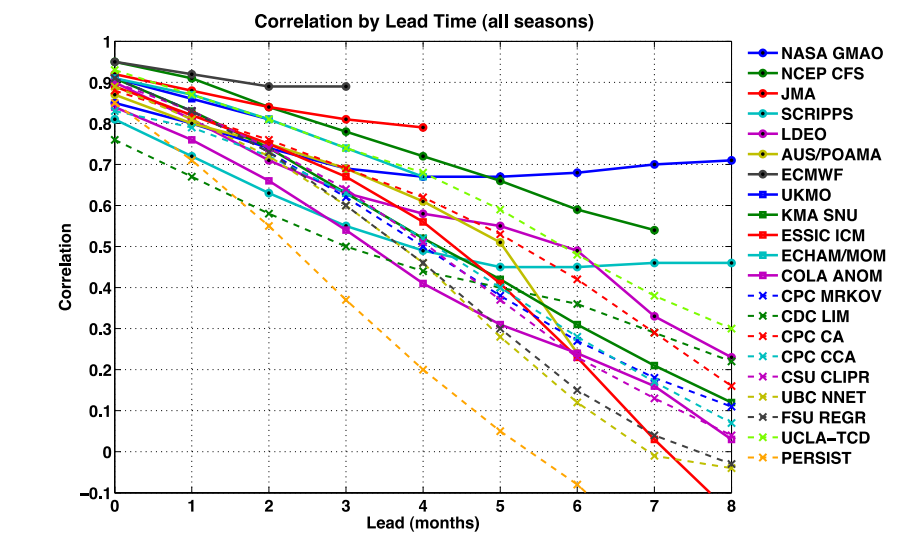
Real-time Prediction

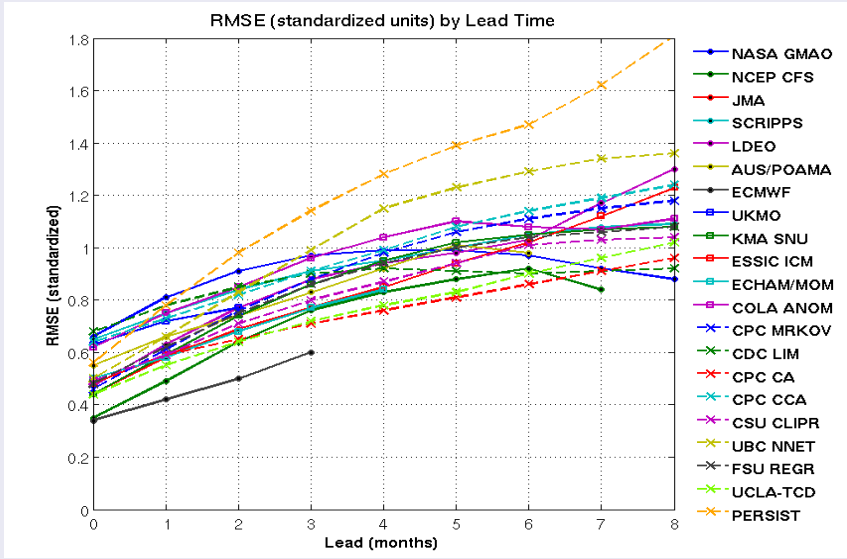




- IRI/CPC coordinated effort: statistical/dynamical multi-model plume to predict Nino3.4.
- UCLA-TCD: Quadratic, 2-level EMR-ENSO model for 20 leading EOFs of Equatorial SSTs [Kondrashov et al. 2005, J. Clim].

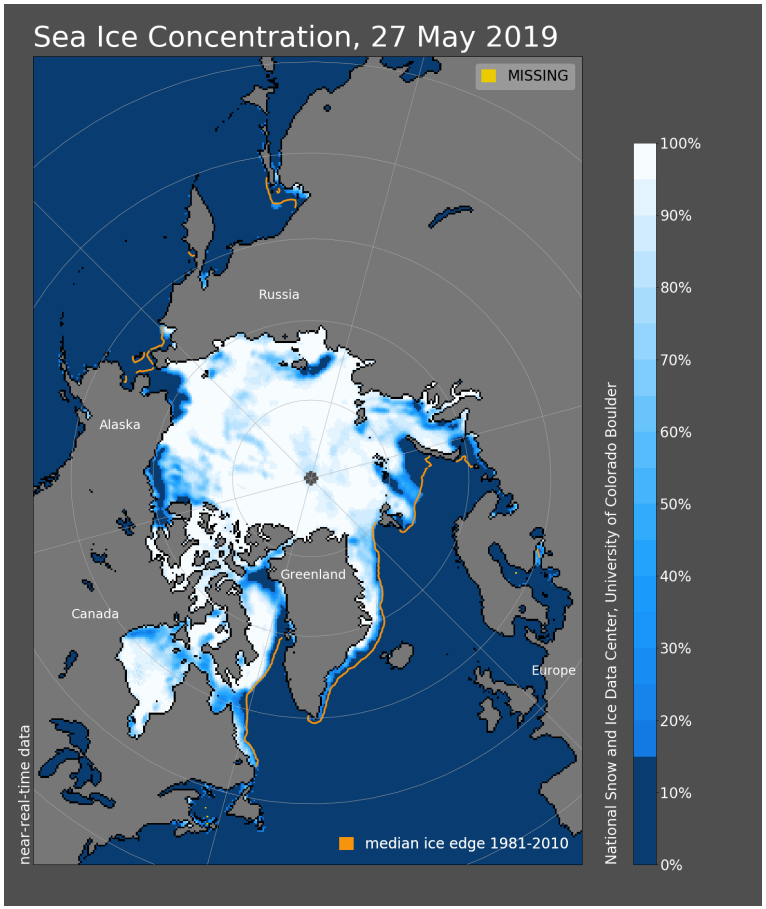
Real-time Prediction Skill in 2002–2011

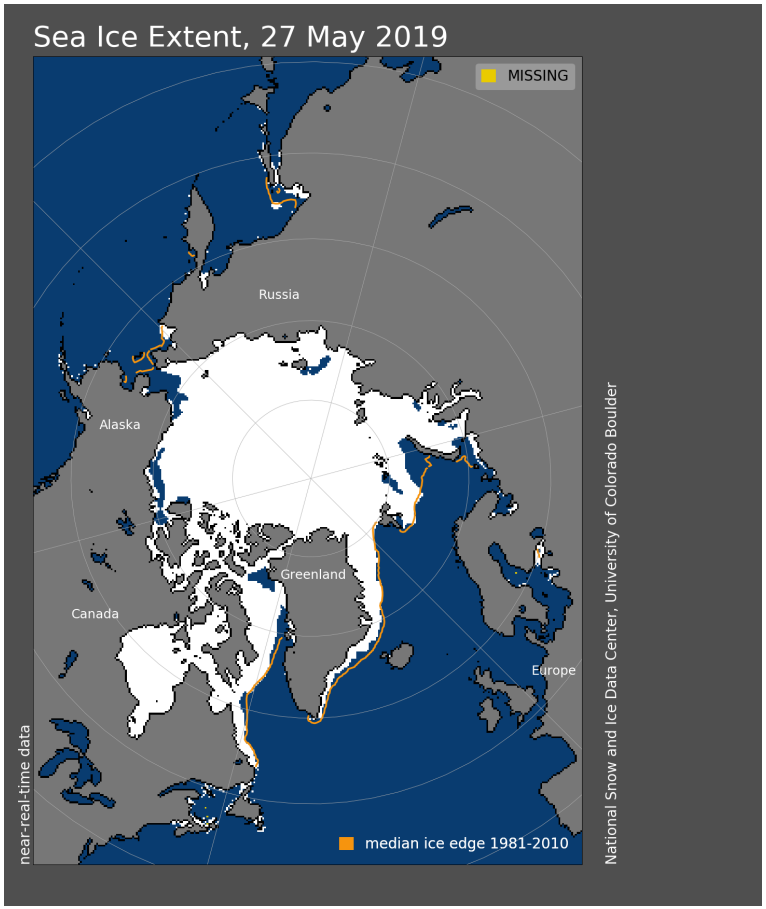




- Figures from Barnston et al.: "Skill of Real-Time Seasonal ENSO Model Predictions during 2002–11: Is Our Capability Increasing?." *Bull. Amer. Meteor. Soc.*, **93**, 631– 651, 2012
- "UCLA-TCD prediction has the highest seasonally combined correlation skill among the statistical models exceeded by only a few dynamical modes [...] as well as one of the smallest RMSE."
- the key to predictive success – conveying memory effects by multiple layer structure, which also helps to reduce spring barrier for prediction [Chen et al., 2016 *J. Climate*].

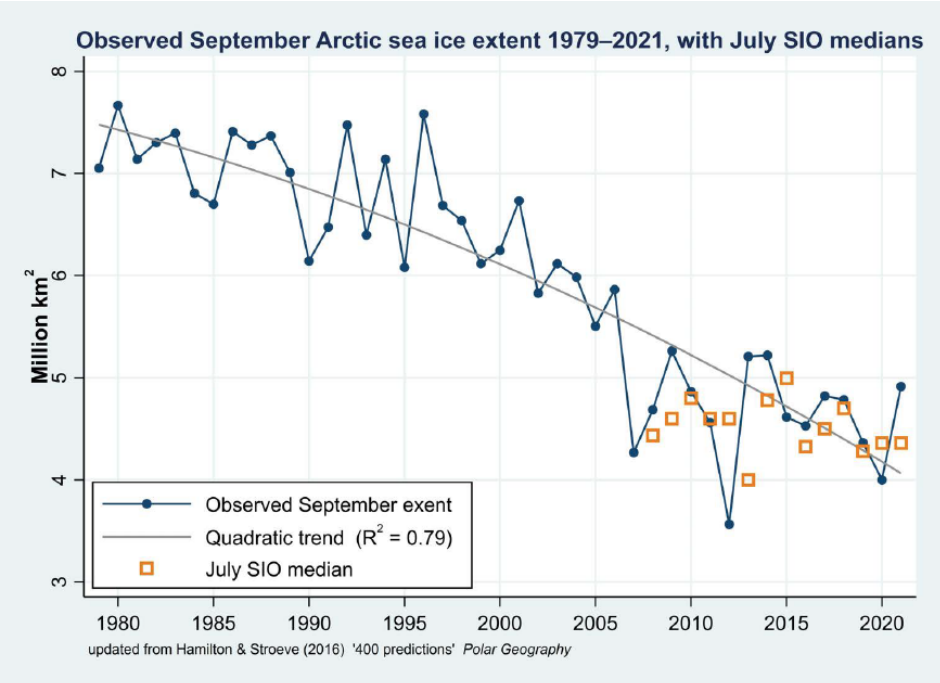
Arctic Sea Ice

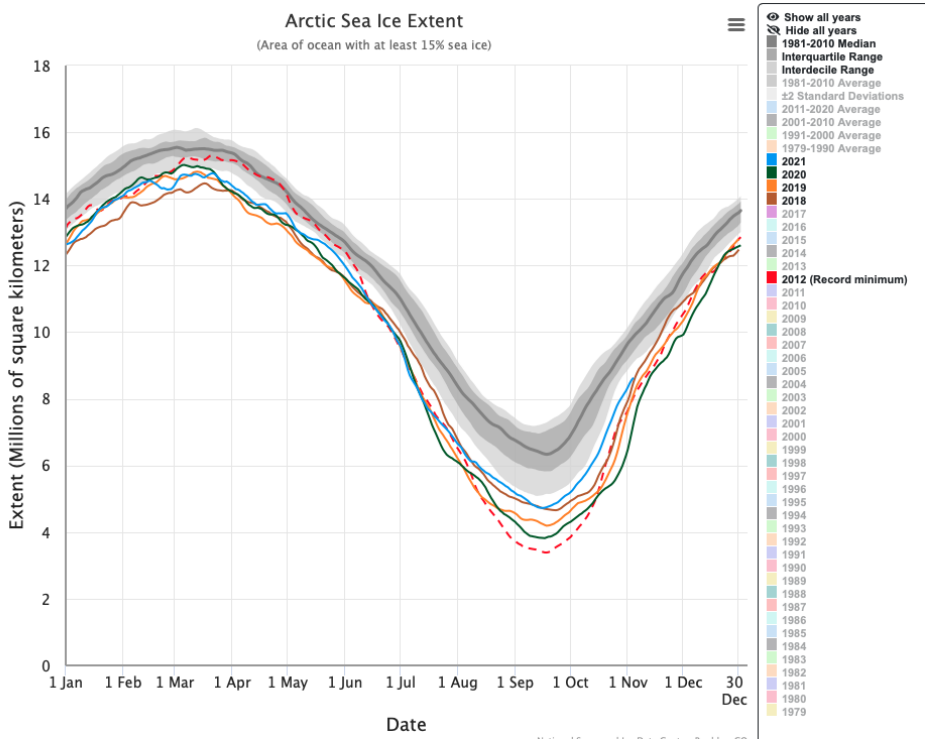




- Sea ice concentration (SIC) – relative area covered by ice (0% – 100%)
- SIC is derived from passive microwave imaging by satellites.
- Sea Ice Extent (SIE) is area covered by ice – when SIC$\geq 15\%$.
- SIE decline due to global warming has profound socio-economic implications, i.e. opening of summertime shipping Europe-to-Asia routes.
- Sea Ice Outlook (SIO) – coordinated effort for multi-model prediction, similar to ENSO.

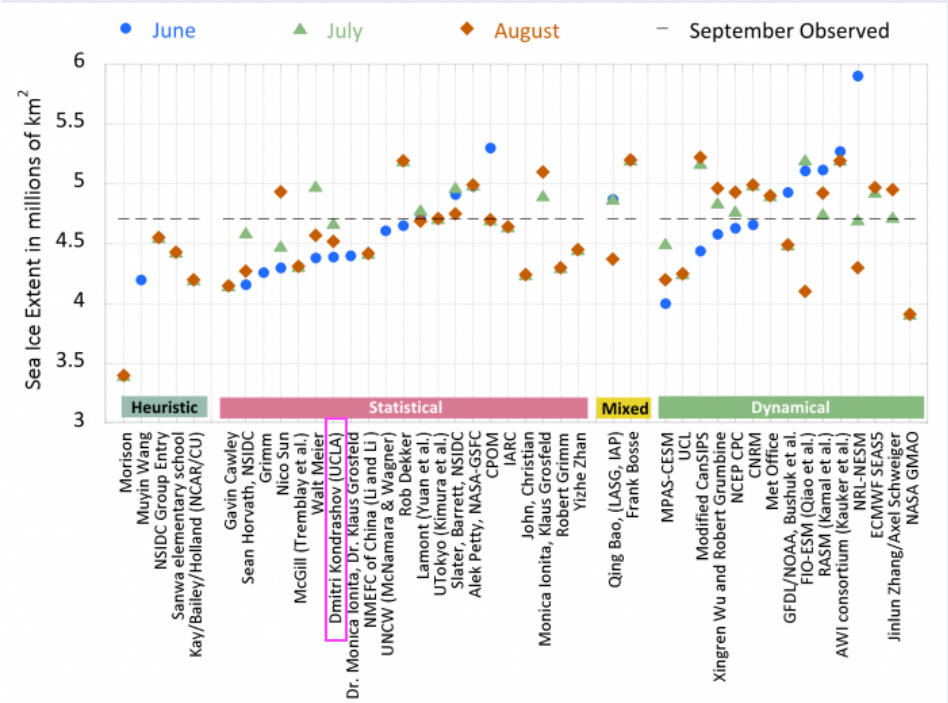
Summertime Arctic Sea Ice

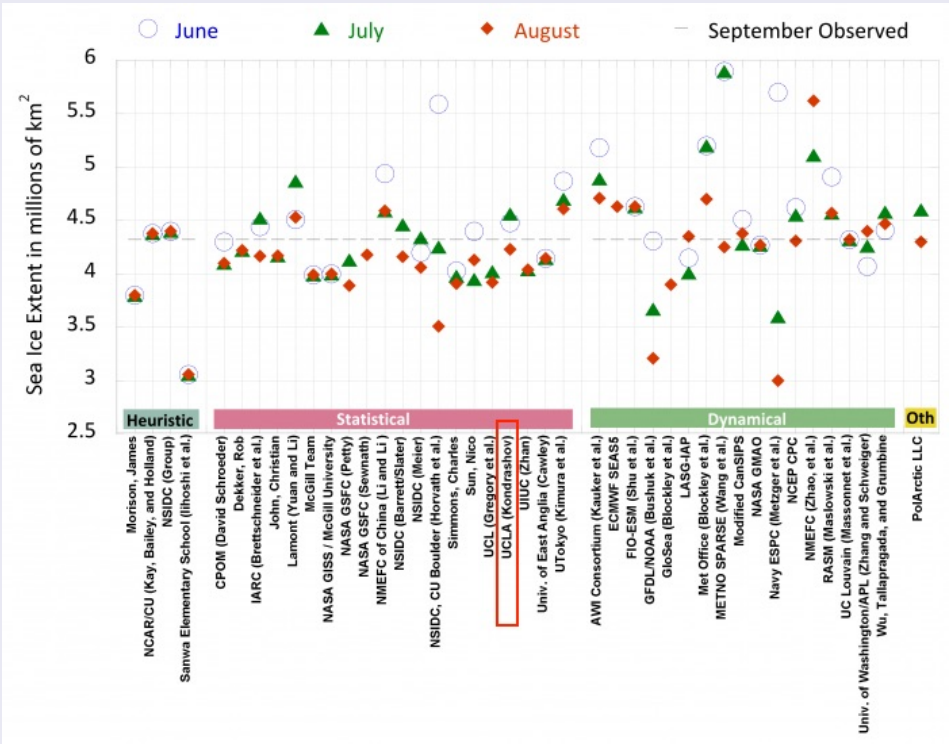




- **Summertime forecasting of September Sea Ice Extent (SIE) is very challenging**: high variability of O&A over Arctic, shortness of observational record, shortcomings of the physics-based models to simulate sea-ice dynamics; also **spring predictability barrier**!
- For **accurate prediction** need to consider dynamics of **SIE anomalies** over **Arctic regions**.
- **Data-adaptive Harmonic decomposition (DAHD)**
- **The key feature**: **DAHD** unites **data-adaptive decomposition** and **inverse modeling**.

2018 and 2019 DAHD prediction of September SIE





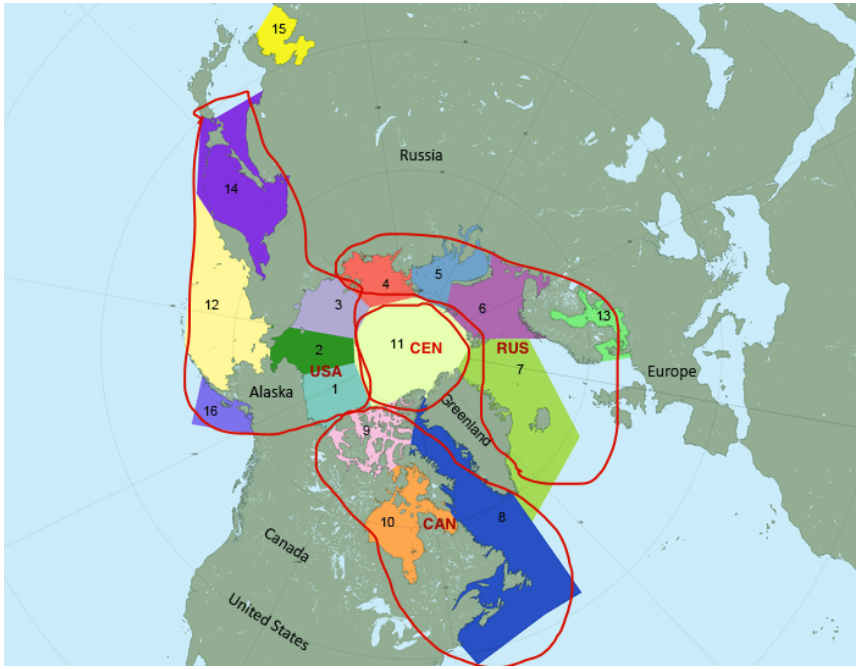
- JJA predictions of September SIE by DAHD are usually within ≈ 0.2 Mkm^2 of the obs: 2020: 4.40/3.92 (Mkm^2), 2019: 4.42/4.32 (Mkm^2) (right panel), 2018: 4.53/4.71 (Mkm^2) (left panel); 2017: 4.57/4.80 (Mkm^2); 2016: 4.9/4.72 (Mkm^2).

Kondrashov, D., M. D. Chekroun, and M. Ghil, 2018:

Data-adaptive harmonic decomposition and prediction of Arctic sea ice extent,

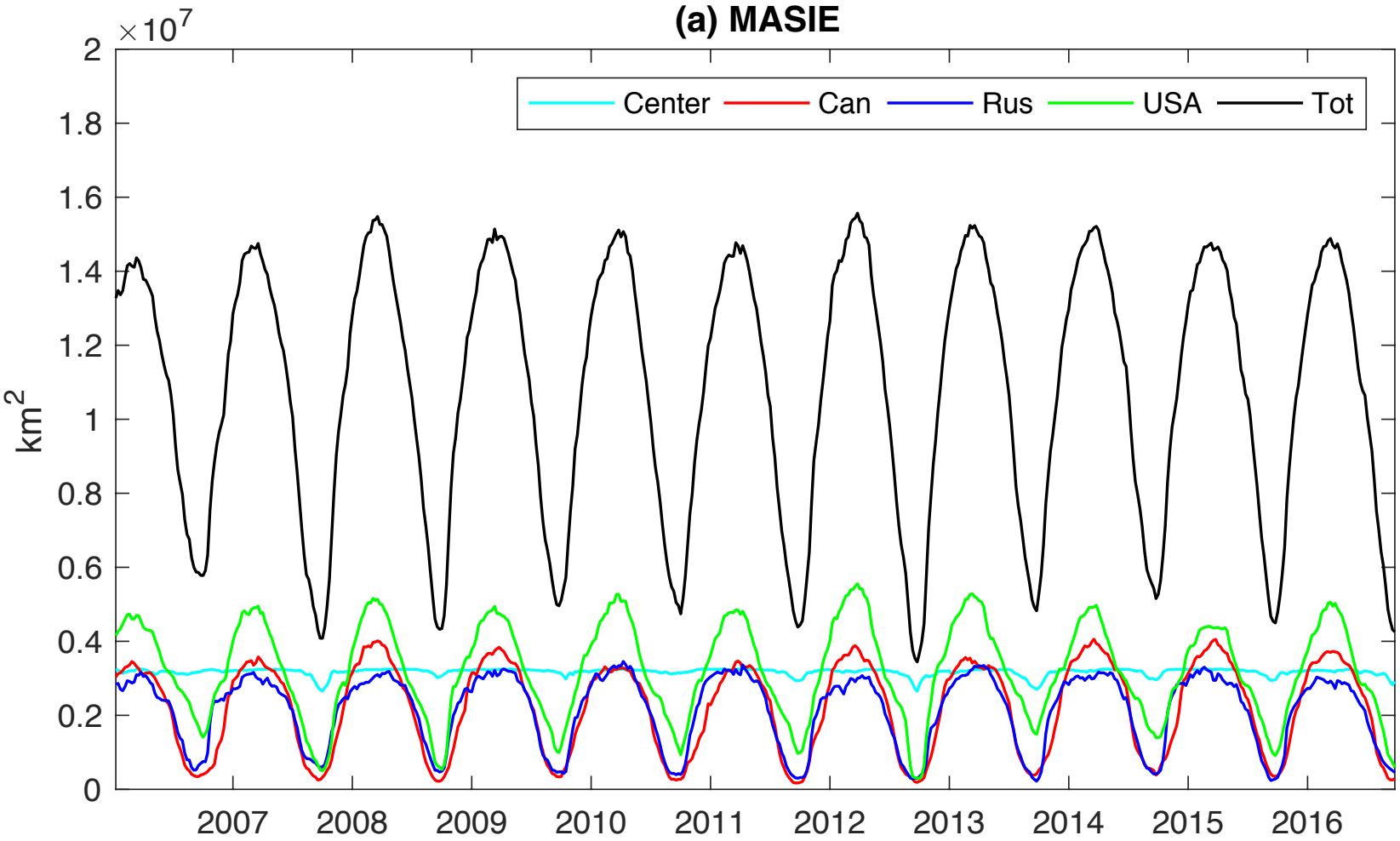
Dynamics and Statistics of the Climate System, doi:10.1093/climsys/dzy001.

Multisensor Analyzed Sea Ice Extent -I



- Multisensor Analyzed Sea Ice Extent (MASIE) – manual (non-automatic) data fusion using variety of sources. It provides daily ice conditions to support navigation and operational forecast models. 2006 – present, daily data of SIE for 16 Arctic regions in **real-time.**
- The daily MASIE data was aggregated into a weekly resolution with 52 weeks in each calendar year, and combined into **four Arctic sectors**: Center, RUS, CAN, and USA.

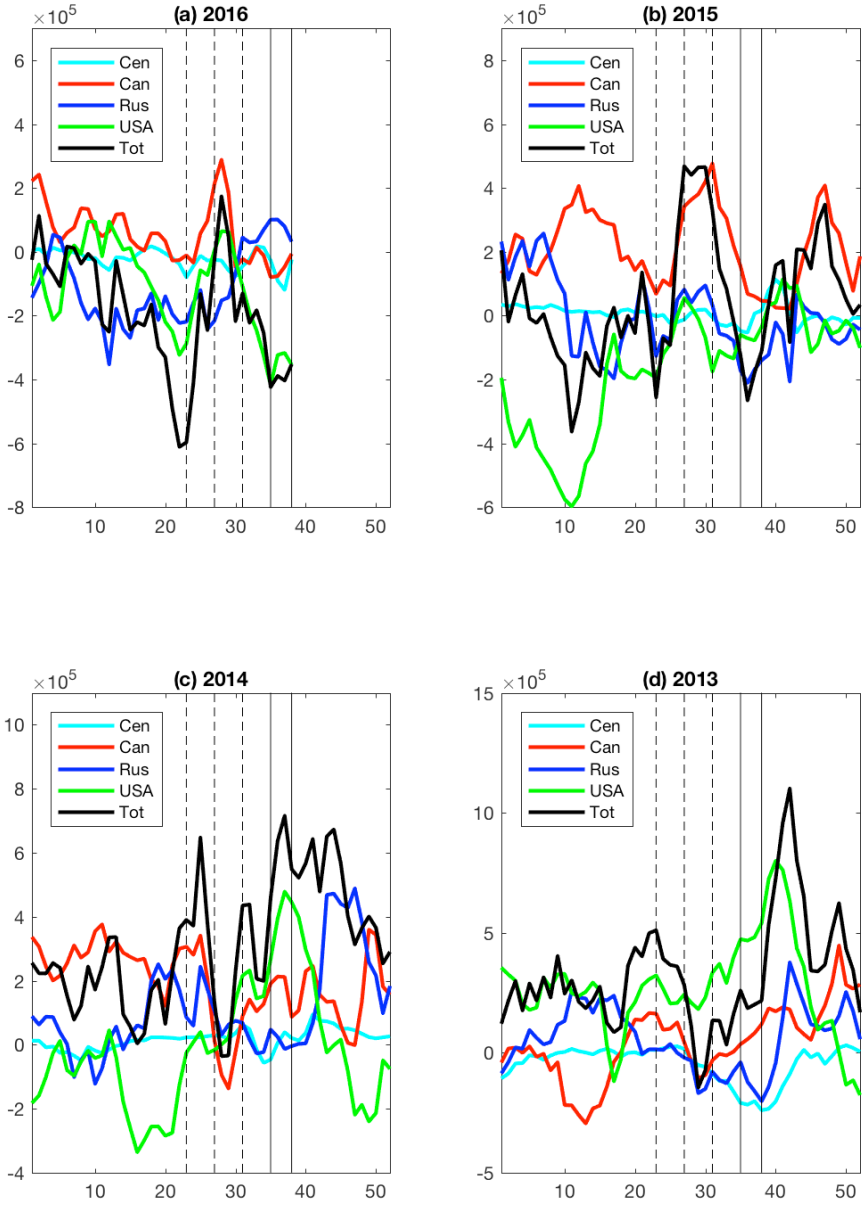
Multisensor Analyzed Sea Ice Extent –II



- Pronounced seasonal cycle; we are interested in **anomalies**.

Multisensor Analyzed Sea Ice Extent –III

- Different Arctic sectors are not isolated and necessarily coupled due to physical processes: advection, winds, O/A processes...
- Complex mixture of temporal scales (subseasonal-to-seasonal-to-intraseasonal) and interactions between the sectors.
- Relative contributions of the four sectors changes from year to year and abruptly during the summer; involve delays between the sectors and may act in opposite.
- Complex and highly nonlinear predictors are needed...?
- **DAHD** techniques help to characterize such complex time series evolution by a system of coupled nonlinear harmonic stochastic oscillators.



Data-Adaptive Harmonic Decomposition (DAHD) -I

- DAHD is a spectral time series analysis technique that utilizes spatio-temporal correlations [Chekroun and Kondrashov, *Chaos*, 2017; Kondrashov et al., *Fluids*, 2018].
- Given d-channel time series $\mathbf{X}(t_n) = (X_1(t_n), \ldots, X_d(t_n))$, $n = 1, \ldots, N$, compute two-sided cross-correlation coefficients $\rho_\tau^{(p,q)}$ at lag τ between channels p and q, where $-M+1 \leq \tau \leq M-1$, and M is time-embedding window.
- Form Hankel matrix by left-shift: l-circ$(\rho_{-M+1}^{(p,q)}, \cdots, \rho_{-1}^{(p,q)}, \rho_0^{(p,q)}, \rho_1^{(p,q)} \cdots, \rho_{M-1}^{(p,q)})$

$$\mathbf{H}^{(p,q)} = \begin{pmatrix} \rho_{-M+1}^{(p,q)} & \rho_{-M+2}^{(p,q)} & \cdots & \rho_0^{(p,q)} & \rho_1^{(p,q)} & \cdots & \rho_{M-1}^{(p,q)} \\ \rho_{-M+2}^{(p,q)} & \iddots & \iddots & \iddots & \iddots & \iddots & \rho_{-M+1}^{(p,q)} \\ \vdots & \iddots & \iddots & \iddots & \iddots & \iddots & \rho_{-M+2}^{(p,q)} \\ \rho_0^{(p,q)} & \iddots & \iddots & \iddots & \rho_{-M+1}^{(p,q)} & \iddots & \vdots \\ \rho_1^{(p,q)} & \iddots & \iddots & \iddots & \rho_{-M+2}^{(p,q)} & \iddots & \rho_0^{(p,q)} \\ \vdots & \rho_{M-1}^{(p,q)} & \rho_{-M+1}^{(p,q)} & \iddots & \iddots & \iddots & \vdots \\ \rho_{M-1}^{(p,q)} & \rho_{-M+1}^{(p,q)} & \rho_{-M+2}^{(p,q)} & \cdots & \rho_0^{(p,q)} & \cdots & \rho_{M-2}^{(p,q)} \end{pmatrix}$$

- Form **symmetric** grand block-Hankel matrix $\mathfrak{C}$ by d^2 blocks of size $(2M-1) \times (2M-1)$:

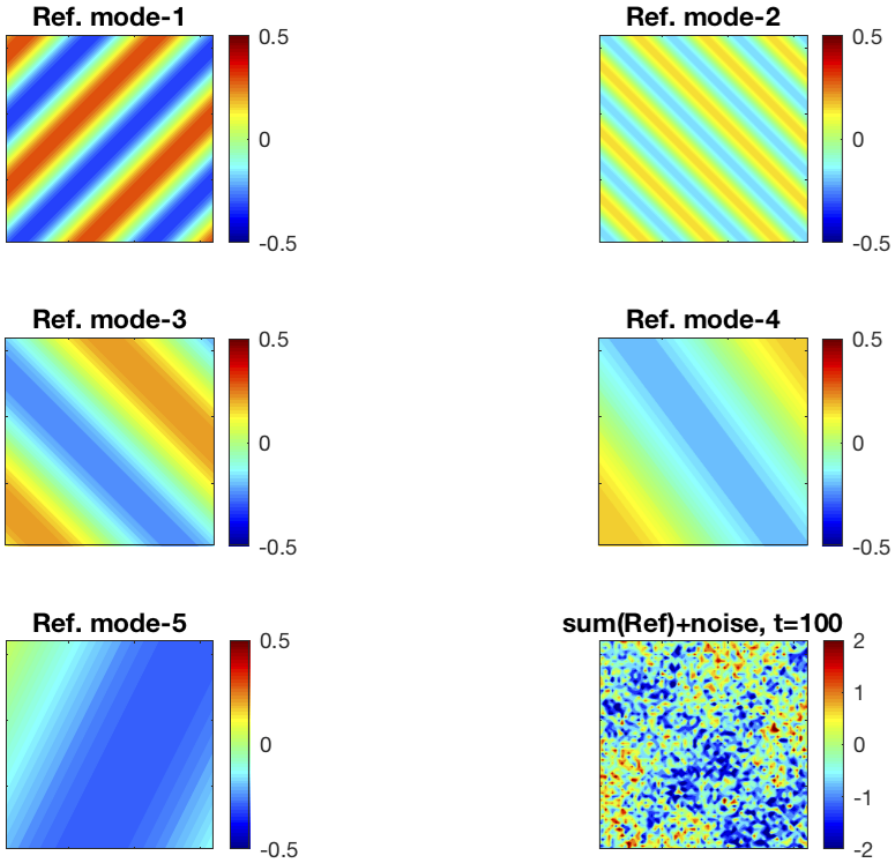
$$\mathfrak{C}^{(p,q)} = \mathbf{H}^{(p,q)}, \text{ if } 1 \leq p \leq q \leq d,$$

$$\mathfrak{C}^{(p,q)} = \mathbf{H}^{(q,p)}, \text{ else.}$$

Data-Adaptive Harmonic Decomposition (DAHD) - II

- Compute eigenvectors (DAHD modes) $\mathbf{W}_j$ and eigenvalues λ_j of $\mathfrak{C}$
- eigenvalues are paired $\lambda_j = -\lambda_{j'}$.
- $\mathbf{W}_j = (\mathbf{E}_1^j, \cdots, \mathbf{E}_d^j)^{\mathrm{T}}$ is **orthonormal** set of spatiotemporal oscillating functions: $\mathbf{E}_k^j(s) = B_k^j \cos(2\pi f s + \theta_k^j)$, $1 \le s \le 2M-1$, $1 \le k \le d$; θ_k^j and B_k^j are data-adaptive, and $f = \frac{(\ell-1)}{2(M-1)}$, $\ell = 1, \ldots, M$
- **each mode pair** is in **exact phase quadrature**, a.k.a. *sin* and *cos*: $\theta_k^{j'} = \theta_k^j + \pi/2$, and are **global space-time filters at their frequency.**
- Using $M = 1$ – yields Principal Component Analysis.
- Using right-shift of correlation sequence – Toeplitz matrix structure: Multichannel Singular Spectrum Analysis (M-SSA) [Ghil et al. 2002, *Rev. Geophys.*]
- **DAHD** provides for **rigorous frequency (temporal scale) separation** in their modes, while **M-SSA and PCA modes mix temporal scales.**
- **DAHD coefficients (DAHCs)**: $\xi_j(t) = \sum_{s=1}^{M'} \sum_{k=1}^{d} X_k(t+s-1)\mathbf{E}_k^j(s)$.
- **Reconstruction:** $R_k^j(t) = \frac{1}{M_t} \sum_{s=L_t}^{U_t} \xi_j(t-s+1)\mathbf{E}_k^j(s)$, $1 \le s \le M'$,
- **frequency-domain** DAHD formulation by eigendecomposition of Hermitian cross-spectral matrix $\mathfrak{S}(f)$: $\mathfrak{S}_{p,q} = \widehat{\rho^{p,q}}(f)$ – Fourier transform of cross-correlation sequence $\rho^{p,q}$ - removes a bias and makes it efficient in high-dimensions [Kondrashov et al. 2020, *Chaos*].

Identification of coherent patterns from noisy data



- We consider here synthetic example of several propagating ocean waves:

$$u_n(x,y,t) = A_n cos(k_n x/L_x + l_n y/L_y + \omega_n t)$$

where A_n are the random weights, while frequency ω_n and wave numbers (k_n, l_n) obey Rossby dispersion relation:

$$\omega_n = -\frac{\beta k_n}{k_n^2 + l_n^2 + R^{-2}}$$

- Total dataset with $N = 999$ points in time, $Nx = Ny = 64$ is the sum of the waves and large-amplitude red noise.

Kondrashov, D., Ryzhov, E.A. and P.S. Berloff, 2020: Data-adaptive harmonic analysis of oceanic waves and turbulent flows, *Chaos*, 30, 061105, doi:10.1063/5.0012077.

Identification of coherent waves from noisy data

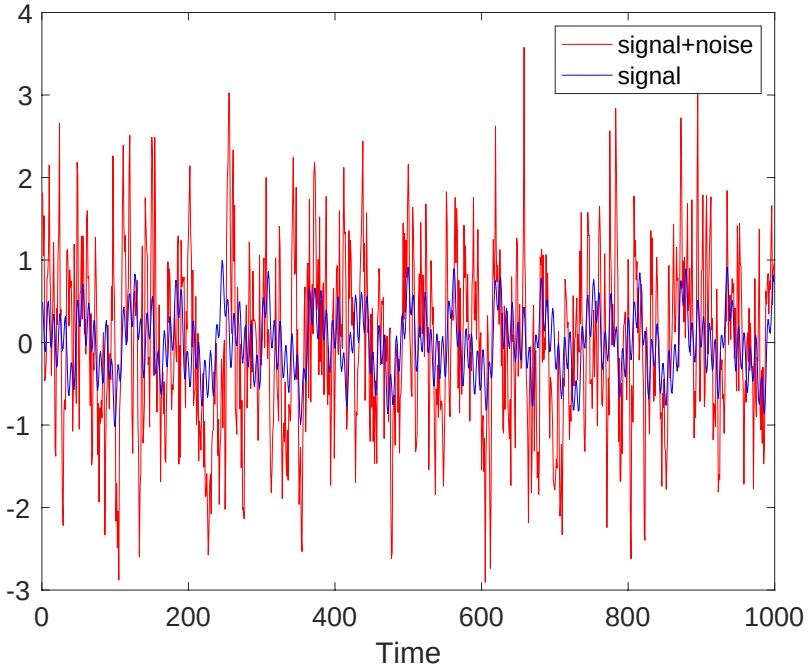


Figure: Time series of the coherent wavy signal and full data with imposed red noise at selected (x,y) point.

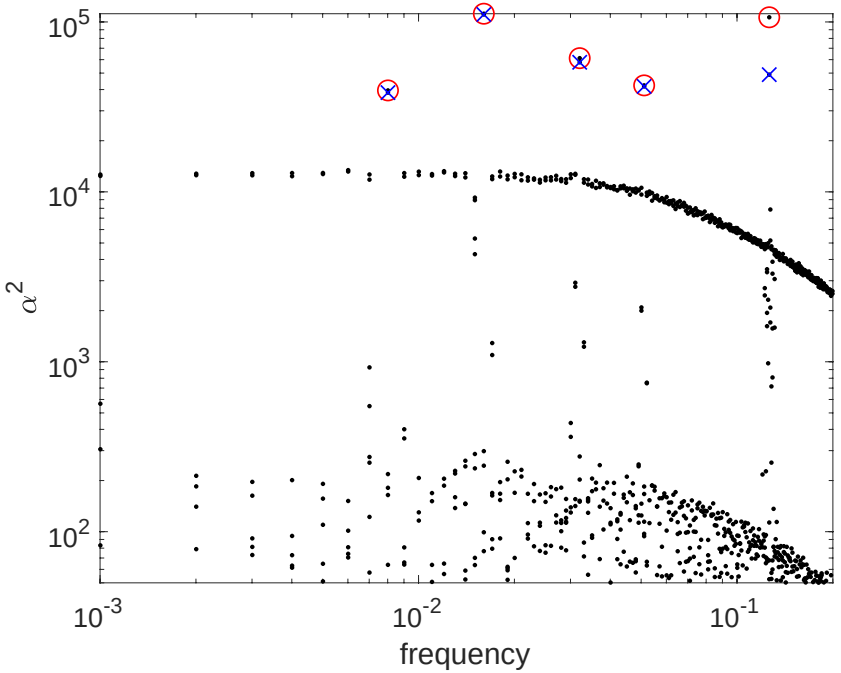
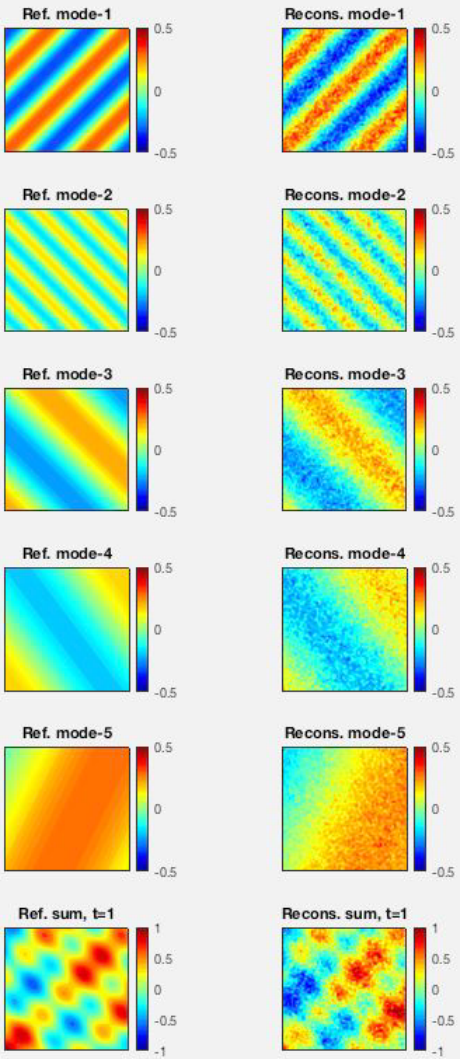


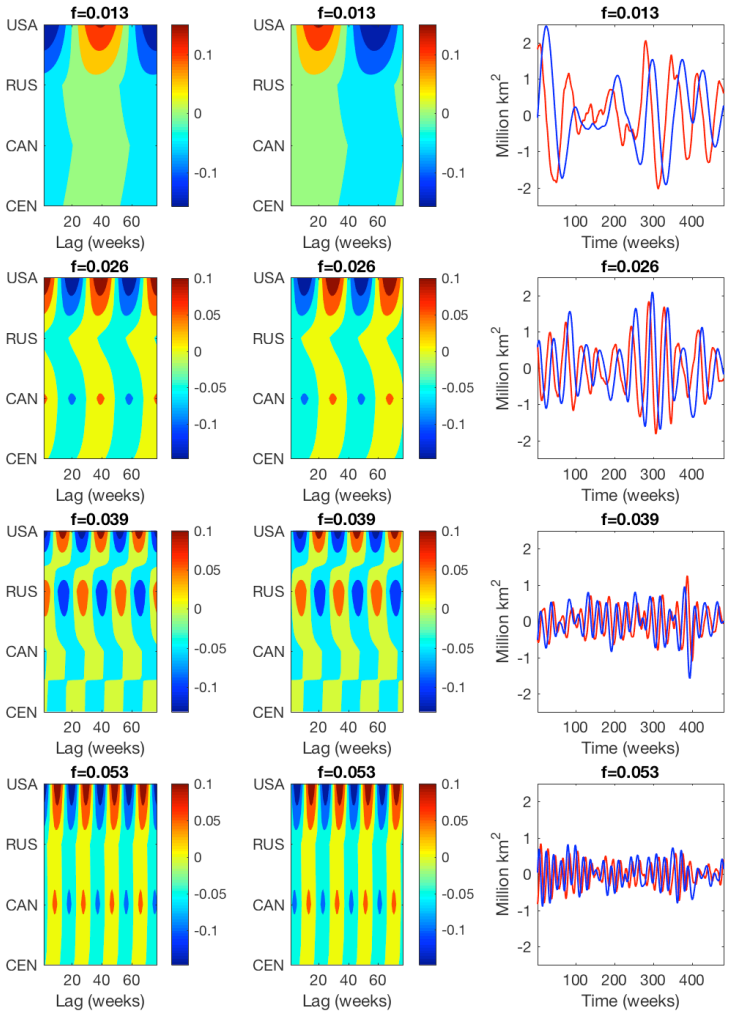
Figure: DAHD energy spectrum: red circles and blue crosses – modes associated with largest eigenvalues at reference wave frequencies

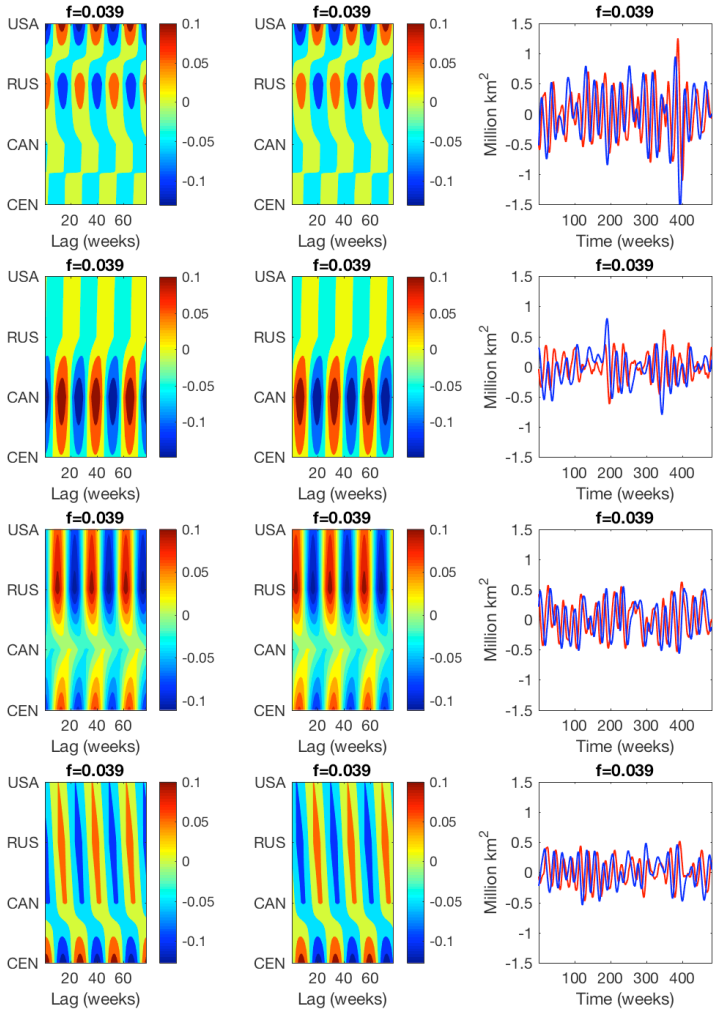
Identification of coherent waves from noisy data



Accurate reconstruction of reference wave patterns is obtained by using modes associated with DAHD spectral peaks

DAHD prediction of Regional Arctic SIE





- Pairs of **DAHD modes** and **DAHD coefficients** are **always** in phase-quadrature: **left panel** – **leading spectral pair** at a given frequency; **right panel** – **top-to-bottom spectral pairs** at given frequency. x-axis – time, y-axis – regions.
- Channel-wise phase and amplitude modulations of the modes are **data-adaptive**!

- If $(x(t), y(t))$ are **pair of DAHD coefficients – narrow-band oscillatory time series** in phase-quadrature and associated with a dominant frequency f, **Stuart-Landau (SL) models** with additive noise are generic class of models to capture **(i) the frequency** f and **(ii) amplitude modulations**:

$$\dot{z} = (\mu + i\gamma)z - (1 + i\beta)|z|^2 z + \eta_t, \quad z \in \mathbb{C},$$

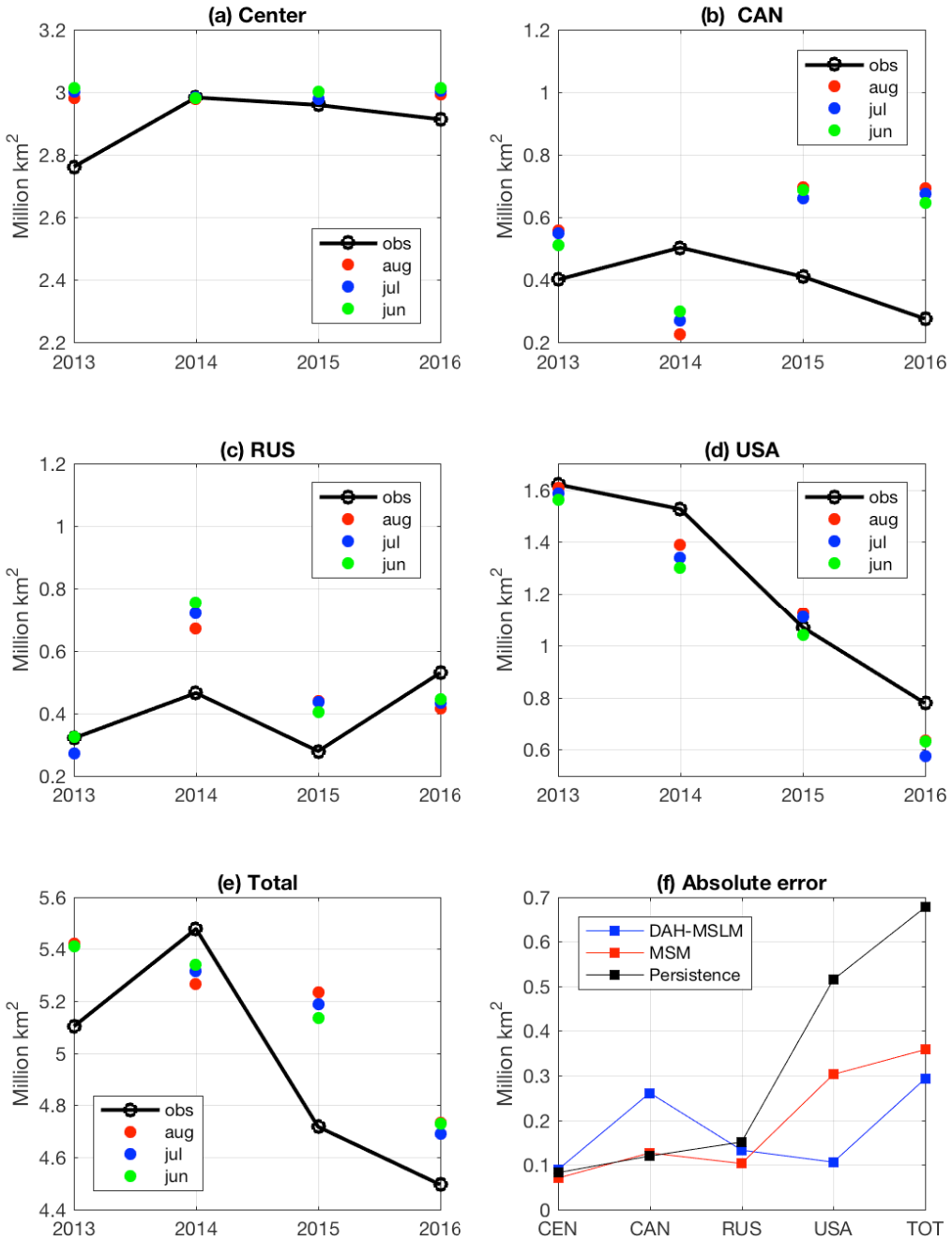
- μ, γ, β - real parameters, η_t - **"noise"** – are estimated from time history of $z(t) = x + iy$.
- System of coupled SL oscillators to model all DAHD pairs (x_j, y_j) at a given frequency f:

$$\dot{x_j} = \beta_j(f)x_j - \alpha_j(f)y_j + \sigma_j(f)x_j(x_j^2 + y_j^2) + \sum_{i \neq j}^{N} a_{ij}(f)x_i + \sum_{i \neq j}^{N} b_{ij}(f)y_i + \eta_j^x,$$

$$\dot{y_j} = \alpha_j(f)x_j + \beta_j(f)y_j + \sigma_j(f)y_j(x_j^2 + y_j^2) + \sum_{i \neq j}^{N} c_{ij}(f)x_i + \sum_{i \neq j}^{N} d_{ij}(f)y_i + \eta_j^y,$$

- Diagnostic by Ruelle–Pollicott resonances justifies use of Stuart-Landau oscillators [Kondrashov, Chekroun and Berloff, *Fluids*, 2018].
- The **model coefficients are estimated** *in parallel* **for each frequency**, by successive regressions with **linear constraints** to impose SL structure, i.e. $\beta_j(f), \alpha_j(f), \sigma_j(f)$.
- Stochastic models are run *in parallel* and coupled across the frequencies by the noise.

DAHD Forecasting of September Sea Ice Extent



- 2013–2016 regional retrospective forecasts ("no look-ahead") of **September SIE** (stochastic ensemble mean) from June (green), July (blue) and August (red).
- Consistent June-to-August outlooks emphasize stable predictive content captured; **ensemble spread (std.dev)** is also fairly small: ≈ 0.1 (M km^2)
- USA region is predicted best and it leads to skillful forecast over the whole Arctic.
- CAN sector is most difficult to predict and may be due to being inland.
- Including some key ocean-atmosphere variables (air temperature, sea-level pressure,...), sea ice thickness could help with **spring predictability barrier**.

Conclusions

- Data-driven stochastic-dynamic climate modeling has undergone remarkable development in recent years with deep learning and neural nets offering new exciting opportunities.
- Such models continue to be useful as statistical benchmarks, also for for empirical probabilistic diagnostics of internal variability in GCMs [Chen et al., 2016, 2017].

1. Kondrashov, D., M. D. Chekroun, and M. Ghil, 2018: Data-adaptive harmonic decomposition and prediction of Arctic sea ice extent, *Dynamics and Statistics of the Climate System*, doi:10.1093/climsys/dzy001.
2. M.D. Chekroun and D. Kondrashov, 2017: Data-adaptive Harmonic Spectra and Stochastic-dynamic Inverse Stuart-Landau Models, *Chaos*, **27**, 093–110.
3. D. Kondrashov, M.D. Chekroun, and M. Ghil, 2015: Data-driven non-Markovian closure models, *Physica D*, 297, 33–55.
4. S. Kravtsov, D. Kondrashov, and M. Ghil, 2009: Empirical model reduction and the modeling hierarchy in climate dynamics, *Stochastic Physics and Climate Modeling* (T. N. Palmer and P. Williams, eds.), Cambridge Univ. Press, pp. 35–72.
5. D. Kondrashov, S. Kravtsov, A. W. Robertson, and M. Ghil, 2005: A hierarchy of data-based ENSO models, *J. Climate*, **18**, 4425–4444.
6. Chen, C., M. Cane, N. Henderson, D. Lee, D. Chapman, D. Kondrashov, M. Chekroun, 2016: Diversity, nonlinearity, seasonality and memory effect in ENSO simulation and prediction using empirical model reduction, *J. Climate*, **29**, 1809–1830.
7. Chen, C., M. Cane, A. Wittenberg, D. Chen, 2017: ENSO in the CMIP5 Simulations: Life Cycles, Diversity, and Responses to Climate Change, *J. Climate*, **30**, 775–801.

DAHD and Stochastic Modeling toolbox: http://research.atmos.ucla.edu/tcd/dkondras/

- Ocean modeling is a challenging problem, because model solutions are usually critically sensitive to the spatial numerical grid resolutions.
- If the solution lacks the dynamics produced by unresolved sub-grid processes, then the resolved processes are also affected due to the involved nonlinearity.
- A notable example is oceanic mesoscale eddies which affect the large-scale circulation (Berloff and McWilliams, 1999; Kravtsov et al., 2006; Berloff et al., 2007; Kirtman et al. 2012, Shevchenko et al., 2016).
- Therefore, it is important to properly account for the dynamical eddy effects in non-eddy-resolving models.
- It can be done by using information inferred either from high-resolution eddy-resolving model simulation (this study) (or observations).

Idealized Midlatitude Ocean Model

Classical double-gyre ocean model setting, where the governing equations describe evolution of the quasi-geostrophic (QG) potential vorticity (PV) in 3 stacked layers ($i = 1..3$ from top to bottom) with densities ρ_i and heights H_i, and forced by the wind stress W in the upper layer:

$$\frac{\partial q_i}{\partial t} + J(\psi_i, q_i) + \beta \frac{\partial \psi_i}{\partial x} = \frac{W(x,y)}{\rho_i H_i}\delta_{1i} - \gamma \Delta \psi_i \delta_{3i} + \nu \Delta^2 \psi_i \,,$$

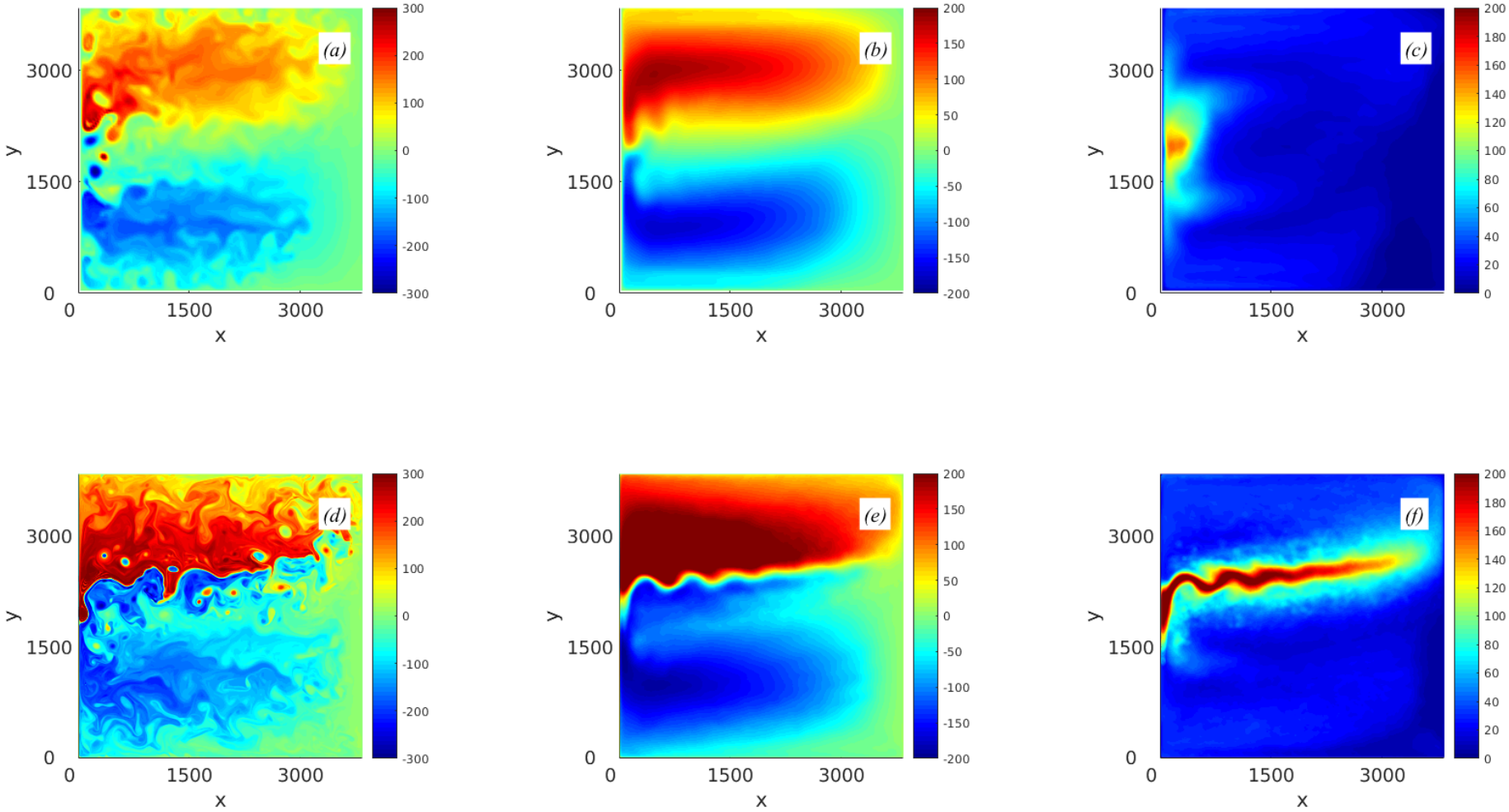
The PV anomaly is inverted to obtain the streamfunctions, according to:

$$\begin{aligned} q_1 &= \Delta\psi_1 + S_1(\psi_2 - \psi_1)\,, \\ q_2 &= \Delta\psi_2 + S_{21}(\psi_1 - \psi_1) + S_{22}(\psi_3 - \psi_2)\,, \\ q_3 &= \Delta\psi_3 + S_3(\psi_2 - \psi_3)\,, \end{aligned}$$

- The model produces a classical double-gyre flow pattern in eddy-resolving high-resolution solution (512x512), characterized by a well-developed and turbulent eastward jet extension of the western boundary currents with its adjacent recirculation zones.
- Non-eddy resolving low-resolution solution (128x128) lacks these features.

Statistical differences between the low- and high-resolution PV solutions

Instantaneous | Time-mean | Std.dev



- A well-developed meandering eastward jet and its ambient eddy field are present in the high-resolution flow (lower panels) and absent in the low-resolution solution (upper panels).
- high-res - pronounced interdecadal temporal variability, entirely absent in low-res.

Eddy forcing inference from the high-resolution solution

- Our main goal is to dynamically augment the low-resolution model by the missing dynamical information of small-scale eddies, so that its solution recovers the key features of the high-resolution solution: **the eastward jet extension and the interdecadal LFV**.
- We assume that low-resolution model are capable of partially resolving slowly varying large-scale dynamics but fails to resolve more transient small-scale dynamics, whereas high-resolution models resolve everything.
- We assume that the low- and high-resolution large-scale dynamics are similar and implement a scale decomposition of the high-resolution Ψ and Q:

$$\Psi = \overline{\Psi} + \Psi', \;\; Q = \overline{Q} + Q',$$

where $(\overline{\Psi}, \overline{Q})$ are large-scale components, and (Ψ', Q') are small-scale (eddy) components. By substitution, we obtain equation that couples the large-scale and eddy dynamics:

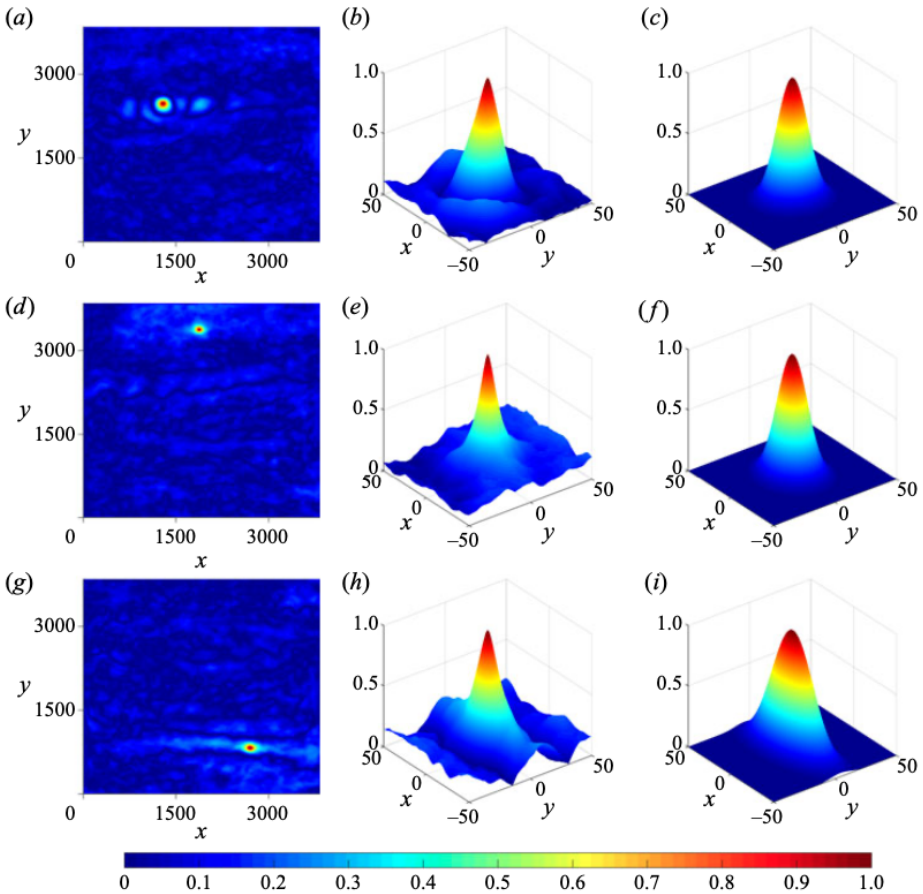
$$\frac{\partial \overline{Q}}{\partial t} + J(\overline{\Psi}, \overline{Q}) \approx \mathcal{F}\left(\overline{\Psi}, \overline{Q}, \Psi', Q'\right) + \mathcal{H}(\overline{\Psi}, \overline{Q}),$$

where the operator $\mathcal{H}$ contains all the nonconservative terms involving only the large-scale components, whereas

$$\mathcal{F} = -\left(J(\overline{\Psi}, Q') + J(\Psi', \overline{Q}) + J(\Psi', Q')\right) = -\left(J(\Psi, Q) - J(\overline{\Psi}, \overline{Q})\right) \;,$$

is the eddy forcing exerted by the nonlinear coupling between the eddy and large-scale flow components, as well as by the eddy nonlinearity.

Correlation-based decomposition (CBD)



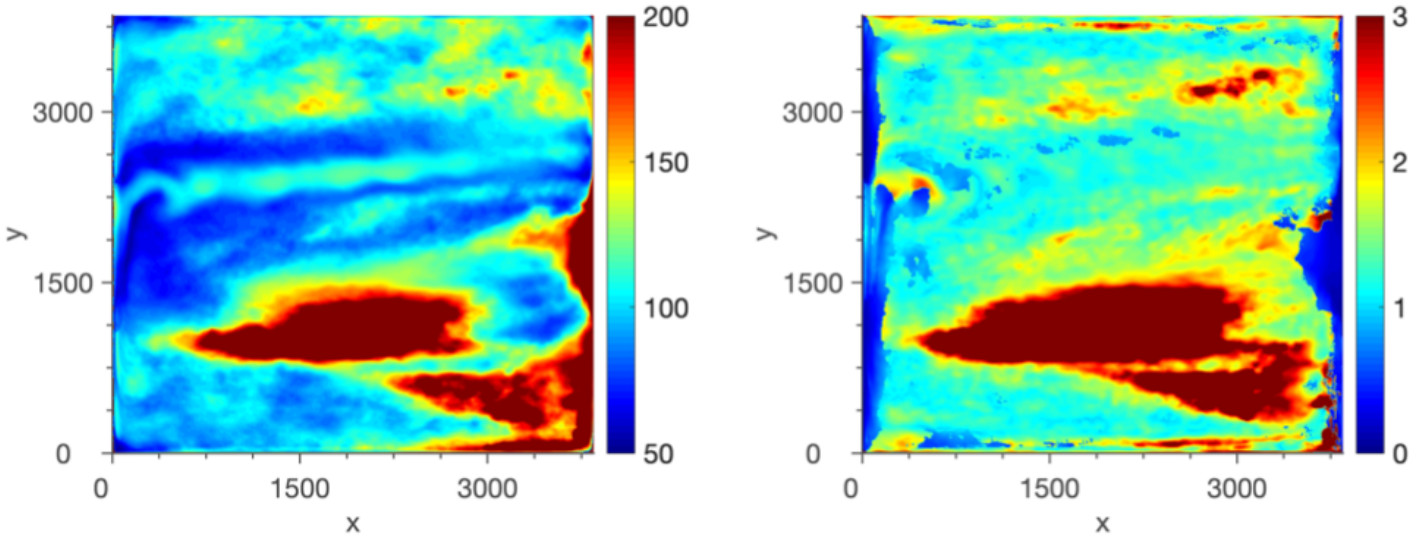
- Filter high-res solution at each location according **to the local length scale of the spatial correlation**, as opposed to using a **fixed kernel size (LES-type)**, and thus extract the large-scale flow component; the small-scale component is obtained as the residual.
- $|C(q_1(\mathbf{x}_0, \mathbf{x})|$ contours (column 1), zoomed-in 3D view around the reference location $\mathbf{x}_0$ (column 2), and the fitted Gaussian function (column 3) for three reference locations.

Gaussian function $f(x, y; x_0, y_0) = \exp(-X^2/a^2)\exp(-Y^2/b^2)$ where X and Y are the rotated and translated coordinate axes:

$$\begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x - x_0 \\ y - y_0 \, , \end{bmatrix}$$

N. Argawal, Ryzhov, E.A., Kondrashov, D., and P.S. Berloff, 2021, *Journal of Fluid Mechanics*,

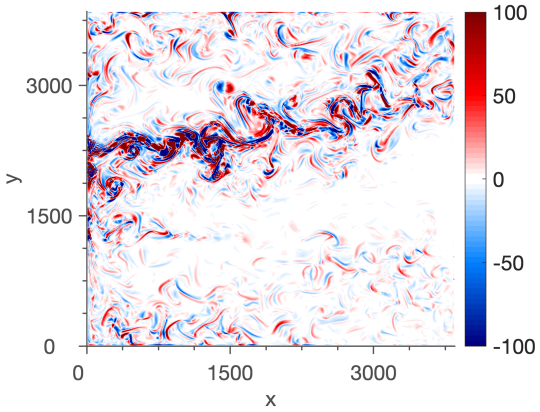
Correlation-based decomposition (CBD)



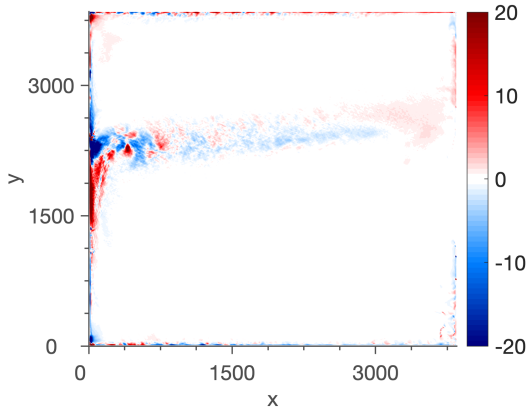
- Maps of correlation length scale $\mathcal{L} = \sqrt{ab}$ (left column) and the correlation anisotropy $\mathcal{A}(\mathbf{x}) = a/b$ (right column) for the upper layer in the ocean model.
- **The length scale and anisotropy maps reveal the complex and multiscale nature of the flow**, and suggest that a fixed-size kernel can substantially underfilter or overfilter eddies depending on the location; therefore, a **differential filter size over the domain is justified**!
- Each spatial location of the PV fields is filtered by using normalized Gaussian kernel with the estimated parameters a, b, and θ: $\overline{q}(x_i, y_j, t) = \sum_{i'} \sum_{j'} G(x_{i'}, y_{j'}; x_i, y_j) q(x_{i'}, y_{j'}, t)$; $G(x_{i'}, y_{j'}; x_i, y_j) = f(x_{i'}, y_{j'}; x_i, y_j)/A_f$ and $A_f = \sum_{i'} \sum_{j'} f(x_{i'}, y_{j'}; x_i, y_j) \equiv \pi ab$ is the sum of the Gaussian weights.

CBD eddy forcing

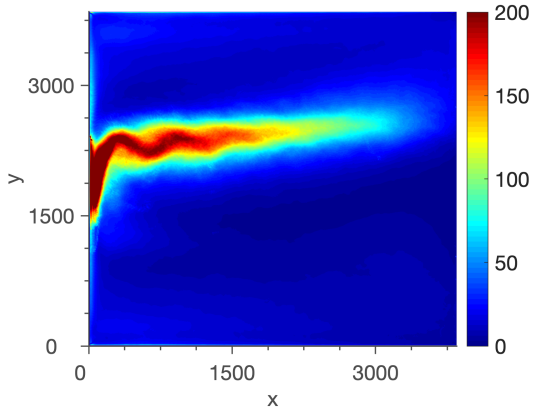
Instantaneous



Time-mean

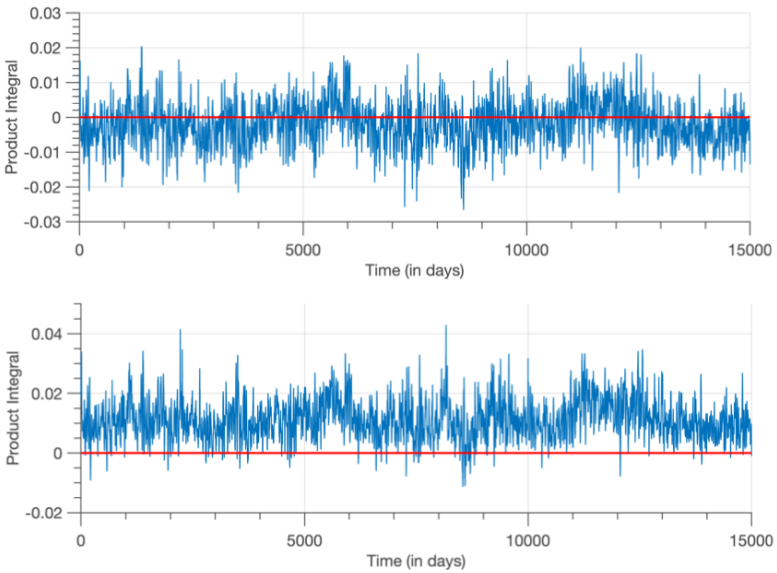


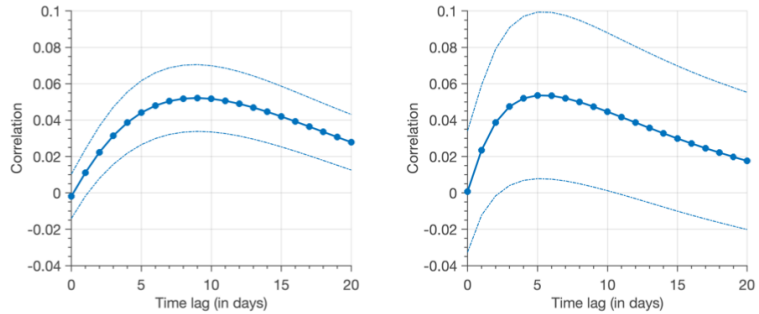
Std.dev



- The resulting eddy forcing $\mathcal{F}$ is most intense around the eastward jet, while its time mean is small relative to the fluctuations
- Spatial pattern is characterized by complex small-scale transient features of backscatter

dynamical diagnosis of eddy forcing - product integral (PI)





- The dynamical impacts of the resulting eddy forcing on the large-scale flow in terms of their mutual time-lagged spatial correlations, formulated as product integral (PI) characteristics.

$$I(t,\tau) = \frac{1}{A}\iint_{\Omega} \overline{q}_1(x,y;t+\tau) \odot \mathcal{F}_1(x,y;t)\,\mathrm{d}x\,\mathrm{d}y\,, \tag{1}$$

- PI temporal statistics uncover robust causality between the eddy forcing and the induced large-scale potential vorticity anomalies – eddy backscatter.
- The results also prove the significance of the transient eddy forcing and the time lag dependence of the eddy backscatter, that are to be considered by parametrization schemes.

Augmentation of the low-resolution numerical model

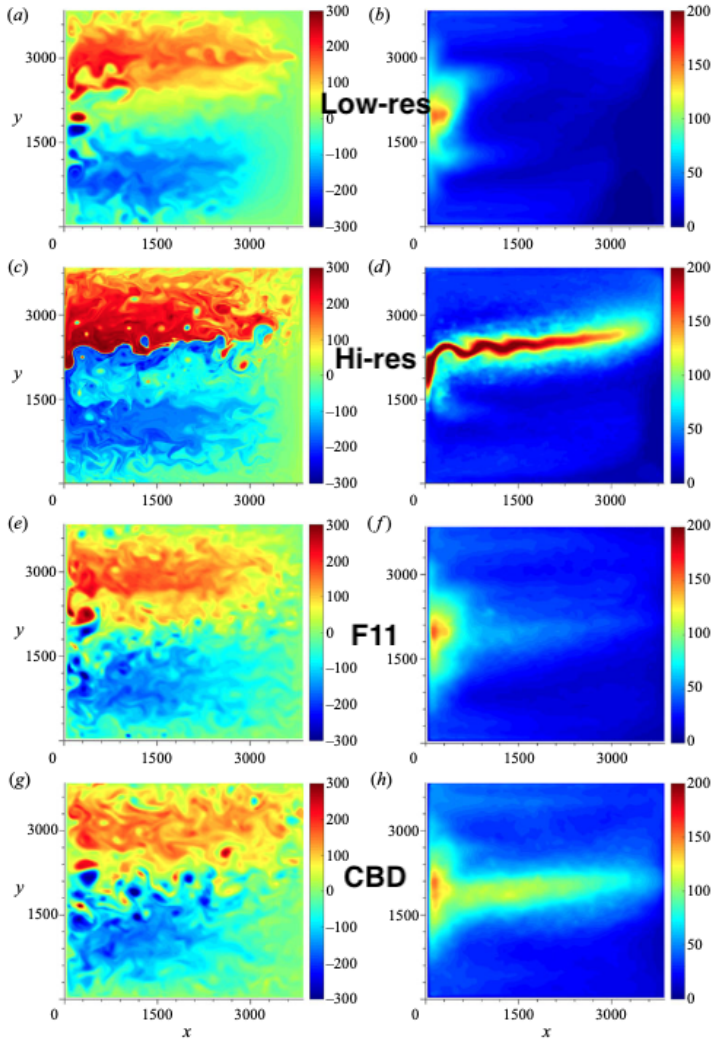
- Update eddy forcing during online integration of the low-res model:

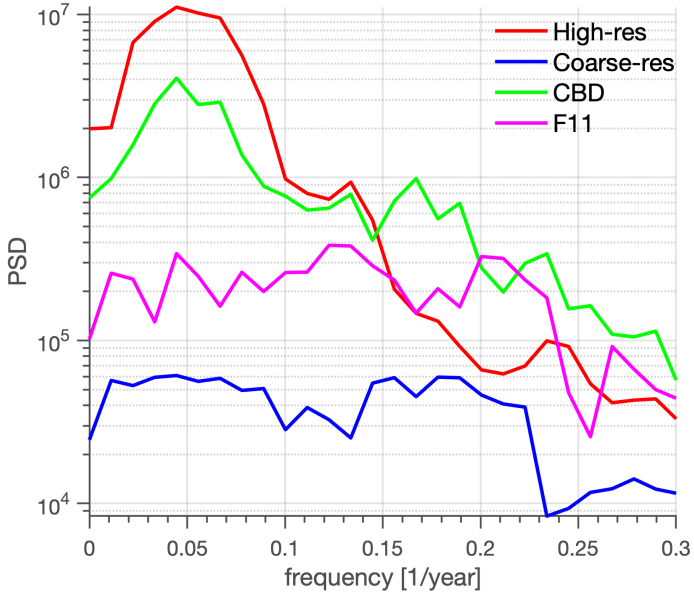
$$\frac{\partial Q_i}{\partial t} + J(\Psi_i, Q_i) = \mathcal{F}_i\left(\Psi_i, Q_i, \Psi_i', Q_i'\right) + \mathcal{H}_i(\Psi_i, Q_i), \tag{2}$$

  where the **small-scale fields** Ψ_i', Q_i' are either supplied from hi-res simulation or obtained by **stochastic ML emulator**, while **prognostic low-resolution variables** Ψ_i, Q_i are updated during numerical integration.
- Linear stochastic emulator
- Nonlinear frequency-ranked stochastic DAHD emulator

Ryzhov, E.A., D. Kondrashov, N. Agarwal, J.C.McWilliams and P.S. Berloff, 2020:
On data-driven induction of the low-frequency variability in a coarse-resolution ocean model,
Ocean Modelling, 153, 101664, doi:10.1016/j.ocemod.2020.101664.

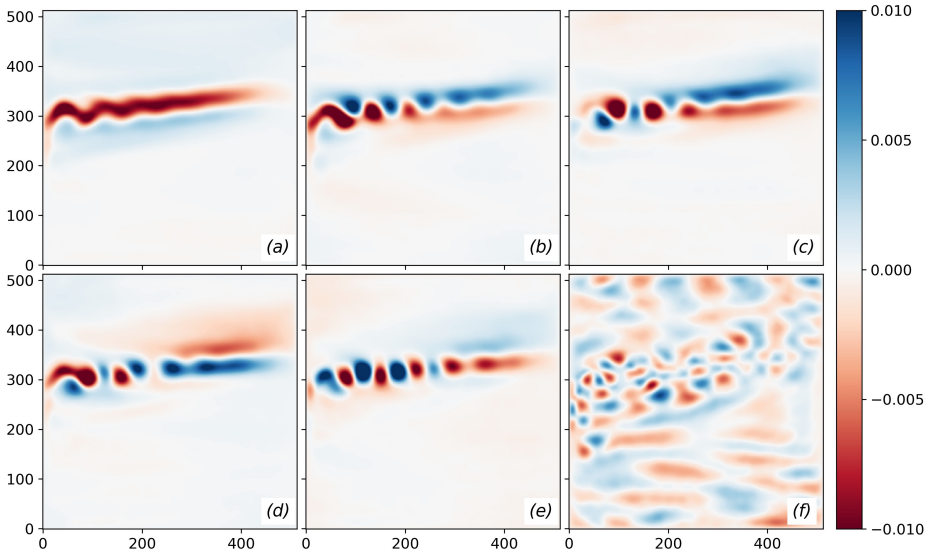
Augmentation of the low-resolution numerical model

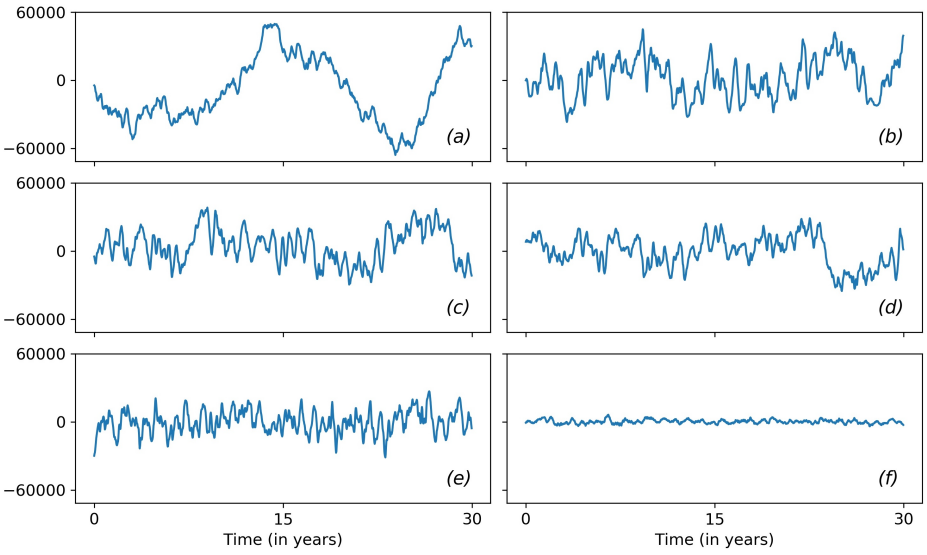




CBD-augmented solution recovers the missing eastward jet extension, the eddies around it, and the interdecadal low-frequency variability, **much better** than when using fixed-size filter (F11).

Methods comparison for reduced-order emulation of ocean circulation





- Systematic comparison of different methods: linear regression, standalone neural nets (ANN, LSTM) and their various hybrid stochastic formulations.
- Subspace of leading PCs from PCA (thus no CNN).
- 150 PCs of upper-layer from high-res double-gyre ocean model simulation
- Short-term prediction
- Long-term simulation: summary statistics, biases in climatology and temporal variability

N. Argawal, Kondrashov, D., Dueben, P., Ryzhov, E.A., and P.S. Berloff, 2021:
A comparison of data-driven approaches to build low-dimensional ocean models,
Journal of Advances in Modelling Earth Systems, doi:10.1029/2021MS002537.

Methods comparison for reduced-order emulation of ocean circulation

Table 2
An Overview of All the Methods Considered in This Work

Method	Abbreviation	Section	Input	Output	Cost function	Memory	State-dependent noise
Linear Regression	LR	3.1	State	tendency	OLS	NA	NA
LR + Additive White Noise	LR-AWN	3.2	State	tendency	OLS	×	×
Multi-level Linear Regression	ML-LR	3.3	state, LR residuals	tendency	OLS	✓	✓
Artificial Neural Network (+White Noise)	ANN (- AWN)	3.4	State	State	MAE	×	×
Long Short Term Memory (+White Noise)	LSTM (- AWN)	3.5	State	State	MAE	✓	×
LR + ANN Hybrid (+White Noise)	LR-ANN (-AWN)	3.6	state, LR residuals	tendency	OLS, MAE	×	✓
LR + LSTM Hybrid (+White Noise)	LR-LSTM (-AWN)	3.6	state, LR residuals	tendency	OLS, MAE	✓	✓

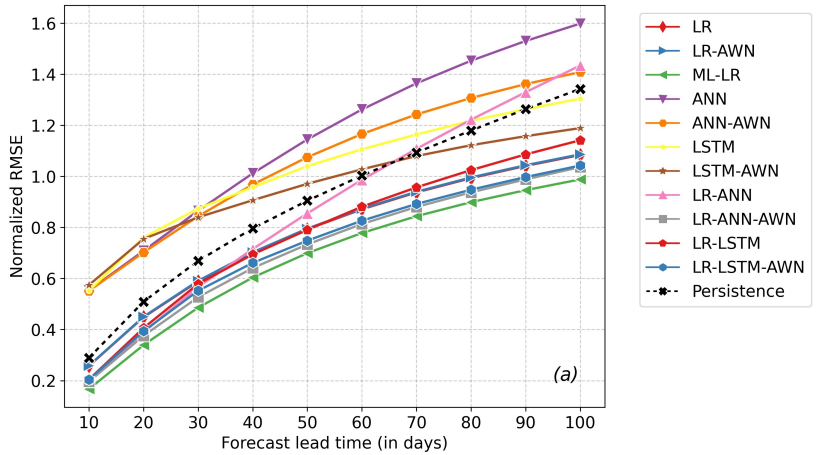
- Hybrid model for PC's tendencies with linear regression (LR) core, while deep-learning is used as nonlinear correction, state-dependent noise and memory effects in LR residuals:

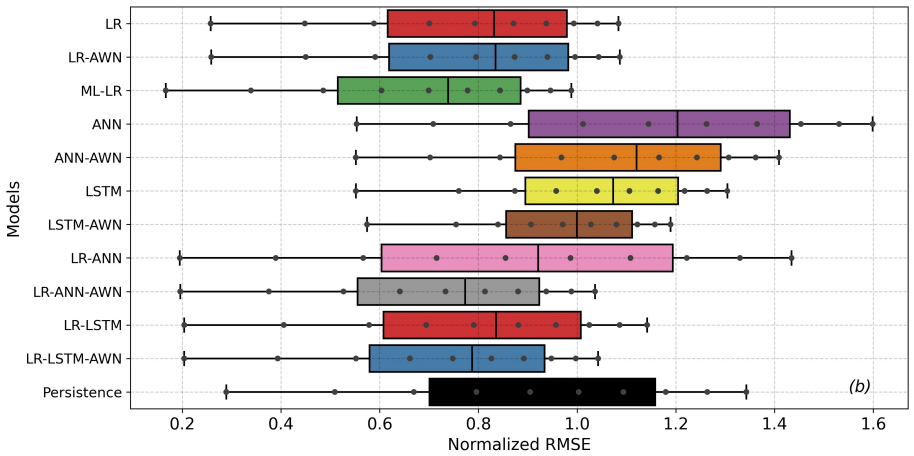
$$\begin{aligned} \mathbf{u_{k+1}} - \mathbf{u_k} &= \mathbf{A u_k}\delta t + \mathbf{r}_k^0 \delta t, \\ \mathbf{r}_{k+1} &= \Phi(\mathbf{r}_k, \mathbf{u}_k) + \boldsymbol{\xi}, \end{aligned}$$

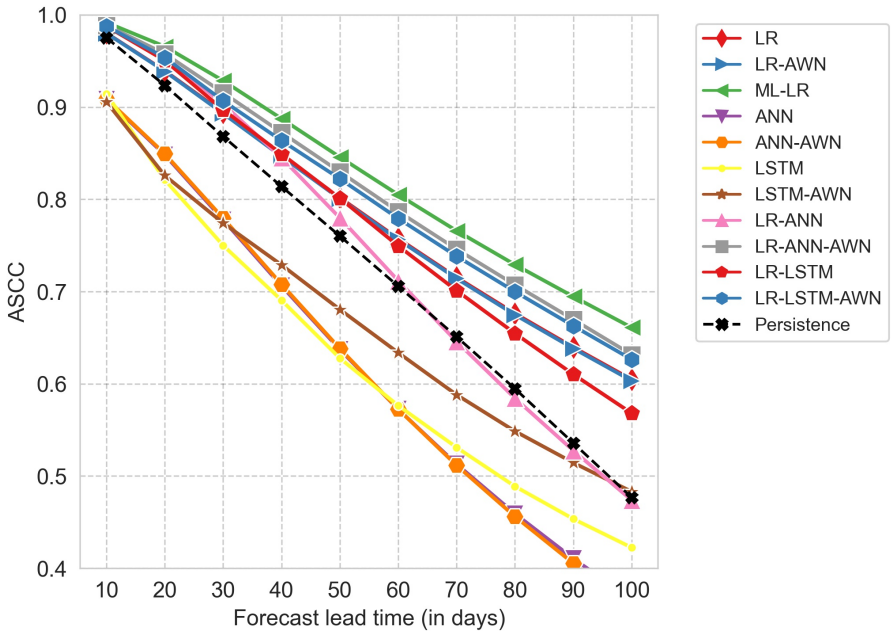
  where Φ is LSTM or ANN, and $\boldsymbol{\xi} \sim \mathcal{N}(0, \mathbf{Q}\mathbf{Q}^T)$ approximates deep-learning residual.
- Muti-level LR model for tendencies:

$$\begin{aligned} \mathbf{u_{k+1}} - \mathbf{u_k} &= \mathbf{A u_k}\delta t + \mathbf{r}_k^0 \delta t, \\ \mathbf{r}_{k+1}^{m-1} - \mathbf{r}_k^{m-1} &= \mathbf{L}^m \left[\mathbf{u_k}, \mathbf{r}_k^0, ..., \mathbf{r}_k^{m-1}\right] \delta t + \mathbf{r}_k^m \delta t,\ 1 \le m \le M, \end{aligned}$$

Short-term prediction

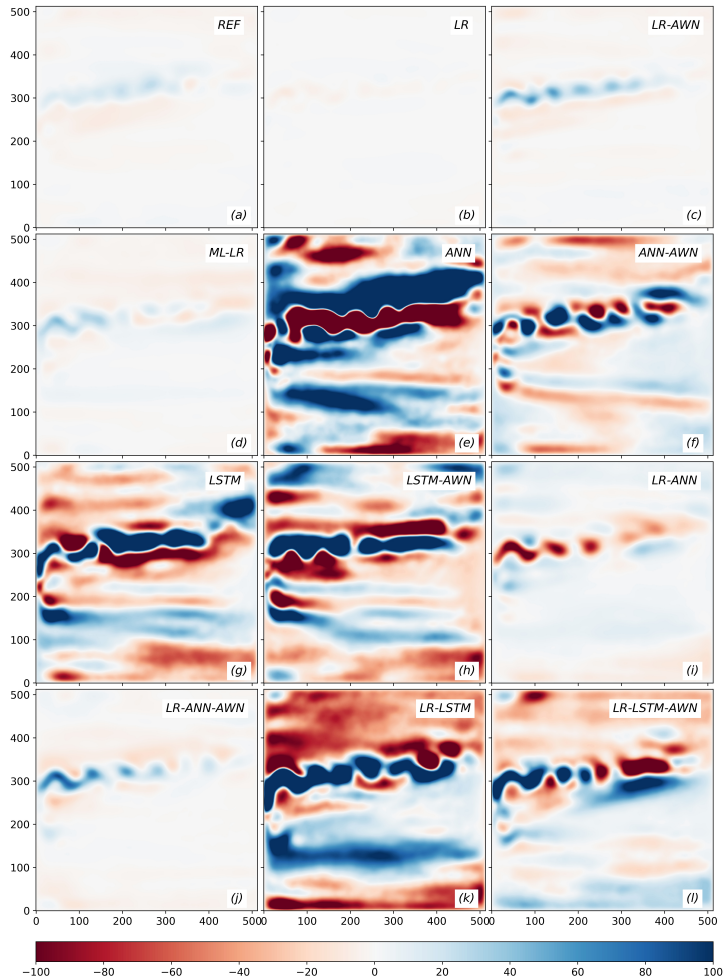


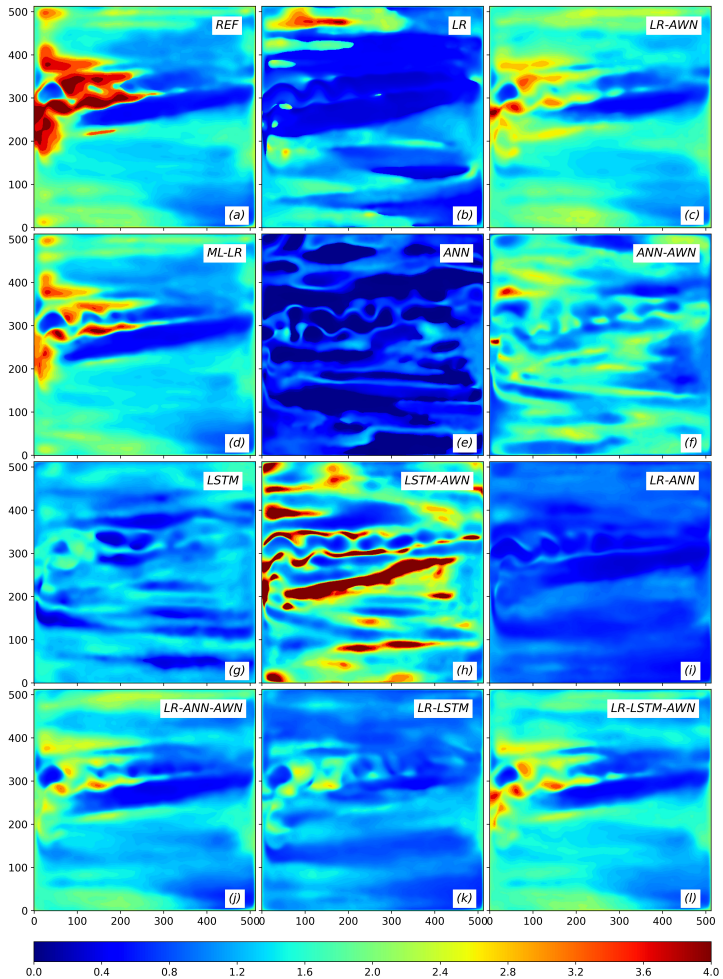




- skill: RMSE and ACC computed in gridded space.
- ML-LR forecasts are the best, followed by the stochastic hybrid models.

Long-term statistics





- gridded skill of the emulated long-timescale solutions for streamfunction anomalies: climatology bias and frequency map for spectral characteristics – 1/(decorrelation time).
- overall ML-LR is the best, followed by the stochastic hybrid models.

Conclusions -Part 2

- Correlation-based flow decomposition is an attractive alternative to commonly-used fixed-size filter for the purposes of subgrid-scale parameterizations in ocean models.
- Multi-level LR stochastic models with memory effects, and hybrid models with linear dynamical core augmented by additive stochastic terms learned via deep learning by LSTM or ANN, are more practical, accurate, and cost-effective option for emulation of complex multicsale oceanic flow, than standalone deep-learning solutions.

1. N. Argawal, Kondrashov, D., Dueben, P., Ryzhov, E.A., and P.S. Berloff, 2021: A comparison of data-driven approaches to build low-dimensional ocean models, *Journal of Advances in Modelling Earth Systems*, doi:10.1029/2021MS002537.
2. N. Argawal, Ryzhov, E.A., Kondrashov, D., and P.S. Berloff, 2021: Correlation-based flow decomposition and statistical analysis of the eddy forcing, *Journal of Fluid Mechanics*, 924, A5, doi:10.1017/jfm.2021.604.
3. Kondrashov, D., Ryzhov, E.A. and P.S. Berloff, 2020: Data-adaptive harmonic analysis of oceanic waves and turbulent flows, *Chaos*, 30, 061105, doi:10.1063/5.0012077.
4. Ryzhov, E.A., D. Kondrashov, N. Agarwal, J.C. McWilliams, and P.S. Berloff, 2020: On data-driven induction of the low-frequency variability in a coarse-resolution ocean model, *Ocean Modelling*, 101664.
5. Ryzhov, E.A., D. Kondrashov, N. Agarwal, and P.S. Berloff, 2019: On data-driven augmentation of low-resolution ocean model dynamics, *Ocean Modelling*, 142, 101464.
6. Kondrashov, D., M. D. Chekroun and P. Berloff, 2018: Multiscale Stuart-Landau Emulators: Application to Wind-Driven Ocean Gyres, *Fluids*, 3(1), 21.