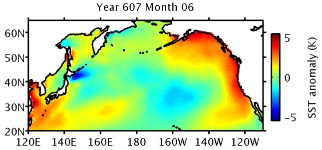
Kernel methods for nonparametric analog forecasting

Dimitris Giannakis
Center for Atmosphere Ocean Science
Courant Institute of Mathematical Sciences

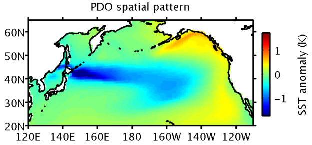
Stochastic Weather Generators
Vannes, May 19 2016

Motivating example

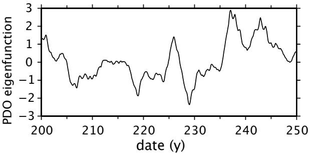


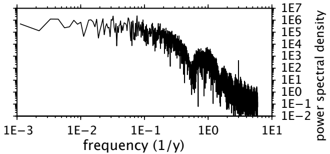
North Pacific sea surface temperature

- Spatiotemporal field with power on subseasonal to multidecadal timescales



Given past SST observations, predict the evolution of the **Pacific decadal oscillation**

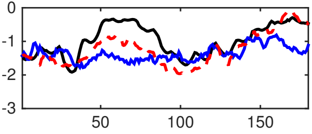




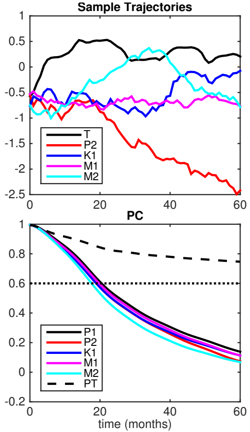
Why nonparametric forecasting?

In many low-order modeling scenarios, it is difficult to construct a parametric model that fits the training data well and also has high forecast skill

PDO fit in training phase



Hindcast phase

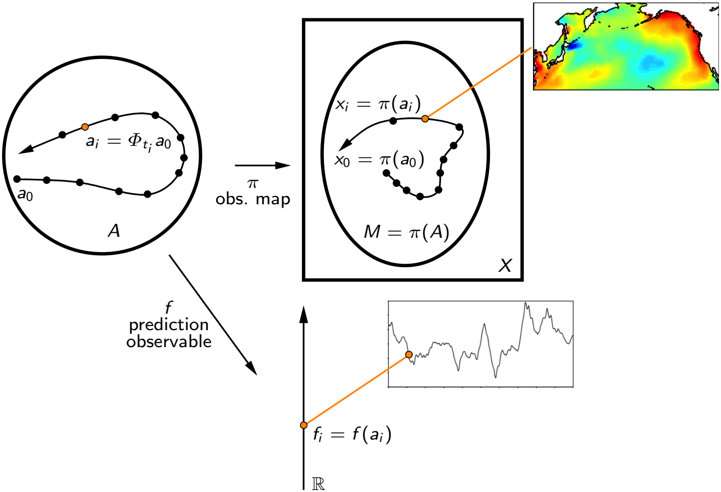


- Stationary and nonstationary autoregressive models fit the PDO training time series well, but fail to beat persistence forecast in hindcast phase (Comeau et al. 2016)

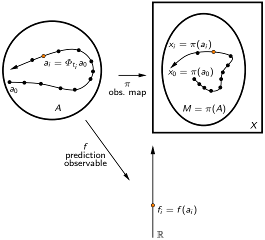
Outline

1. Conventional analog forecasting
2. Basic kernel analog forecasting
3. KAF with Nyström extension
4. Dynamics-aware kernels

Collaborators. Romeo Alexander, Mitch Bushuk, Darin Comeau, Andy Majda, Joanna Slawinska, Eniko Szekely, Jane Zhao

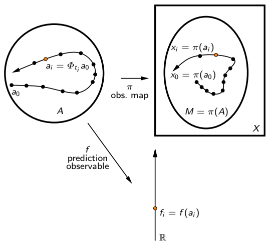

$a_i = \Phi_{t_i} a_0$
a_0
A
π
obs. map
$x_i = \pi(a_i)$
$x_0 = \pi(a_0)$
$M = \pi(A)$
X
f
prediction
observable
$f_i = f(a_i)$
$\mathbb{R}$

Problem setup



- Ergodic dynamical system $(A, \mathcal{A}, \Phi_t, \alpha)$ with evolution map Φ_t and invariant prob. measure α
- Measurable observation map $\pi : A \mapsto X$ into a metric space X (space of initial data)
- Square-integrable prediction observable $f \in L^2(A, \alpha)$
- Koopman operators $U_t : L^2(A, \alpha) \mapsto L^2(A, \alpha)$, $U_t f(a) = f(\Phi_t(a))$

Problem setup



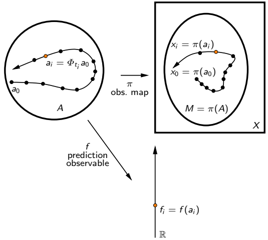
- Ergodic dynamical system $(A, \mathcal{A}, \Phi_t, \alpha)$ with evolution map Φ_t and invariant prob. measure α
- Measurable observation map $\pi : A \mapsto X$ into a metric space X (space of initial data)
- Square-integrable prediction observable $f \in L^2(A, \alpha)$
- Koopman operators $U_t : L^2(A, \alpha) \mapsto L^2(A, \alpha)$, $U_t f(a) = f(\Phi_t(a))$

Objective. Given time-ordered pairs

$$\{(x_0, f_0), (x_1, f_1), \ldots, (x_{N-1}, f_{N-1})\},$$
$$x_i = \pi(a_i), \quad f_i = f(a_i), \quad a_i = \Phi_{t_i}(a_0), \quad t_i = (i-1)\tau,$$

construct a function $\hat{F}_t : X \mapsto R$ that predicts f over lead time t

Problem setup



- Ergodic dynamical system $(A, \mathcal{A}, \Phi_t, \alpha)$ with evolution map Φ_t and invariant prob. measure α
- Measurable observation map $\pi : A \mapsto X$ into a metric space X (space of initial data)
- Square-integrable prediction observable $f \in L^2(A, \alpha)$
- Koopman operators $U_t : L^2(A, \alpha) \mapsto L^2(A, \alpha)$, $U_t f(a) = f(\Phi_t(a))$

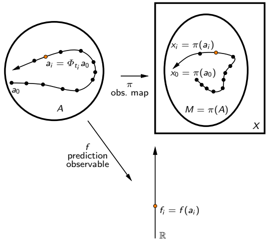
Objective. Given time-ordered pairs

$$\{(x_0, f_0), (x_1, f_1), \dots, (x_{N-1}, f_{N-1})\},$$
$$x_i = \pi(a_i), \quad f_i = f(a_i), \quad a_i = \Phi_{t_i}(a_0), \quad t_i = (i-1)\tau,$$

construct a function $\hat{F}_t : X \mapsto R$ that predicts f over lead time t

- A "perfect" forecast has $\hat{F}_t(\pi(b)) = U_t f(b)$ for all $b \in A$.

Analog prediction (Lorenz 1969)



- **Training data:** Time-ordered pairs $\{(x_i, f_i)\}_{i=0}^{N-1}$
- **Input:** Previously unseen sample $y \in X$, distance function D
- **Output:** Forecast $\hat{F}_t(y)$ of f at lead $t = q\tau$, $q \in \mathbb{N}$

Model-free, two-step technique:

1. **Analog identification:**

$$i = \underset{j \in \{0, \dots, N-1\}}{\operatorname{argmin}} D(y, x_j)$$

2. **Prediction:**

$$\hat{F}_t(y) = U_t f(a_i) = f_{i+q}$$

- Model error is avoided if $\{(x_i, f_i)\}$ are observations of nature
- $\{x_i\}$ does not have to be a Markovian time series

Kernel analog forecasting

Potential shortcomings of conventional analog forecasting:

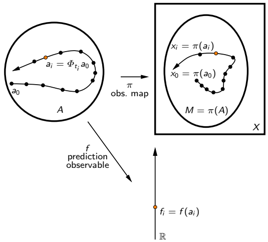
- $\hat{F}_t(y)$ depends discontinuously on the initial data even if f is continuous
- The forecast step uses information from only a single state in the training data

Kernel analog forecasting (Zhao & G. 2016, Comeau et al. 2016) addresses these deficiencies using kernel out-of-sample extension techniques for functions

Two variants of the technique based on:

- **Nyström extension** (Coifman & Lafon 2006)
  Performs best for data-driven bandlimited observables
- **Laplacian pyramids** (Coifman & Rabin 2012, Fernández et al. 2014)
  Works for arbitrary observables, is able to cope with forecast biases in the case of partially observed systems

Basic properties of kernels



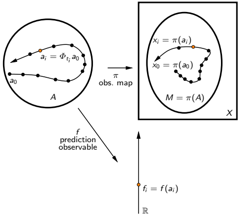
A **kernel** $K : X \times X \mapsto \mathbb{R}$ is a pairwise measure of similarity in data space with "nice" properties

- $0 \leq K \leq C < \infty$
- $0 < c \leq \int_M K(y, \cdot)\, d\mu < \infty$, $y \in X$, $\mu = \alpha \circ \pi^{-1}$
- $K(\cdot, x)$ is continuous for all $x \in M$.

Example (radial Gaussian kernel)**.** X is compact and

$$K(y, x) = e^{-D^2(y,x)/\epsilon}, \quad \epsilon > 0$$

Basic properties of kernels



For data generated by ergodic dynamical systems, kernels naturally lead to **averaging operators**

$$\hat{P} f(b) = \int_A \hat{p}(b, a) f(a)\, d\hat{\alpha}(a) = \frac{1}{N} \sum_{i=0}^{N-1} \hat{p}(b, a_i) f(a_i)$$

- $\hat{P}$ is an operator on $L^2(A, \hat{\alpha})$ for the sampling measure $\hat{\alpha} = N^{-1} \sum_{i=0}^{N-1} \delta_{a_i}$
- $\hat{p}$ is computed by **normalization** of K, e.g.,

$$\hat{p}(b, a_i) = \frac{K(\pi(b), \pi(a_i))}{\hat{q}(b)}, \quad \hat{q}(b) = \frac{1}{N} \sum_{i=0}^{N-1} K(\pi(b), \pi(a_i))$$

- By the pointwise ergodic theorem, as $N \to \infty$, $\hat{P} f(b)$ converges α-a.s. to

$$Pf(b) = \int_A p(b, a) f(a)\, d\alpha(a),$$

$$p(b, a) = \frac{K(\pi(b), \pi(a))}{q(b)}, \quad q(b) = \int_A K(\pi(b), \pi(a))\, d\alpha(a)$$

Basic kernel analog forecasting

Conventional analog forecast at lead $t = q\tau$ can also be expressed as

$$\hat{F}_t y = \int_A U_t f \, d\delta_{a_i} = f_{i+q}, \quad i = \operatorname*{argmin}_{j \in \{0,\dots,N-1\}} D(y, x_j)$$

Kernel-based formulation:

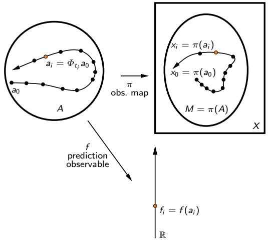
1. Map the initial data y to the probability measure

$$\hat{\nu}_y = \frac{1}{N} \sum_{i=0}^{N-1} \hat{\rho}(y, x_i) \delta_{a_i}, \quad \hat{\rho}(y, x_i) = \frac{K(y, x_i)}{\hat{q}(y)}$$

2. Predict using the expectation value

$$\hat{F}_t(y) = \int_A U_t f \, d\hat{\nu}_y = \frac{1}{N} \sum_{i=0}^{N-1} \hat{\rho}(y, x_i) f_{i+q}$$

Basic kernel analog forecasting



$$\hat{F}_t(y) = \int_A U_t f \, d\hat{\nu}_y = \frac{1}{N} \sum_{i=0}^{N-1} \hat{\rho}(y, x_i) f_{i+q}$$

- There exists a unique function $\hat{f}_t \in L^2(A, \alpha)$ lying above $\hat{F}_t$ in the sense that $\hat{f}_t(b) = \hat{F}_t(\pi(b))$ for all $b \in A$:

$$\hat{f}_t(b) = \frac{1}{N} \sum_{i=0}^{N-1} \hat{p}(b, a_i) f_{i+q}$$

- As $N \to \infty$, $\hat{F}_t(y)$ converges to

$$F_t(y) = \int_A U_t f \, d\nu_y, \quad \frac{d\nu_y}{d\alpha} = \rho(y, \pi(\cdot)) = \frac{K(y, \pi(\cdot))}{q(y)},$$

  and there exists a unique $f_t \in L^2(A, \alpha)$ lying above F_t
- If the dynamics is mixing, $f_t \xrightarrow[t \to \infty]{} \bar{f} = \int_A f \, d\alpha$

Basic kernel analog forecasting

Pros:

- $\hat{F}_t$ is continuous if K is continuous in its first argument
- Multiple states in the training data are weighted

Cons:

- The forecast can be biased, i.e.,

$$\int_A r_t \, d\alpha \neq 0, \quad r_t = f_t - U_t f$$

- The forecast can be overly diffusive, in the sense that even if P is a "good" averaging operator (positive semidefinite, self-adjoint on $L^2(A, \alpha)$), $\|r_t\| \geq \gamma \|f - \bar{f}\|$, where γ is the spectral gap of P, and $\|r_t\|$ can be arbitrarily close to $\|f - \bar{f}\|$

Reproducing kernel Hilbert spaces

A **positive-semidefinite kernel** $\sigma : X \times X \mapsto \mathbb{R}$ has the integrability and boundedness properties of non-symmetric kernels plus

- Symmetry: $\sigma(y, x) = \sigma(x, y)$ for all $x, y \in X$
- Non-negativity: $\sum_{i,j=1}^{n} c_i \sigma(x_i, x_j) c_j \geq 0$ for all $n \in \mathbb{N}$, $x_1, \ldots, x_n \in X$, $c_1, \ldots, c_n \in \mathbb{R}$

σ induces a positive-semidefinite kernel on A, $s(b, a) = \sigma(\pi(b), \pi(a))$

Theorem (Moore-Aronszajn). There exists a unique Hilbert space $\mathcal{H}$ of functions $A \mapsto \mathbb{R}$ s.t. for all $b \in A$,

$$f(b) = \langle s(b, \cdot), f \rangle_{\mathcal{H}}$$

Reproducing kernel Hilbert spaces

Given a normalized non-symmetric kernel $\hat{\rho}$, we can form a positive-semidefinite kernel via "right" symmetrization

$$\hat{\sigma}(y, x) = \frac{1}{N} \sum_{i=0}^{N-1} \hat{\rho}(y, x_i,)\hat{\rho}(x, x_i)$$

[similarly for "left" symmetrization]

As $N \to \infty$, $\hat{\sigma}$ converges α-a.s. to

$$\sigma(y, x) = \int_M \rho(y, \cdot)\rho(x, \cdot)\, d\mu$$

- $\hat{\sigma}$ is computable from finite datasets whereas σ is not

Reproducing kernel Hilbert spaces

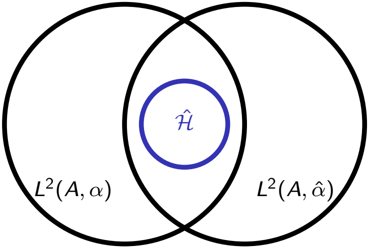
The RKHS $\hat{\mathcal{H}}$ associated with $\hat{s}$ consists of functions of the form

$$f = \sum_{i=0}^{N-1} c_i \hat{p}(\cdot, a_i), \quad \frac{1}{N} \sum_{i=0}^{N-1} c_i^2 < \infty$$

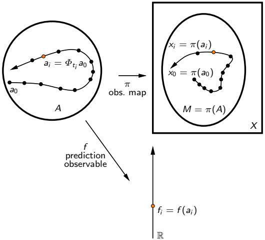
- Because $\hat{p}(b, a_i) = \hat{\rho}(\pi(b), \pi(a_i))$, $f(b)$ is constant on $\pi^{-1}(\pi(b))$

$\hat{\mathcal{H}}$ lies in the intersection of $L^2(A, \alpha)$ and $L^2(A, \hat{\alpha})$, i.e.,

$$f \in \hat{\mathcal{H}} \implies \|f\|_\alpha = \int_A f^2 \, d\alpha < \infty, \quad \|f\|_{\hat{\alpha}} = \frac{1}{N} \sum_{i=0}^{N-1} f^2(a_i) < \infty$$



Nyström extension



There exist subspaces $B_l \subset \hat{\mathcal{H}}$ of **bandlimited observables** which can be evaluated at arbitrary points on A given their values on the sampled states

$$B_l = \operatorname{span}\{\psi_0, \dots, \psi_{l-1}\}$$

$$\psi_k = \frac{1}{\lambda_k^{1/2}} \hat{S}\phi_k, \quad \lambda_k > 0$$

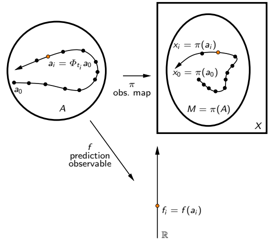
$$\hat{S} : L^2(A, \hat{\alpha}) \mapsto \hat{\mathcal{H}}, \quad \hat{S}f = \int_A \hat{s}(\cdot, a)\, d\hat{\alpha}(a) = \frac{1}{N} \sum_{i=0}^{N-1} \hat{s}(\cdot, a_i) f(a_i)$$

- $\{\phi_k\}_{k=0}^{N-1}$ is an orthonormal basis of $L^2(A, \hat{\alpha})$ consisting of eigenfunctions of $\hat{S}^* \hat{S}$
- In the case of left (right) symmetrization, the $\phi_k(a_i)$ are given by the right (left) singular vectors of $\mathbf{P} = N^{-1}[\hat{p}(a_i, a_j)]$, and $\lambda_k^{1/2}$ are equal to the corresponding singular values
- In the context of out-of-sample extension of functions on manifolds, the ψ_k are known as **geometric harmonics** (Coifman & Lafon 2006)

Nyström extension

- The ψ_k are orthonormal on $\hat{\mathcal{H}}$
- For $f = \sum_{k=0}^{l-1} c_k \psi_k$, then, *without approximation*,

$$c_k = \frac{1}{\lambda_k^{1/2}} \langle \phi_k, f \rangle_{\hat{\alpha}} = \frac{1}{N\lambda_k^{1/2}} \sum_{i=0}^{N-1} f(a_i)\phi_k(a_i)$$



- ψ_k lies above a unique function $\Psi_k : X \mapsto \mathbb{R}$ s.t.

$$\Psi_k(y) = \frac{1}{N} \sum_{j=0}^{N-1} \hat{\sigma}(y, x_i)\phi_k(a_i)$$

- Similarly, f lies above $F = \sum_{k=0}^{l-1} c_k \Psi_k$

- Given $g = \sum_{k=0}^{N-1} d_k \phi_k \in L^2(A, \hat{\alpha})$, the RKHS norm $\|g\|_{\hat{\mathcal{H}}}^2 = \sum_{k=0}^{l-1} d_k^2/\lambda_k$ measures its **roughness** with respect to the kernel $\hat{s}$

Kernel analog forecasting with Nyström extension

1. Expand the time-shifted forecast observable $U_t f$, $t = q\tau$, as

$$U_t f = \sum_{k=0}^{l-1} c_k(t)\psi_k + r_t,$$

where r_t is in the orthogonal complement $B_l^\perp$ of B_l in $L^2(A, \alpha)$

2. Compute the expansion coefficients

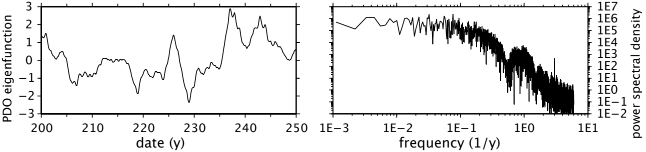
$$c_k(t) = \langle \psi_k, U_t f \rangle_{\hat{\mathcal{H}}} = \frac{1}{N\lambda_k^{1/2}} \sum_{i=0}^{N-1} \phi_k(a_i) f_{i+q}$$

3. Given initial data $y \in X$, evaluate the forecast using

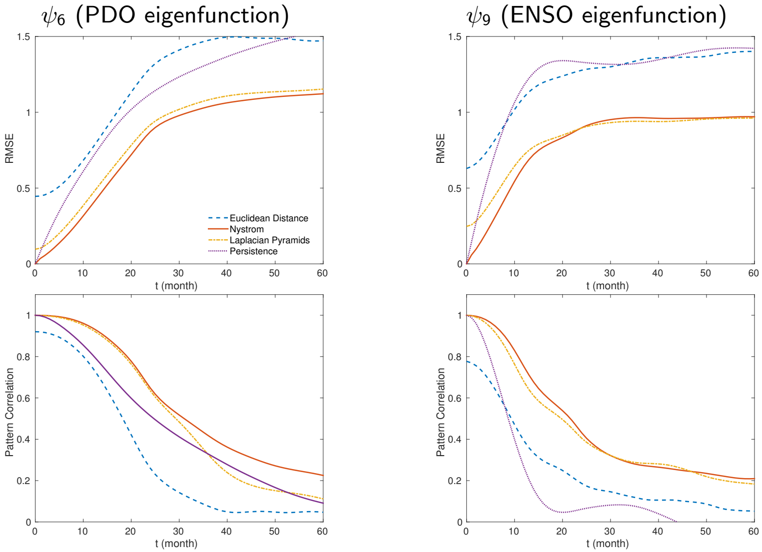
$$\hat{F}_t(y) = \sum_{k=0}^{l-1} c_k(t)\Psi_k(y)$$

Kernel analog forecasting with Nyström extension

- If $U_t f$ lies in B_l for all lead times of interest, then, by construction, analog forecasting with Nyström extension has vanishing forecast error
- In practice, $U_t f$ will have a nonzero residual in $B_l^\perp$ at $t = 0$ and/or at $t > 0$ since B_l is not a U_t-invariant subspace of $L^2(A, \alpha)$
- Yet, for appropriately constructed kernels, B_l contains physically meaningful **data-driven observables**

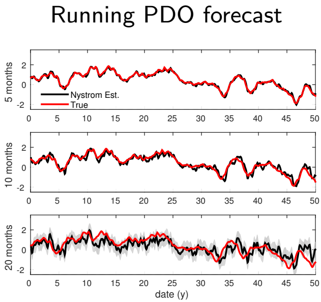


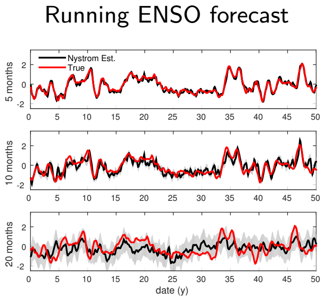
Analog prediction of low-frequency eigenfunctions of North Pacific SST



(Zhao & G. 2016, Comeau et al. 2016)

Analog prediction of low-frequency eigenfunctions of North Pacific SST

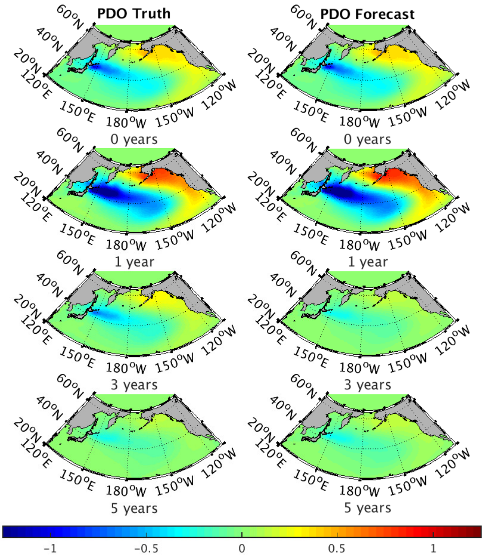




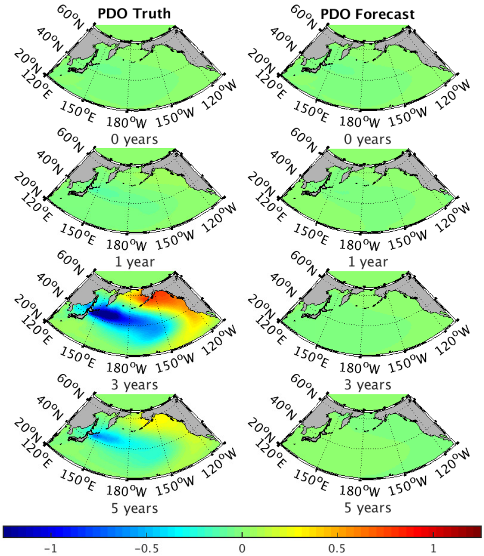
(Rudimentary) **uncertainty quantification:**
We compute two-sided error bars $\pm\varepsilon_t(y)$ through a local average of the squared residual norm in the training dataset

$$\epsilon_t^2(y) = \frac{1}{N}\sum_{i=0}^{N-1}\hat{\rho}(y, x_i)|r_t(x_i)|^2, \quad r_t(x_i) = \hat{F}_t(x_i) - f_{i+q}, \quad t = q\tau$$

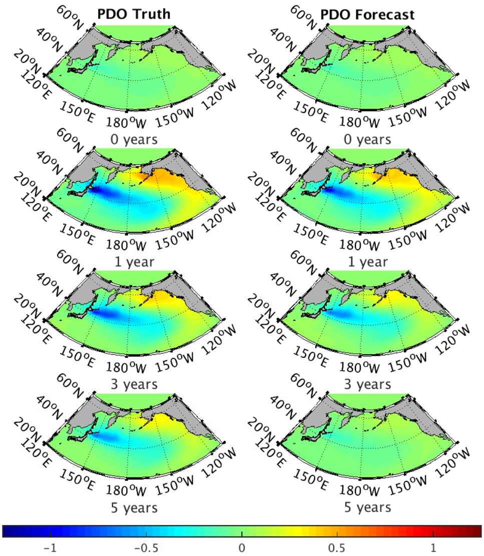
Prediction of spatial PDO patterns



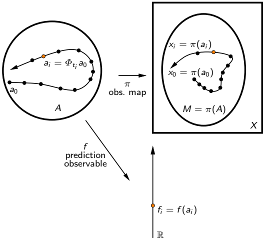
Prediction of spatial PDO patterns



Prediction of spatial PDO patterns



Choice of kernel



Restrict attention to the case $X = \mathbb{R}^d$, $M = \pi(A) \subset X$ is compact, and K_ϵ is a **symmetric, exponentially decaying kernel** with a **bandwidth parameter** ϵ, e.g.,

$$K_\epsilon(y, x) = e^{-\|y-x\|^2/\epsilon}$$

Diffusion maps normalization (Coifman & Lafon 2006, Berry & Sauer 2015)

1. "Right" normalization:

$$\hat{K}'_\epsilon(y, x) = \frac{K_\epsilon(y, x)}{\hat{q}_\epsilon^{1/2}(x)}, \quad \hat{q}_\epsilon(x) = \frac{1}{N} \sum_{i=0}^{N-1} K_\epsilon(x, x_i),$$

2. "Left" normalization:

$$\hat{\rho}_\epsilon(y, x) = \frac{\hat{K}'_\epsilon(y, x)}{\hat{d}_\epsilon(y)}, \quad \hat{d}_\epsilon(y) = \frac{1}{N} \sum_{i=0}^{N-1} \hat{K}'_\epsilon(y, x_i)$$

By ergodicity, $\hat{\rho}_\epsilon(y, x) \xrightarrow[N \to \infty]{} \rho_\epsilon(y, x)$, given by replacing $N^{-1} \sum_i$ in the above by $\int_M d\mu$

Choice of kernel

Associated with $\hat{\rho}_\epsilon$ is an (ergodic) **Markov semigroup** $\{\hat{P}_\epsilon^n\}_{n\in\mathbb{N}}$ on $L^2(A,\hat{\alpha})$ with

$$\hat{P}_\epsilon f(b) = \frac{1}{N}\sum_{i=0}^{N-1} \hat{p}_\epsilon(b,a_i) f(a_i), \quad \hat{p}_\epsilon(b,a) = \hat{\rho}_\epsilon(\pi(b),\pi(a))$$

As $N\to\infty$, $\hat{P}_\epsilon f(b) \to P_\epsilon f(b)$ (given again by replacing sums with integrals)

Recall that the $\{\phi_i\}$ basis of $L^2(A,\hat{\alpha})$ used in Nyström extension is determined by the SVD of $\hat{P}_\epsilon$

- Can study the asymptotic behavior of the scheme in the limit of large data through the properties of P_ϵ

Small-bandwidth asymptotics

If A is a smooth manifold diffeomorphic to M (i.e., the observations are **full**), then, uniformly on A (Coifman & Lafon 2006, Berry & Sauer 2015),

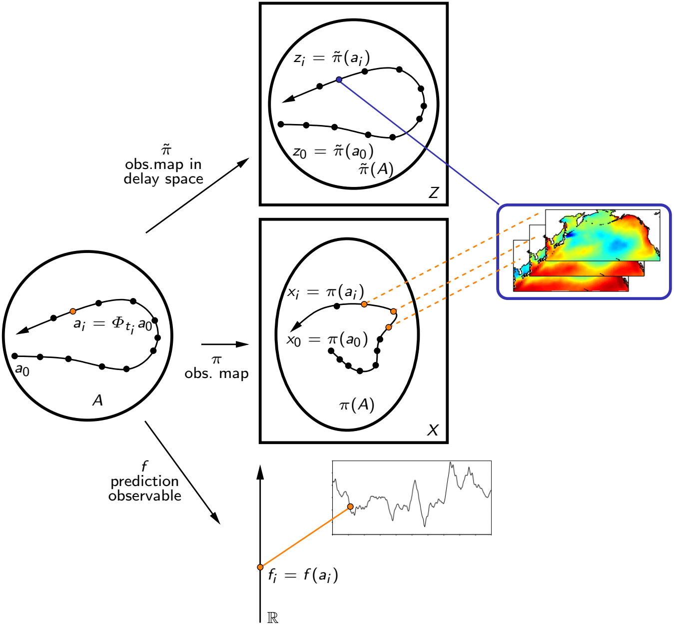
$$Pf(b) = f(b) - \epsilon \mathcal{L} f(b) + O(\epsilon^2)$$

- $\mathcal{L}$ is the generator of a **gradient flow** for a Riemannian metric g that depends on K_ϵ and a potential $e^{-\theta}$, $\theta = d\alpha / d\,\mathrm{vol}_g$:

$$\mathcal{L} f = \Delta f - \frac{\Delta \theta^{1/2}}{\theta^{1/2}} f, \quad \Delta f = -\operatorname{div}_g \operatorname{grad}_g f$$

- In the limit $\epsilon \to 0$, $N \to \infty$, the ϕ_k converge to the eigenfunctions of $\mathcal{L}$, which have a geometrical interpretation as the extrema of the Rayleigh quotient

$$\frac{E(f)}{\|f\|^2}, \quad E(f) = \int_A \|\operatorname{grad}_g f\|^2 \, d\alpha$$



$z_i = \tilde{\pi}(a_i)$
$z_0 = \tilde{\pi}(a_0)$
$\tilde{\pi}(A)$
Z
$\tilde{\pi}$
obs.map in
delay space
$x_i = \pi(a_i)$
$x_0 = \pi(a_0)$
$\pi(A)$
X
$a_i = \Phi_{t_i} a_0$
a_0
A
π
obs. map
f
prediction
observable
$f_i = f(a_i)$
$\mathbb{R}$

Dynamics-aware kernels

The following features enhance the **timescale separation** and **physical interpretability** of the eigenfunctions:

- **Delay-coordinate embeddings** (G. & Majda 2011–2014, Berry et al. 2013)
  Concatenate snapshots over a running window to form "videos"

$$z_i = \tilde{\pi}(a_i) = (x_i, x_{i-1}, \dots, x_{i-(q-1)}), \quad x_i = \pi(a_i)$$

  Kernels evaluated on Z become increasingly biased towards stable Lyapunov directions (Berry et al. 2013)

$$K_\epsilon(z_i, z_j) = e^{-\|z_i - z_j\|^2/\epsilon}$$

- **"Cone" kernels** (G. 2015)

$$K_{\epsilon,\zeta}(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{\epsilon\|\xi_i\|\|\xi_j\|}[(1 - \zeta\cos^2\theta_i)(1 - \zeta\cos^2\theta_j)]^{1/2}\right)$$

$$\xi_i = \frac{x_{i+1} - x_{i-1}}{2}, \quad \cos\theta_i = \frac{(x_j - x_i)\cdot\xi_i}{\|\xi_i\|\|x_j - x_i\|}, \quad \zeta < 1$$

  Directional dependence on ξ_i acts as an intrinsic **low-pass filter** favoring slowly-varying observables

Recovering slow intrinsic timescales

Cone kernels are associated with a modified **Riemannian geometry** on A with the metric tensor

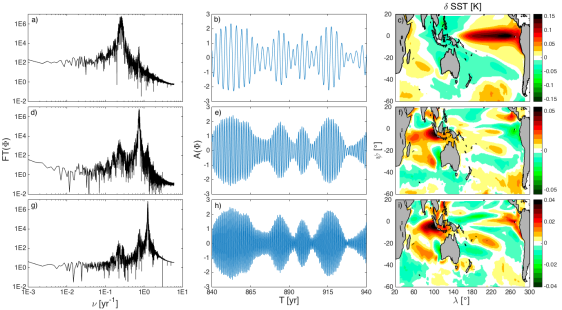
$$h_\zeta = \frac{1}{\|v\|_g^2}\left(g - \zeta \frac{v^\flat \otimes v^\flat}{\|v\|_g^2}\right), \quad v = \left.\frac{dU_t}{dt}\right|_{t=0}$$

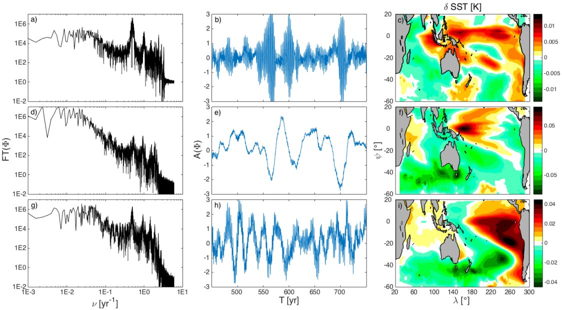
- v is the **vector field** on A generating the dynamics
- As $\zeta \to 1$, h_ζ increasingly contracts lengths along v
- The associated gradient flow becomes increasingly biased along the integral curves of the dynamics

$$E_\zeta(f) = \frac{1}{(1-\zeta)} \int_A (v(f))^2 \, d\alpha + O((1-\zeta)^0)$$

- The leading eigenfunctions with small $E_\zeta(\phi_k)$ have **small directional derivative** $v(\phi_k)$, i.e., are slowly varying

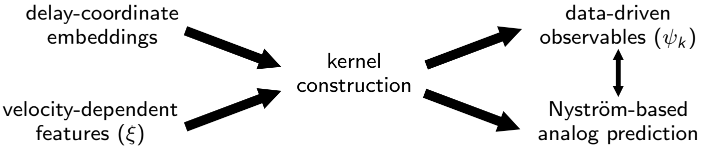
Recovering multiple intrinsic timescales





Interannual and decadal patterns of Indo-Pacific SST extracted via cone kernels (Slawinska & G. 2016)

Summary of kernel analog forecasting with Nyström extension



- Technique is expected to perform well for bandlimited forecast observables (e.g., kernel eigenfunctions)
- Potential shortcomings include biases and/or poor conditioning for broadband observables in the kernel eigenfunction basis (in such cases, use Laplacian pyramids technique)

Summary & outlook

- Kernel analog forecasting provides a flexible nonparametric modeling approach, which does not require prior knowledge of the equations of motion, or an appropriate parametric model structure
- Variants of the technique based on Nyström extension and Laplacian pyramids are applicable for bandlimited and broadband observables (in the RKHS sense), respectively
- Interplay between "dynamics-aware" kernels and data-driven observables constructed through kernel eigenfunctions
- Promising results in SST and sea-ice prediction and forecasts of tropical intraseasonal oscillations (Alexander et al. 2016; not discussed here)

Summary & outlook

Several open questions for future research, including:

- Establishment of convergence rates and associated truncation criteria to prevent overfitting
- Improved UQ
- Kernel design: Can we construct kernels tailored to the given prediction observable?
- Incorporation of prior information about the equations of motion
- Study effects of observational noise
- Extension to nonautonomous and/or stochastic dynamics
- Extension to vector-valued prediction observables

References

- Z. Zhao, D. Giannakis (2016). Analog forecasting with dynamics-adapted kernels. *Nonlinearity*. In minor revision

- D. Comeau, Z. Zhao, D. Giannakis, A. J. Majda (2016). Data-driven prediction strategies for low-frequency patterns of North Pacific climate variability *Climate Dyn.* In press

- D. Giannakis, A. J. Majda (2012). Nonlinear Laplacian spectral analysis for time series with intermittency and low-frequency variability. *Proc. Natl. Acad. Sci.*, 109(7), 2222–2227

- D. Giannakis (2015). Dynamics-adapted cone kernels. *SIAM J. Appl. Dyn. Sys.*, 14(2), 556–608

- R. Alexander, Z. Zhao, E. Szekely, D. Giannakis (2016). Kernel analog forecasting of tropical intraseasonal oscillations. *J. Atmos. Sci.* In review

- J. Slawinska, D. Giannakis (2016). Timescale separation in the Indo-Pacific: Revealing intrinsic modes of SST variability on seasonal to multidecadal timescales. *J. Climate*. In revision

Additional slides

Kernel analog forecasting with Laplacian pyramids

- Let $P_0, P_1, \ldots, P_l$ be a sequence of averaging operators constructed from kernels K_{ϵ_i} with bandwidth parameters $\epsilon_0 > \epsilon_1 > \cdots > \epsilon_l$
- Consider the basic kernel analog forecast formula

$$f_{t,0}(b) = P_0 U_t f(b) = \int_A p_0(b, a) U_t f(a) \, d\alpha(a)$$

Kernel analog forecasting with Laplacian pyramids

- Let $P_0, P_1, \dots, P_l$ be a sequence of averaging operators constructed from kernels K_{ϵ_i} with bandwidth parameters $\epsilon_0 > \epsilon_1 > \cdots > \epsilon_l$
- Consider the basic kernel analog forecast formula

$$f_{t,0}(b) = P_0 U_t f(b) = \int_A p_0(b,a) U_t f(a) \, d\alpha(a)$$

- In LP (Coifman & Rabin 2012, Fernández et al. 2014), the residual $r_{t,0}(b) = U_t f(b) - P_0 U_t f(b)$ is viewed as a new observable to be approximated via

$$g_{t,1}(b) = P_1 r_{t,0}(b) = \int_A p_1(b,a) r_{t,1}(a) \, d\alpha(\alpha)$$

Kernel analog forecasting with Laplacian pyramids

- Let $P_0, P_1, \dots, P_l$ be a sequence of averaging operators constructed from kernels K_{ϵ_i} with bandwidth parameters $\epsilon_0 > \epsilon_1 > \cdots > \epsilon_l$
- Consider the basic kernel analog forecast formula

$$f_{t,0}(b) = P_0 U_t f(b) = \int_A p_0(b,a) U_t f(a)\, d\alpha(a)$$

- In LP (Coifman & Rabin 2012, Fernández et al. 2014), the residual $r_{t,0}(b) = U_t f(b) - P_0 U_t f(b)$ is viewed as a new observable to be approximated via

$$g_{t,1}(b) = P_1 r_{t,0}(b) = \int_A p_1(b,a) r_{t,1}(a)\, d\alpha(\alpha)$$

- Iterating this procedure l times yields the l-level **Laplacian pyramids approximation** of the forecast observable f at lead time t:

$$f_{t,l} = f_{t,0} + \sum_{i=1}^{l} g_{t,i}$$

Kernel analog forecasting with Laplacian pyramids

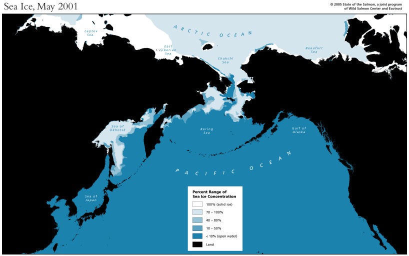
- LP-based kernel analog forecasting is better suited than the Nyström-based approach to deal with non-bandlimited forecast observables for the given kernel family
- At each level of approximation, the refinement $g_{t,i} = P_i r_{t,i-1}$ lies in a RKHS $\mathcal{H}_i$ associated with p_i
- If the P_i are positive-semidefinite, self-adjoint on $L^2(A, \alpha)$,

$$\|r_{t,i}\| \leq \|r_{t,i-1}\|$$

- LP-based forecasts with general averaging operators can be biased,

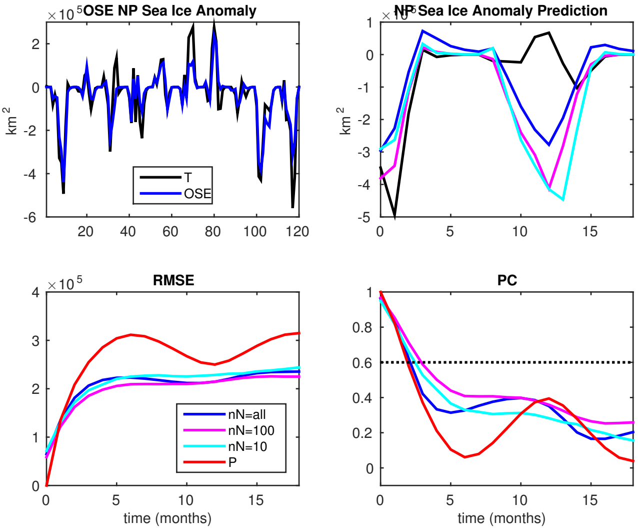
$$\int_A r_{t,l} \, d\alpha \neq 0$$

North Pacific sea-ice prediction



- Kernel analog forecasts of North Pacific total sea-ice area from extended control integration of the CCSM4 climate model and HadISST satellite observations (Comeau et. al 2015)
- Signal dominated by seasonal cycle; anomalies relative to seasonal cycle are highly intermittent and have heavy-tailed distributions
- Multivariate predictors (SST and sea ice concentration) used
- Significant trend (non-autonomous dynamics) present in nature (HadISST dataset)

Kernel analog forecast results—CCSM4 data



Kernel analog forecast results—HadISST data