Application and validation of polynomial chaos methods to quantify uncertainties in simulating the Gulf of Mexico circulation using HYCOM.

Mohamed Iskandarani **Matthieu Le Hénaff** **Carlisle Thacker**
University of Miami
Guotu Li **Omar Knio**
Duke University
Ashwanth Srinivasan
Tendral LLC

June 2, 2015

Outline

- Use Polynomial Chaos (PC) expansions to quantify uncertainties in oceanic forecasts
  - Approximate model with an accurate surrogate
  - Compute series coefficients via an ensemble
  - Validate the accuracy of the surrogate
  - Mine the series for statistical information in lieu of model
- Initial Condition uncertainties in the Gulf of Mexico
  - How to perturb a field?
  - How to localize perturbations around a dynamical process?
  - Series Validation
  - Statistical Outputs
- Uncertainty Experiments to date
  - Initial conditions uncertainties (small ensemble)
  - Initial conditions and wind forcing uncertainties

HYCOM surrogate idea

- $M(\boldsymbol{x}, t, \xi)$ is a model output
- ξ is a stochastic variable that represents the dependence of M on the uncertain input data
- ξ is characterized by its probability density function $\rho(\xi)$
- The mean of M is: $\overline{M}(\boldsymbol{x}, t) = \int M(\boldsymbol{x}, t, \xi)\, \rho(\xi)\, \mathrm{d}\xi$
- Its variance is

$$\sigma^2(M) = \overline{(M - \overline{M})^2} = \int (M - \overline{M})^2\, \rho(\xi)\, \mathrm{d}\xi$$

- A surrogate embodies the dependence of M on the uncertain data ξ via a spectral series in ξ:

$$M(\boldsymbol{x}, t, \xi) = \sum_{n=0}^{P} \widehat{M}_n(\boldsymbol{x}, t)\, \psi_n(\xi)$$

PC surrogate series $M(\boldsymbol{x}, t, \boldsymbol{\xi}) = \sum_{n=0}^{P} \widehat{M}_n(\boldsymbol{x}, t)\ \psi_n(\boldsymbol{\xi})$

- $\widehat{M}_n(\boldsymbol{x}, t)$: series coefficients
- $\psi_k(\boldsymbol{\xi})$: orthonormal basis functions w.r.t. $\rho(\boldsymbol{\xi})$

$$\overline{\psi_m \psi_n} = \int \psi_m(\boldsymbol{\xi})\psi_n(\boldsymbol{\xi})\rho(\xi)\mathrm{d}\boldsymbol{\xi} = \delta_{m,n}$$

- $\psi_m(\boldsymbol{\xi})$ consits of Legendre polynomials when $\rho(\boldsymbol{\xi})$ is a uniform distribution
- mean: $\overline{M} = \sum_{n=0}^{P} \widehat{M}_n(\boldsymbol{x}, t)\overline{\psi_n, \psi_0} = \widehat{M}_0(\boldsymbol{x}, t)$
- Variance: $\overline{(M - \overline{M})^2} = \sum_{n=1}^{P} \widehat{M}_n^2(\boldsymbol{x}, t)$
- Where to truncate the series, P ? Monitor Variance
- How to determine the coefficients $\widehat{M}_n$?

Calculating the PC series coefficients

- Minimize the norm of the approximation error by using either Galerkin projection, interpolation, least square, or compressed sensing. All can be implemented via ensemble.
- The least square approaches are useful when model response includes model noise.
- Galerkin projection exploits orthogonality

$$\widehat{M} = \overline{M\psi_m} = \int M(\boldsymbol{x}, t, \boldsymbol{\xi})\psi_m(\boldsymbol{\xi})\rho(\boldsymbol{\xi})\mathrm{d}\boldsymbol{\xi}$$

$$\approx \sum_{q=1}^{Q} M(\boldsymbol{x}, t, \boldsymbol{\xi}_q)\psi_m(\boldsymbol{\xi}_q)\omega_q$$

- $\boldsymbol{\xi}_q, \omega_q$ are appropriate quadrature roots,weights
- $M(\boldsymbol{x}, t, \boldsymbol{\xi}_q)$ requires an ensemble at the quadrature roots

Gulf of Mexico Circulation

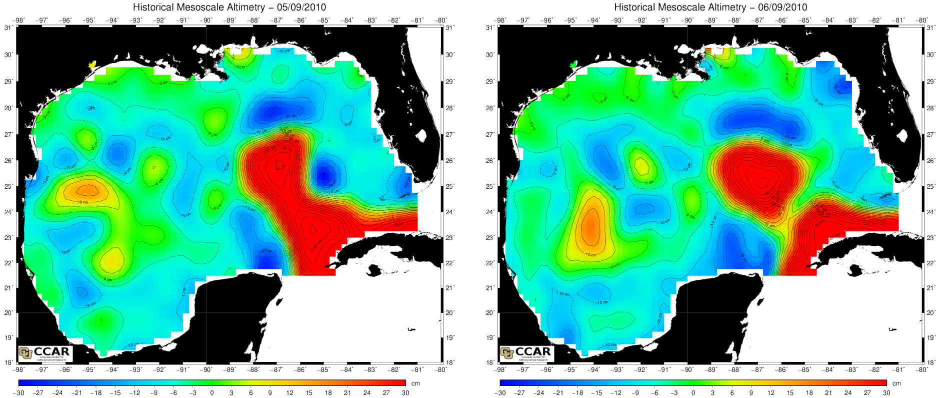


Figure: Sea Surface Height in cm from AVISO

Target Loop Current Eddy separation during May-June 2010.
Capture the role of the frontal anticyclones in eddy detachment.

Uncertainty in Initial Boundary Conditions

- Computational challenge: Number of sample grows exponentially with the number of stochastic variables.
- Rely on EOFs to characterize uncertainty and reduce the number of stochastic variables. For 2 EOFs mode we have:

$$M(\boldsymbol{x}, 0, \xi_1, \xi_2) = \overline{M}(\boldsymbol{x}, 0) + \left[\sqrt{\lambda_1}\mathcal{M}_1\xi_1 + \sqrt{\lambda_2}\mathcal{M}_2\xi_2\right] \quad (1)$$

- $(\lambda_k, \mathcal{M}_k)$: are eigenvalues/eigenvectors of covariance matrix obtained from free-run simulation
- $\overline{M}$: unperturbed initial condition
- $M(\boldsymbol{x}, 0, \xi)$: Stochastic initial condition input
- ξ_1, ξ_2 are the amplitudes of the perturbations

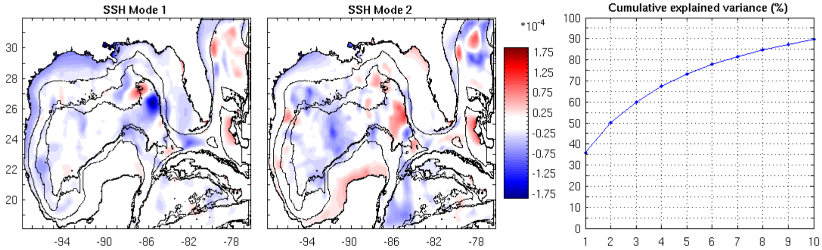


Figure: First and Second SSH modes from a 14-day series. The 2 modes account for 50% of variance during these 14 days.

- Characterize local uncertainty: get perturbation from short, 14-day, simulation.
- Uncertainty dominated by Loop Current (LC) dynamics
- Mode 1 seems associated with a frontal eddy

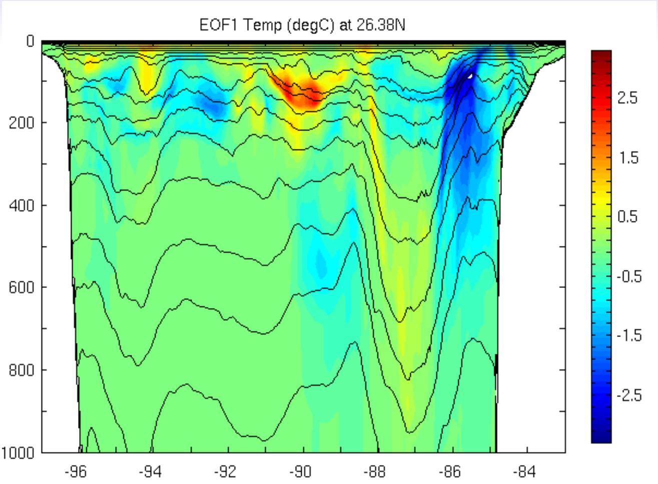


Figure: Vertical slice along 26.4N showing Temperature perturbations. The first mode shows a strong 2.5°C cooling in the vicinity of the frontal cyclones. The "warm" perturbation around 90W is at the southern edge of a small anticyclone NW of the LC.

PC representation

- (ξ_1, ξ_2) independent and uniformly distributed random variables
- PC basis: Legendre polynomials of max degree 6, $P = 28$
- Ensemble of 49 realizations for Gauss-Legendre quadrature

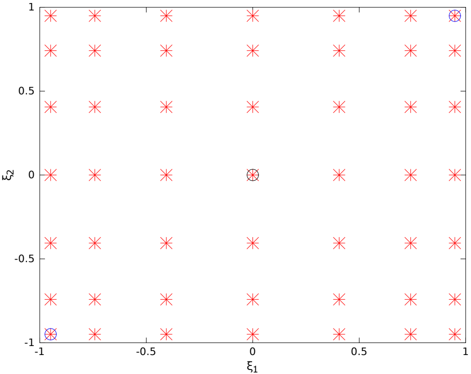
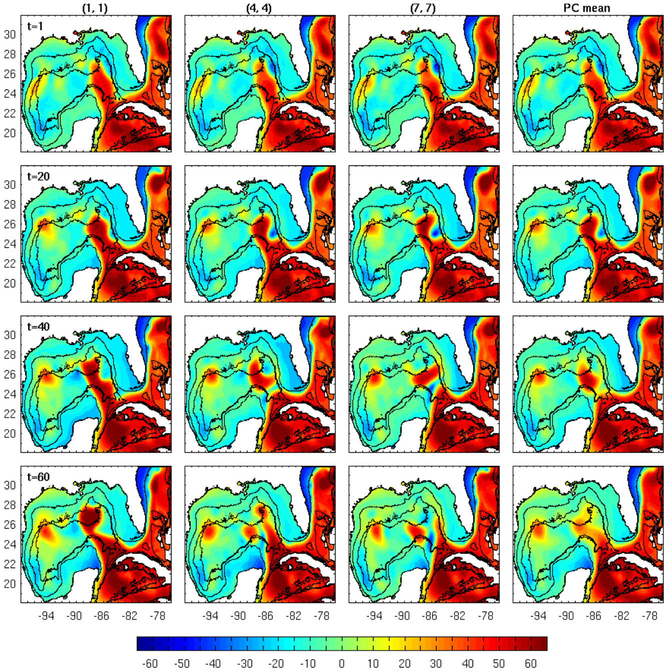


Figure: Quadrature/Sample points in ξ_1, ξ_2 space. Center black circle corresponds to unperturbed run, while blue circles correspond to largest negative and positive perturbations.

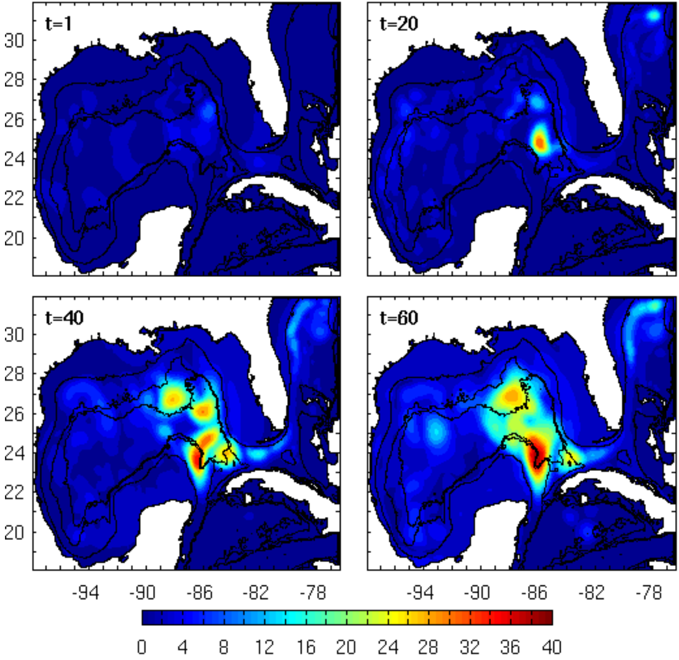


Col 1: SSH of realization (1,1) with weakest frontal eddy

Col 2: SSH of unperturbed realization (4,4) has medium strength frontal eddy

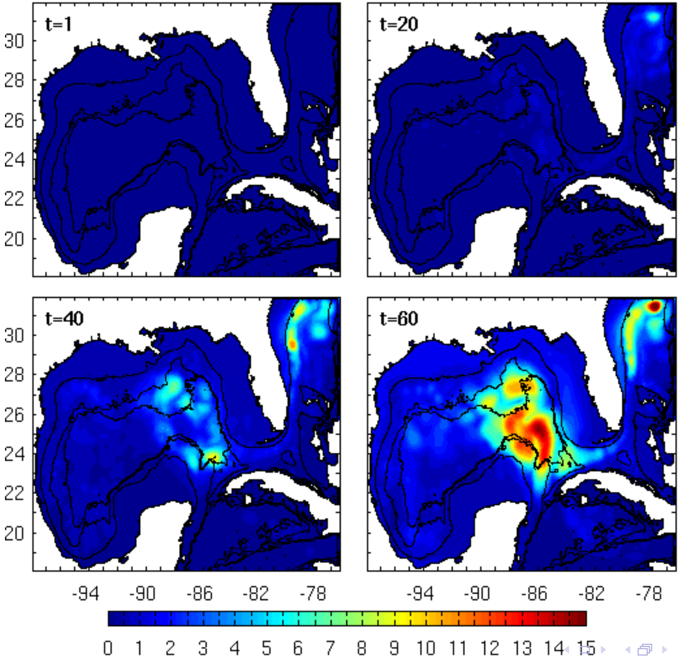
Col 3: SSH of realization (7,7) has strogest frontal eddy and earliest LC separation

Col 4: Loop current edge in ensemble



SSH stddev (cm) grows in time with maximum in LC region

PC-error: $\|\epsilon\|_2^2 = \sum_q \left[\eta(\vec{x}, t, \xi_q) - \eta_{PC}(\vec{x}, t, \xi_q)\right]^2 \omega_q$



SSH PC-errors (cm) grow in time with maxima in LC region

On day 60 PC-error is about 38% of stddev

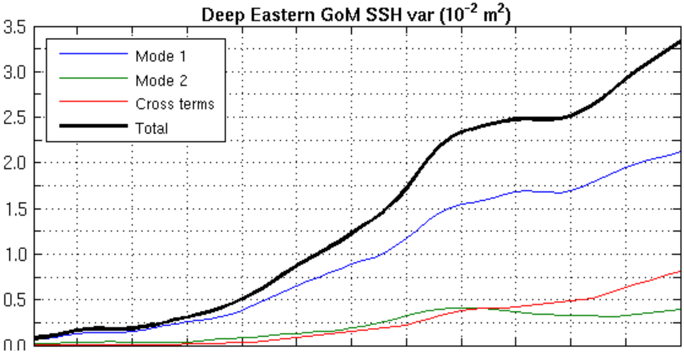


Figure: **Variance Analysis:**The majority of the variance in the deep part of the Eastern Gulf of Mexico can be attributed to the 1st EOF, while the second mode plays a secondary role, particularly during the time span when the series is reliable ($<$ 40 days, x-axis tick marks interval is 5-days).

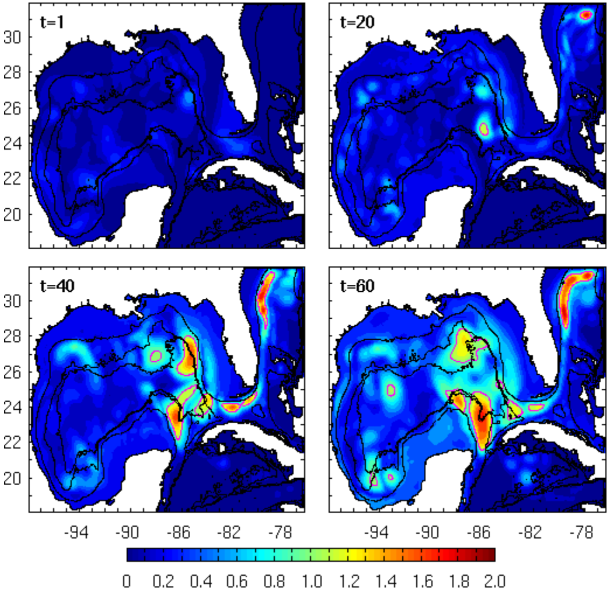


Figure: **Predictability Limit**:Spatial distribution of the ratio of the forecast standard deviation to climatology standard deviation for SSH (from AVISO). The magenta lines show areas where the ratio > 1.

Initial Conditions & Wind Forcing Uncertainties

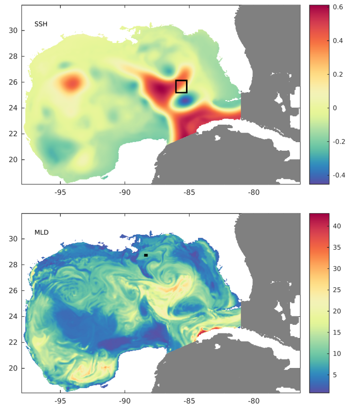


Figure: Field snapshots on day 30: (Top) SSH; (Bottom) MLD

Focus on the following two Quantities of Interest (QoIs):

- Sea Surface Height (SSH) averaged over a square area near the loop current (LC) region: $[-86.04^\circ, -85.20^\circ]$ in longitude and $[25.19^\circ, 26.23^\circ]$ in latitude
- Mixed Layer Depth (MLD) averaged over the DeepWater Horizon (DWH) region: $[-88.44^\circ, -88.28^\circ]$ in longitude and $[28.68^\circ, 28.79^\circ]$ in latitude

Our HYCOM simulations start from 05-01-2010 to 05-30-2010. Fig.7 on the left shows HYCOM results on the last day of simulation using unperturbed initial and wind forcing fields.

Variance Analysis

1st Order Sensitivity:

$$\mathcal{S}_i = \frac{\sum_{\alpha \in \mathbb{S}_1^i} c_\alpha^2 < \Psi_\alpha, \Psi_\alpha >}{\sum_{i=1}^{P} c_i^2 < \Psi_i, \Psi_i >}$$

Total Order Sensitivity:

$$\mathcal{T}_i = \frac{\sum_{\alpha \in \mathbb{S}_T^i} c_\alpha^2 < \Psi_\alpha, \Psi_\alpha >}{\sum_{i=1}^{P} c_i^2 < \Psi_i, \Psi_i >}$$

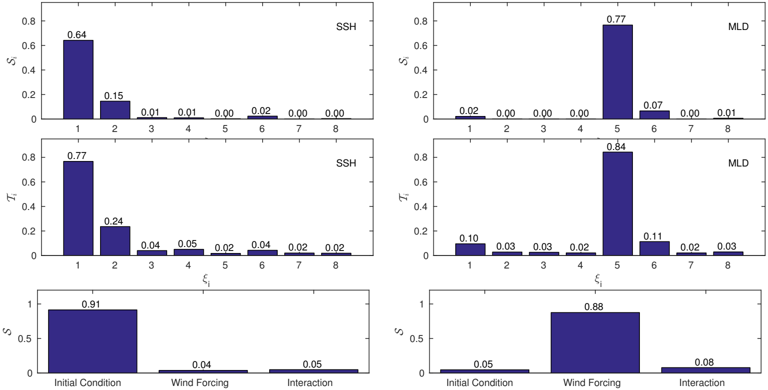
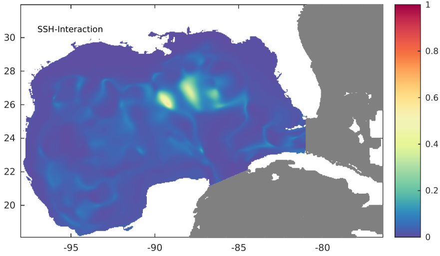
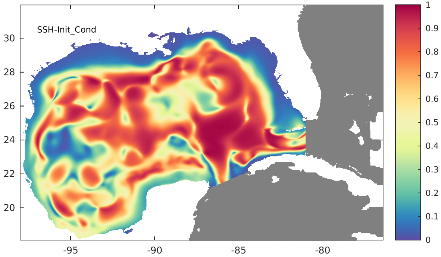


Figure: Sensitivities: (Top) 1st Order; (Middle) total Order; (Bottom) Initial and Wind forcing sensitivities.

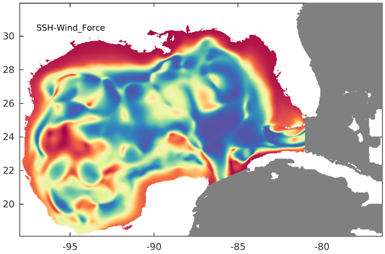
SSH: Joint Sensitivity



(a) Interaction

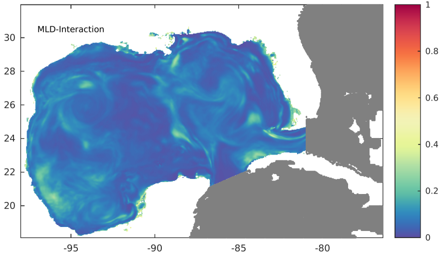


(b) Initial Condition

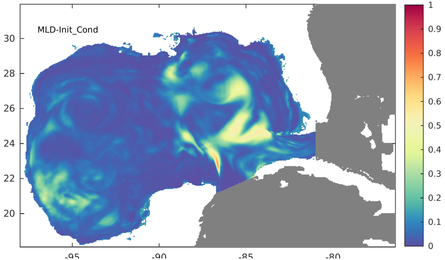


(c) Wind Forcing

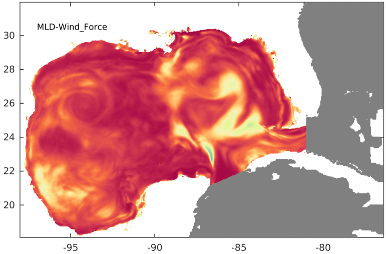
MLD: Joint Sensitivity



(d) Interaction



(e) Initial Condition



(f) Wind Forcing

Conclusions

- Polynomial Chaos are a promising approach in analyzing oceanic uncertainties
- It can monitor representation fidelity and adequacy of sampling
- The 14-day time series EOF analysis helped in designing perturbations focused on a Loop Current separation event.
- The PC-Series performs reasonably well in the 20–40 days range
- Predictability in SSH lost after about 20 days
- It is probably enough to perturb only the 2 leading EOF modes
- Including more uncertainty sources is more important for the current experiment than including more modes.
- SSH uncertainty due mainly to initial conditions except in shelf areas
- MLD uncertainty mainly caused by winds

Publications

- A. Srinivasan, J. Helgers, C. B. Paris, M. LeHenaff, H. Kang, V. Kourafalou, M. Iskandarani, W. C. Thacker, J. P. Zysman, N. F. Tsinoremas, and O. M. Knio. Many task computing for modeling the fate of oil discharged from the deep water horizon well blowout. In *Many-Task Computing on Grids and Supercomputers (MTAGS), 2010 IEEE Workshop on*, pages 1–7, November, 2010. IEEE.
- W. C. Thacker, A. Srinivasan, M. Iskandarani, O. M. Knio, and M. Le Henaff. Propagating oceanographic uncertainties using the method of polynomial chaos expansion. *Ocean Modelling*, **43–44**, pp 52–63, 2012.
- A. Alexanderian, J. Winokur, I. Sraj, M. Iskandarani, A. Srinivasan, W. C. Thacker, and O. M. Knio, Global sensitivity analysis in an ocean general circulation model: a sparse spectral projection approach, *Computational Geosciences*, **16**, 757–778, 2012.
- I. Sraj, M. Iskandarani, A. Srinivasan, W. C. Thacker, A. Alexanderian, C. Lee and S. S. Chen and O. M. Knio, Bayesian Inference of Drag Parameters Using AXBT Data from Typhoon Fanapi, *Monthly Weather Review*, **141**, Issue 7, pp. 2347-2367, 2013.
- J. Winokur, P. Conrad, I. Sraj, O. M. Knio, A. Srinivasan, W. Carlisle Thacker, Y. Marzouk, and M. Iskandarani, A Priori Testing of Sparse Adaptive Polynomial Chaos Expansions Using an Ocean General Circulation Model Database, *Computational Geosciences*, **17**, No 6, pp 899–911, 2013.
- I. Sraj, M. Iskandarani, A. Srinivasan, W. C. Thacker and O. M. Knio, Drag Parameter Estimation using Gradients and Hessian from a Polynomial Chaos Model Surrogate *Monthly Weather Review*, **142**, No 2, pp 933941, 2013.