Physics-Based Covariance Models for Gaussian Processes with Multiple Outputs

Emil M. Constantinescu, Argonne National Laboratory

Context and background

- Gaussian process analysis of processes with multiple outputs is limited by the fact that far fewer good classes of covariance functions exist compared with the scalar (single-output) case.
- We explore models that appropriately account for the major features in the data, such as physical and nonstationarity characteristics.
- Models that incorporate such information are suitable when performing uncertainty quantification or inferences on multidimensional processes with partially known relationships among different variables, also known as co-kriging.
- Covariance structure has a large impact on the uncertainty quantification and forecast efficiency.
- Current techniques rely on local fitting of ad hoc statistical models and may not reveal a robust characterization of the data statistics.

Approach and theoretical considerations

- We introduce analytical and numerical auto-covariance and cross-covariance models that are consistent with physical constraints or can incorporate automatically sensible assumptions about the process generating the data.
- Process nonstationarity is addressed implicitly by the physical parameterization.
- We determine high-order closures, which are required for nonlinear dependencies among the observables.
- These models are applied to Gaussian process regression for processes with multiple outputs and latent processes (i.e., processes that are not directly observed and predicted but interrelate the output quantities).

"Physical" processes under consideration:

$$y_1 = g(y_2) + \eta \quad \text{(explicit)}$$

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = f(y_1,\, y_2) + \psi = \begin{bmatrix} f_1(y_1,\, y_2) + \psi_1 \\ f_2(y_1,\, y_2) + \psi_2 \end{bmatrix} \text{(*)} \quad \text{(implicit)}$$

$$\begin{aligned} y_1 &= f_1(y_1,\, y_2,\, \dots,\, y_m) + \psi_1 \\ y_2 &= f_2(y_1,\, y_2,\, \dots,\, y_m) + \psi_2 \\ \dots &= \dots \\ y_m &= f_m(y_1,\, y_2,\, \dots,\, y_m) + \psi_m \end{aligned} \quad \text{(**)} \quad \text{(general)}$$

- Covariance models for explicit processes:

LEMMA 1. *If two processes y_1 and y_2 satisfy a physical constraint given by $y_1 = g(y_2) + \eta$ with $g(\cdot) \in C^2$, and $\mathrm{Cov}(y_2, y_2) = K_{22}$, then the covariance matrix formed by the elements of the two vectors satisfies*

$$\mathrm{Cov}\left([y_1^T, y_2^T]^T\right) = \begin{bmatrix} LK_{22}L^T + L\mathrm{Cov}(y_2,\eta) + \mathrm{Cov}(\eta, y_2)L^T + \mathrm{Cov}(\eta,\eta) & LK_{22} + \mathrm{Cov}(\eta, y_2) \\ K_{22}L^T + \mathrm{Cov}(y_2,\eta) & K_{22} \end{bmatrix} + \mathcal{O}\left(\delta y_2^3\right),$$

where $L = \left.\frac{\partial g}{\partial y}\right|_{y=\overline{y}}$ is the Jacobian matrix of g evaluated at $\mathrm{E}\{y_2\}$.

- Covariance models for implicit processes:

PROPOSITION 1. *Consider a process driven by the implicit separable system (**). Then under the second-order closure assumptions the block covariance matrix elements satisfy the following simultaneous algebraic equations:*

$$\mathbb{K} = \mathbb{L}\mathbb{K}\mathbb{L}^T + \mathbb{L}K_{y\psi} + K_{\psi y}\mathbb{L}^T + K_{\psi\psi}\,,\ (\mathbb{K})_{ij} = K_{ij} = \mathrm{Cov}(y_i, y_j)\,,\ (\mathbb{L})_{ij} = L_{ij} = \frac{\partial f_i}{\partial y_j}\,.$$

In addition, given K_{22} and if $(I - L_{11})$ is invertible, then for the reduced system (*), the following hold:

$$\begin{aligned} K_{11} = {} & L_{11}K_{11}L_{11}^T + L_{12}\left((I-L_{11})^{-1}(L_{12}K_{22} + K_{\psi_1,2})\right)^T L_{11}^T + \\ & L_{11}\left((I-L_{11})^{-1}(L_{12}K_{22} + K_{\psi_1,2})\right)L_{12}^T + L_{12}K_{22}L_{12}^T + \\ & K_{\psi_1,1}L_{11}^T + K_{\psi_1,2}L_{12}^T + L_{11}K_{1,\psi_1} + L_{12}K_{2,\psi_1} + K_{\psi_1\psi_1}\,, \\ K_{12} = {} & (I-L_{11})^{-1}(L_{12}K_{22} + K_{\psi_1,2})\,. \end{aligned}$$

- Covariance models with high-order closures:

PROPOSITION 2. *Consider a process driven by $y_1 = g(y_2) + \eta$. Then under the third-order closure assumptions, its covariance matrix takes the following form:*

$$\mathrm{Cov}\left([y_1^T, y_2^T]^T\right) = \begin{bmatrix} K_{11} & LK_{22} + \mathrm{Cov}(\eta, y_2) \\ K_{22}L^T + \mathrm{Cov}(y_2,\eta) & K_{22} \end{bmatrix} + \mathcal{O}\left(\delta y_2^3\right),$$

where L is the Jacobian matrix, H is the Hessian tensor corresponding to g evaluated at $\mathrm{E}\{y_2\}$, and

$$K_{11} = K_{11}^I + \frac{1}{4}\overline{\delta y_2^T H \delta y_2 \delta y_2^T H^T \delta y_2} - \frac{1}{4}\mathrm{tr}(HK_{22})\mathrm{tr}(HK_{22})^T + \frac{1}{2}\overline{\delta y_2^T H \delta y_2 \eta^T} + \frac{1}{2}\overline{\delta\eta\delta y_2^T H^T \delta y_2}\,,$$

where K_{11}^I is the corresponding term using second-order closure assumptions.

Auto- and cross-covariance functions

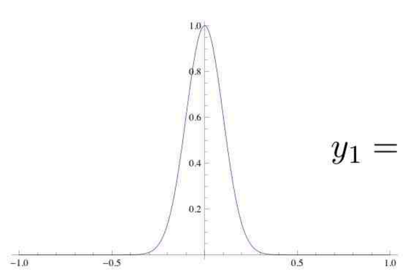
- Squared exponential function and the generated auto- and cross-covariance functions

$K_{22} = \mathrm{Cov}(y_2, y_2)$ squared exponential

$K_{11} = \mathrm{Cov}(y_1, y_1)$ auto-covariance

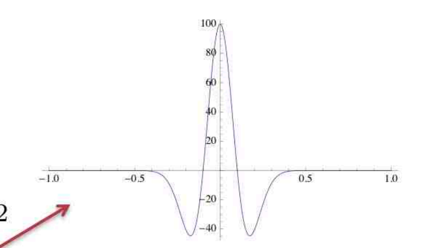
$K_{12} = \mathrm{Cov}(y_1, y_2)$ cross-covariance

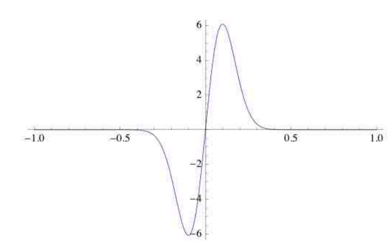
$$k(d) = \sigma^2 \exp\left(-\frac{1}{2}\frac{d^2}{\ell^2}\right)$$



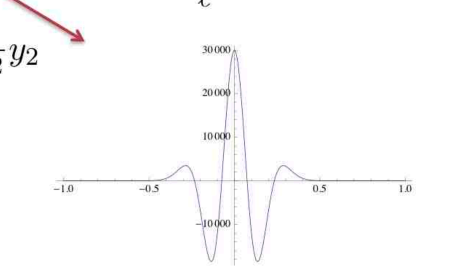
$$y_1 = \frac{\partial}{\partial x}y_2$$

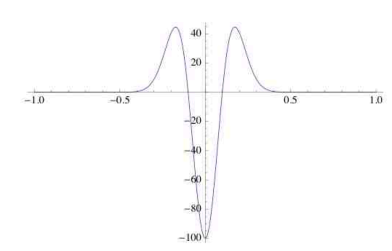
$$y_1 = \frac{\partial^2}{\partial x^2}y_2$$





$$\sigma^2\frac{d^2-\ell^2}{\ell^4}e^{-\frac{d^2}{2\ell^2}} \qquad \sigma^2\frac{d^4-6d^2\ell^2+3\ell^4}{\ell^8}e^{-\frac{d^2}{2\ell^2}}$$





$$\sigma^2\frac{d}{\ell^2}e^{-\frac{d^2}{2\ell^2}} \qquad -\sigma^2\frac{d^2-\ell^2}{\ell^4}e^{-\frac{d^2}{2\ell^2}}$$

- Auto- and cross-covariance functions can also be derived for nonlinear processes or Matérn covariance kernels

$$K_{11} = \mathrm{Cov}(y_1, y_1)$$

$$k(d) = \sigma^2\frac{2^{1-\frac{\nu}{2}}}{\Gamma(\nu)}\left(\frac{d\sqrt{\nu}}{\ell}\right)^{\nu}K_\nu\left(\frac{d\sqrt{2\nu}}{\ell}\right) \rightarrow \sigma^2\frac{2^{1-\frac{\nu}{2}}\nu}{\ell^3\Gamma(\nu)}\left(\frac{d\sqrt{\nu}}{\ell}\right)^{\nu-1}\left(\sqrt{2}\ell K_{\nu-1}\left(\frac{d\sqrt{2\nu}}{\ell}\right) - 2d\sqrt{\nu}K_{\nu-2}\left(\frac{d\sqrt{2\nu}}{\ell}\right)\right)$$

$$y_1 = \frac{\partial}{\partial x}y_2 \qquad K_{12} = \mathrm{Cov}(y_1, y_2)$$

$$\sigma^2\frac{2^{\frac{3}{2}-\frac{\nu}{2}}}{d\Gamma(\nu)}\left(\frac{d\sqrt{\nu}}{\ell}\right)^{\nu+1}K_{\nu-1}\left(\frac{d\sqrt{2\nu}}{\ell}\right)$$

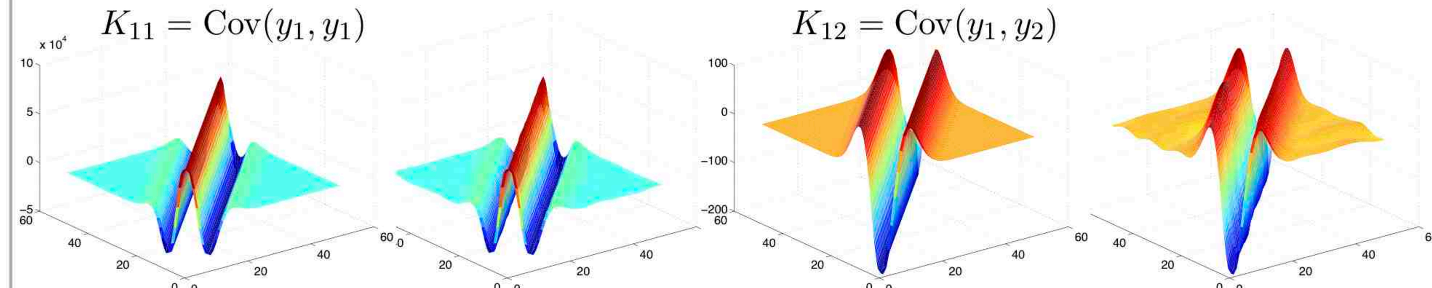
Validation of the auto- and cross-covariance

- Model and sample-based auto- and cross-covariance functions: $y_1 = \frac{\partial^2}{\partial x^2}y_2$

$$y_1 = u(x) = g(y_2(x)) = \alpha\tfrac{\partial}{\partial x}y_2 + \eta,\ y_2 \sim \mathcal{N}(0, C_{se}([\ell_2, \sigma_2^2])),\ \eta \sim \mathcal{N}(0, C_{mat,\nu=5/2}([\ell_\eta, \sigma_\eta^2]))$$

- Auto- and cross-covariance matrices generated by this process:

$K_{11} = \mathrm{Cov}(y_1, y_1)$ $\qquad$ $K_{12} = \mathrm{Cov}(y_1, y_2)$



model $\quad$ 10,000 samples $\quad$ model $\quad$ 10,000 samples

Gaussian process regression with multiple outputs

- Consider the following process: $y_1 = g(y_2) + \eta$
- The joint distribution conf. Proposition 2 is approximated by

$$\begin{bmatrix} y_1 \\ y_2 \\ y_{*1} \\ y_{*2} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} g(\overline{y_2}) + \frac{1}{2}\mathrm{tr}(HK_{22}) \\ \overline{y_2} \\ g(\overline{y_{*2}}) + \frac{1}{2}\mathrm{tr}(HK_{*2,*2}) \\ \overline{y_{*2}} \end{bmatrix}, \right.$$

(observed states: first two entries; predicted states [sub star]: last two entries)

$$\left. \left[\begin{array}{cc|cc} LK_{22}L^T + K_{\eta\eta} & LK_{22} & LK_{2,*2}L^T + K_{\eta,*\eta} & LK_{*2,*2} \\ K_{22}L^T & K_{22} & K_{2,*2}L^T & K_{2,*2} \\ \hline LK_{*2,2}L^T + K_{*\eta,\eta} & LK_{*2,2} & LK_{*2,*2}L^T + K_{*\eta,*\eta} & LK_{*2,*2} \\ K_{*2,*2}L^T & K_{2,*2}^T & K_{*2,*2}L^T & K_{*2,*2} \end{array}\right] + \left[\begin{array}{c|c} \Sigma & 0 \\ \hline 0 & 0 \end{array}\right] \right)$$

(observational error)

$$\Sigma = \begin{bmatrix} K_{\varepsilon_1\varepsilon_1} & 0 \\ 0 & K_{\varepsilon_2\varepsilon_2} \end{bmatrix} = \begin{bmatrix} \sigma_{n,1}^2 I & 0 \\ 0 & \sigma_{n,2}^2 I \end{bmatrix}$$

- The predictive distribution is given by

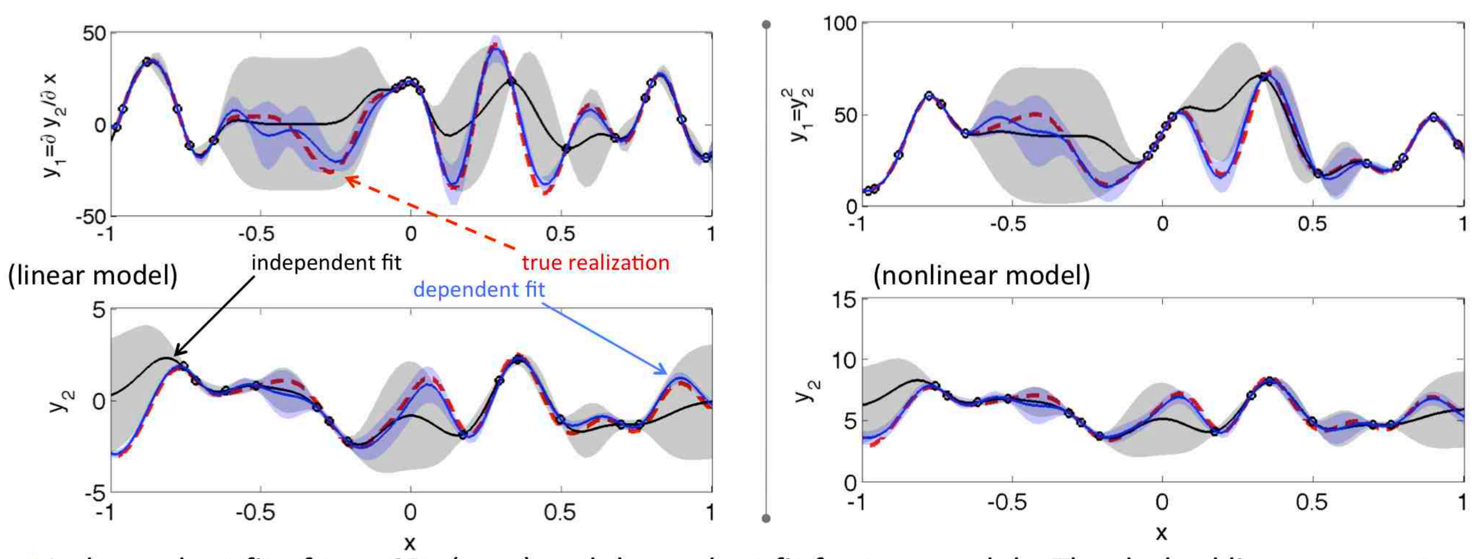
$$\begin{aligned} \overline{\mathbf{y}_*|\mathbf{X}, \mathbf{X}_*, \mathbf{y}} &= \mathbf{m}(\mathbf{X}_*) + \mathbf{K}_{21}\left(\mathbf{K}_{11} + \Sigma\right)^{-1}\left(\mathbf{y} - \mathbf{m}(\mathbf{X})\right)\,, \\ \mathrm{Cov}(\mathbf{y}_*|\mathbf{X}, \mathbf{X}_*, \mathbf{y}) &= \mathbf{K}_{22} - \mathbf{K}_{21}\left(\mathbf{K}_{11} + \Sigma\right)^{-1}\mathbf{K}_{12} \end{aligned}$$

- Hyperparameters are obtained by maximizing the likelihood function:

$$\log(\mathcal{P}(\mathbf{y}|\mathbf{X}, \theta)) = -\frac{1}{2}\left(\mathbf{y} - \mathbf{m}(\mathbf{X})\right)^T\left(\mathbf{K}_{11} + \Sigma\right)^{-1}\left(\mathbf{y} - \mathbf{m}(\mathbf{X})\right) - \frac{1}{2}\log|\mathbf{K}_{11} + \Sigma| - \frac{n_2}{2}\log(2\pi)$$

One-dimensional numerical experiments

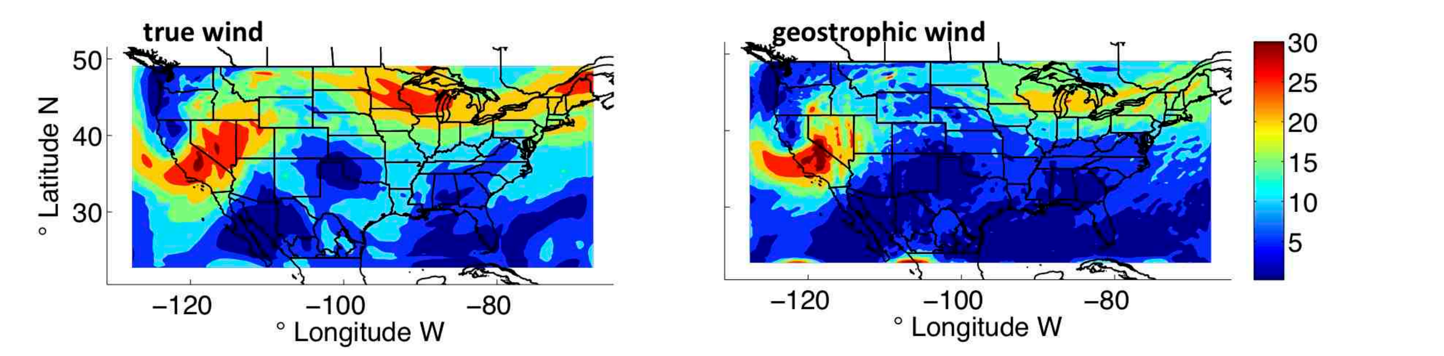
- The fit of two fields that are related through some functional (w/ and w/o considering it)



- Independent fit of two GPs (grey) and dependent fit for two models. The dashed line represents the ``truth,'' with noisy observations denoted by circles. The solid dark line represents the independent fit and the grey shade the point variance. The blue solid line with blue shade represents the dependent fit.

Application: Two-dimensional co-kriging for geostrophic wind

- Geostrophic wind: $u_g = -\alpha_u\dfrac{\partial\phi_p(x,y)}{\partial y}\,,\quad v_g = \alpha_v\dfrac{\partial\phi_p(x,y)}{\partial x}$



- Physical model:

$$\mathbf{y} = [\underbrace{u_g, v_g}_{\mathbf{y}_1}, \underbrace{\phi}_{\mathbf{y}_2}]^T; \quad \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} u_g \\ v_g \end{bmatrix} + \Sigma\,, \quad \text{obs err.: } \Sigma = \mathcal{N}(m_{uv}, \mathbf{K}_{uv})$$

$$\mathbf{y}_1 = \begin{bmatrix} u_g \\ v_g \end{bmatrix} = \begin{bmatrix} (-L_y \otimes I_x) \\ (I_y \otimes L_x) \end{bmatrix}(I_2 \otimes \phi) = \begin{bmatrix} L_u\phi \\ L_v\phi \end{bmatrix} = g(\mathbf{y}_2) = g(\phi)\,,$$

- Statistical model:

hyperparameters: $\theta = \{\ell_2,\, \ell_\eta,\, \sigma_2^2,\, \sigma_\eta^2,\, \sigma_{n,1}^2,\, \sigma_{n,2}^2\}$

$$\begin{aligned} &\phi \sim \mathbf{m}_\phi + \mathcal{M}_{\nu_\phi}(\ell_\phi, \sigma_\phi^2)\,, \\ &U = L_u\phi + \eta\,, \quad \mathbf{m_u} = L_u\mathbf{m}_\phi + \mathbf{m}_\eta\,, \quad \eta \sim \mathcal{M}_{\nu_\eta}(\ell_\eta, \sigma_\eta^2)\,, \\ &V = L_v\phi + \nu\,, \quad \mathbf{m_v} = L_v\mathbf{m}_\phi + \mathbf{m}_\nu\,, \quad \nu \sim \mathcal{M}_{\nu_\nu}(\ell_\nu, \sigma_\nu^2)\,, \end{aligned}$$

$$\mathbf{K}_{11} = \begin{bmatrix} L_{*u}K_{*\phi,*\phi}L_{*u}^T + K_{\eta\eta} + K_{\varepsilon_u\varepsilon_u} & K_{uv} & L_{*u}K_{*\phi,*\phi}H_{\phi,*\phi}^T \\ K_{vu} & L_{*v}K_{*\phi,*\phi}L_{*v}^T + K_{\nu\nu} + K_{\varepsilon_v\varepsilon_v} & L_{*v}K_{*\phi,*\phi}H_{\phi,*\phi}^T \\ H_{\phi,*\phi}K_{*\phi,*\phi}L_{*u}^T & H_{\phi,*\phi}K_{*\phi,*\phi}L_{*v}^T & K_{\phi,\phi} + K_{\varepsilon_\phi\varepsilon_\phi} \end{bmatrix},$$

$$\mathbf{K}_{12} = \begin{bmatrix} L_{*u}K_{*\phi,*\phi}L_u^T + K_{\eta,*\eta} & K_{u,*v} & L_{*u}K_{*\phi,*\phi} \\ K_{v,*u} & L_{*v}K_{*\phi,*\phi}L_v^T + K_{\nu,*\nu} & L_{*v}K_{*\phi,*\phi} \\ H_{\phi,*\phi}K_{*\phi,*\phi}L_u^T & H_{\phi,*\phi}K_{*\phi,*\phi}L_v^T & K_{\phi,*\phi} \end{bmatrix}, \quad \mathbf{K}_{21} = \mathbf{K}_{12}^T\,,$$

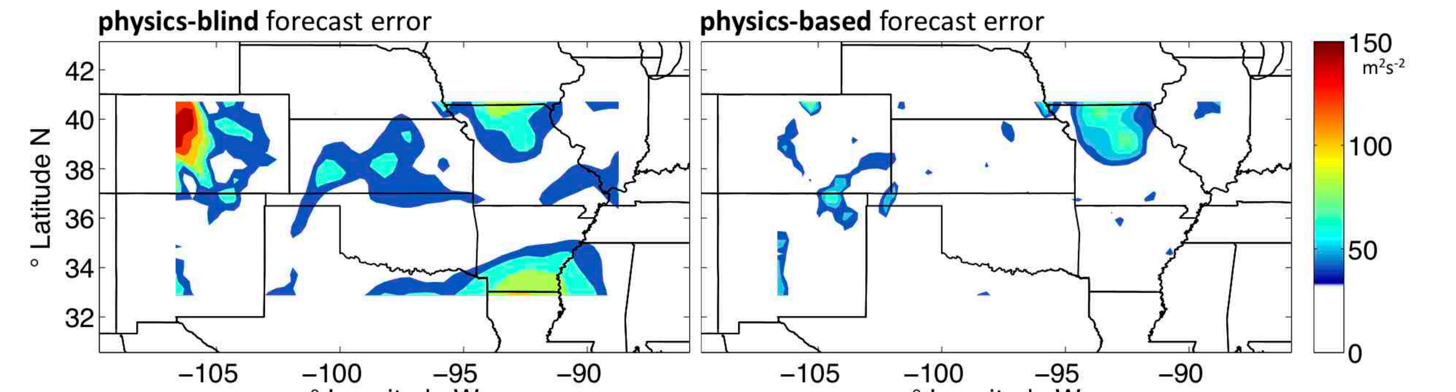
$$\mathbf{K}_{22} = \begin{bmatrix} L_uK_{*\phi,*\phi}L_u^T + K_{*\eta,*\eta} & K_{*u,*v} & L_uK_{*\phi,*\phi} \\ K_{*v,*u} & L_vK_{*\phi,*\phi}L_v^T + K_{*\nu,*\nu} & L_vK_{*\phi,*\phi} \\ (L_uK_{*\phi,*\phi})^T & (L_vK_{*\phi,*\phi})^T & K_{*\phi,*\phi} \end{bmatrix} \qquad K_{uv} = L_{*u}K_{*\phi,*\phi}L_{*v}$$

- Predictive RMSE at unobserved sites:

Observations: 0.15% (East) U & V, and 0.25% (West) ϕ

	Calibration sample			Validation sample		
Model [observations]	u_*	v_*	ϕ_*	u_*	v_*	ϕ_*
Exact model [y1y2]	2.20	1.85	3.23	2.20	2.00	2.98
Independent [y1]	4.08	3.37	-	2.94	3.09	-
Independent [y1y2]	4.08	3.28	5.77	2.91	2.89	4.45
Dependent [y1]	2.98	2.57	-	2.39	2.42	-
Dependent [y1y2]	2.33	1.97	3.85	2.32	2.18	3.74

- Geopotential prediction: OK, TX region yields a better fit when using the physics-based model



References

E.M. Constantinescu and M. Anitescu, **Physics-based covariance models for Gaussian processes with multiple outputs.** International Journal for Uncertainty Quantification, in press, 2011.

E.M. Constantinescu, T. Chai, A. Sandu, and G.R. Carmichael, **Autoregressive models of background errors for chemical data assimilation**; Journal of Geophysical Research, Vol. 112, 2007, pp. 1-14.

*Coauthor: Mihai Anitescu[ANL]; work supported by the Scalable Statistical Analysis of Gaussian Models for Petascale Spatiotemporal Data project through the U.S. Department of Energy under Contract No. DE-AC02-06CH11357

U.S. DEPARTMENT OF ENERGY