A generalized, compactly-supported correlation function for data assimilation applications

Shay Gilpin[1], Tomoko Matsuo[2,1], Stephen E. Cohn[3]

[1]Department of Applied Mathematics, University of Colorado, Boulder, [2]Smead Aerospace Engineering Sciences, University of Colorado, Boulder, [3]Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, Maryland

AMS Annual Meeting, January 10, 2023

Outline

1. Present a new application of forecasting (evolving) estimation error covariances during data assimilation,
2. Introduce a new correlation function that generalizes the Gaspari-Cohn compactly-supported approximation to a Gaussian,
3. Apply this generalized correlation function for evolving covariances.

Motivation: Covariance Evolution in Data Assimilation

At the heart of modern data assimilation is *covariance evolution*, *e.g.*,

$$\boldsymbol{P}_{k+1} = \boldsymbol{M}_{k+1,k} \boldsymbol{P}_k \boldsymbol{M}_{k+1,k}^T + \boldsymbol{Q}_k.$$

Common Issues with Covariance Evolution:

- Computational Expense
- Variance Loss

Covariance Evolution in Data Assimilation

At the heart of modern data assimilation is *covariance evolution*, e.g.,

$$\boldsymbol{P}_{k+1} = \boldsymbol{M}_{k+1,k}\boldsymbol{P}_k\boldsymbol{M}_{k+1,k}^T.$$

Common Issues with Covariance Evolution:

- ~~Computational Expense~~ $\rightarrow$ Computationally efficient
- ~~Variance Loss~~ $\rightarrow$ Evolves variance directly

Alternative: evolve the variance and correlation length, then reconstruct $\boldsymbol{P}$ using a parametric correlation function.

Local Covariance Evolution

Consider covariances $P = P(x_1, x_2, t)$ associated with states $q = q(x, t)$ on the unit circle (S_1^1),

$$q_t + vq_x + v_x q = 0, \qquad P_t + v_1 P_{x_1} + v_2 P_{x_2} + (v_{x_1} + v_{x_2})P = 0,$$
$$q(x, t_0) = q_0(x) \qquad P(x_1, x_2, t_0) = P_0(x_1, x_2)$$

Local Covariance Evolution

Consider covariances $P = P(x_1, x_2, t)$ associated with states $q = q(x, t)$ on the unit circle (S_1^1),

$$q_t + vq_x + v_x q = 0, \qquad P_t + v_1 P_{x_1} + v_2 P_{x_2} + (v_{x_1} + v_{x_2})P = 0,$$
$$q(x, t_0) = q_0(x) \qquad P(x_1, x_2, t_0) = P_0(x_1, x_2)$$

Variance Equation:	Correlation Length Equation:
$\sigma_t^2 + v\sigma_x^2 + 2v_x\sigma^2 = 0,$	$L_t + vL_x - v_x L = 0,$
$\sigma^2(x, t_0) = \sigma_0^2(x)$	$L(x, t_0) = L_0(x)$

Local Covariance Evolution

Consider covariances $P = P(x_1, x_2, t)$ associated with states $q = q(x, t)$ on the unit circle (S_1^1),

$$q_t + vq_x + v_x q = 0, \qquad P_t + v_1 P_{x_1} + v_2 P_{x_2} + (v_{x_1} + v_{x_2})P = 0,$$
$$q(x, t_0) = q_0(x) \qquad P(x_1, x_2, t_0) = P_0(x_1, x_2)$$

Variance Equation:	Correlation Length Equation:
$\sigma_t^2 + v\sigma_x^2 + 2v_x\sigma^2 = 0,$	$L_t + vL_x - v_x L = 0,$
$\sigma^2(x, t_0) = \sigma_0^2(x)$	$L(x, t_0) = L_0(x)$

1. Evolve σ^2 and L from initial condition $P_0(x_1, x_2) = \sigma_0(x_1)C_0(x_1, x_2)\sigma_0(x_2)$,
2. Approximate $P(x_1, x_2, t) = \sigma(x_1, t)C(x_1, x_2, t)\sigma(x, t)$ with evolved σ^2 and L using a parametric correlation function.

The Gaspari and Cohn (1999) Correlation Function

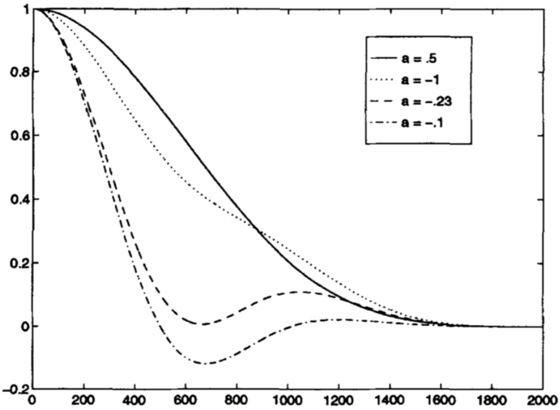


Figure 7. The function $C_0(z, a, c)$ of the example in section 4(c) for $c = 1000$ km and various values of a. See text for explanation.

Figure 1: Figure 7 from Gaspari and Cohn (1999). The function $C_0(z, a, c)$ is the general form of the compactly-supported, fifth-order, piecewise rational correlation function derived in Sec. 4(c). Typically, $a = 1/2$ (solid black).

The Generalized Gaspari-Cohn (GenGC) Correlation Function

Correlation length for the compactly-supported, piecewise rational:

$$L = c\left(\frac{3(22a^2+3a+1)}{40(8a^2-2a+1)}\right)^{1/2}, \quad a = 1/2 \Rightarrow L = \sqrt{0.3}c \qquad (1)$$

Need to generalize this correlation function to allow for *variable* L.

The Generalized Gaspari-Cohn (GenGC) Correlation Function

Correlation length for the compactly-supported, piecewise rational:

$$L = c\left(\frac{3(22a^2+3a+1)}{40(8a^2-2a+1)}\right)^{1/2}, \quad a = 1/2 \Rightarrow L = \sqrt{0.3}c \qquad (1)$$

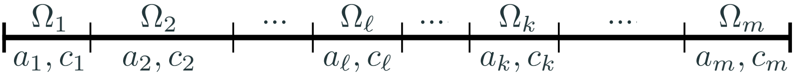
Need to generalize this correlation function to allow for *variable* L.

Generalized Gaspari-Cohn

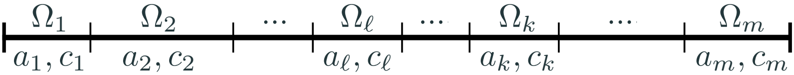
Allow $a = a_k$ and $c = c_k$ to vary over the spatial index k.

Now $L = L_k$ can vary!

The GenGC Correlation Function, cont.



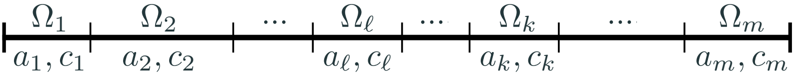
The GenGC Correlation Function, cont.



For $r \in \Omega_k$, $s \in \Omega_\ell$, and each fixed $k, \ell = 1, 2, ..., m$,

GenGC: $C_{k\ell}(r, s; a_k, a_\ell, c_k, c_\ell)$

The GenGC Correlation Function, cont.



For $r \in \Omega_k$, $s \in \Omega_\ell$, and each fixed $k, \ell = 1, 2, ..., m$,

GenGC: $C_{k\ell}(r, s; a_k, a_\ell, c_k, c_\ell)$

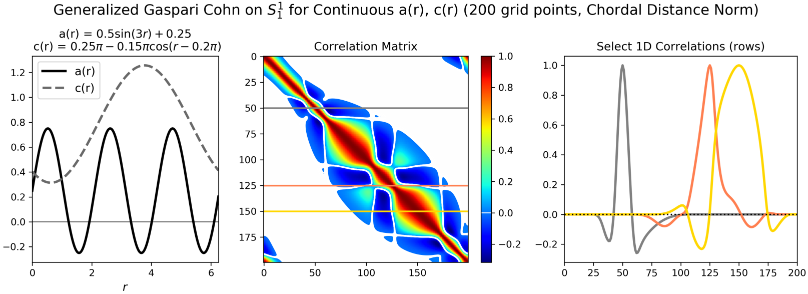


Figure 2: GenGC on the unit circle (S_1^1) for continuous a and c (left), correlation matrix (middle), and selected correlations (right). White regions in the correlation matrix correspond to correlations between -0.003 and 0.003.

GenGC in 2D

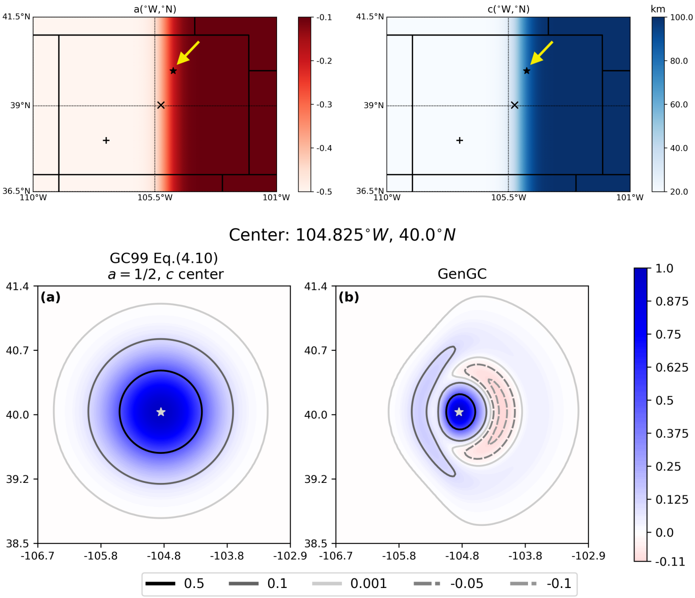


Figure 3: Top: Background a and c fields as functions of longitude and latitude over Colorado. Bottom: 2D correlations with respect to $\star$ for the given a and c fields, plotted using a Mercator projection.

Local Covariance Evolution: Correlation Reconstruction with GenGC

Methodology:

1. Define the initial correlation length from standard Gaspari-Cohn correlation ($a_0 = 1/2$, $c = c_0$), $L_0(x) = \sqrt{0.3}c_0$

Local Covariance Evolution: Correlation Reconstruction with GenGC

Methodology:

1. Define the initial correlation length from standard Gaspari-Cohn correlation ($a_0 = 1/2$, $c = c_0$), $L_0(x) = \sqrt{0.3}c_0$
2. Evolve $L(x, t)$ forward in time by solving its PDE

Local Covariance Evolution: Correlation Reconstruction with GenGC

Methodology:

1. Define the initial correlation length from standard Gaspari-Cohn correlation ($a_0 = 1/2$, $c = c_0$), $L_0(x) = \sqrt{0.3}c_0$
2. Evolve $L(x,t)$ forward in time by solving its PDE
3. At fixed time t, use $L(x,t)$ and $a = a_0$ fixed to compute $c(x,t)$, then reconstruct the correlations with GenGC: for $x_1 \in \Omega_k, x_2 \in \Omega_\ell$

$$C(x_1, x_2, t) = C_{k\ell}\big(x_1, x_2; a_0, a_0, c(x_1, t), c(x_2, t)\big)$$

Correlation Reconstruction with GenGC: Preliminary Results

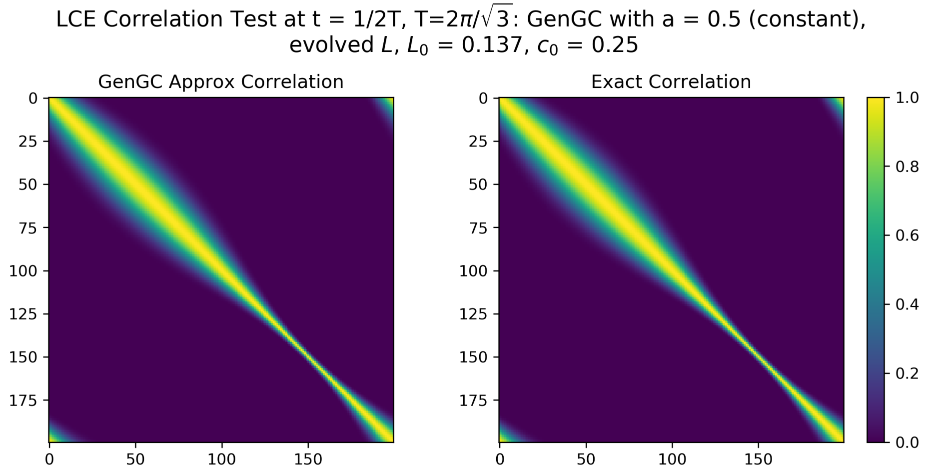


Figure 3: Comparisons of the correlation matrix approximated with GenGC using evolved correlation lengths L and $a = 1/2$ constant (left) with the exact correlation matrix (right). $1D$ experiments on S_1^1 with $v(x) = \sin(x) + 2$ and $c_0 = 0.25$. Errors in the GenGC approximation are between -0.0125 and 0.0125.

In Summary

The added flexibility of GenGC while remaining compactly supported can be a useful tool for the data assimilation community.

Applications of GenGC in Data Assimilation:

- Local covariance evolution, an alternative means of mitigating problems associated with covariance evolution,
- Covariance modeling and tapering (localization) in current data assimilation schemes.

Acknowledgements and Questions

For more information or further discussion, contact Shay at

shay.gilpin@colorado.edu

Look out for our paper,

A generalized, compactly-supported correlation function for data assimilation applications

coming soon to the Quarterly Journal of the Royal Meteorological Society.

The presenter would like to thank the National Science Foundation (NSF) for supporting this work through the NSF Graduate Research Fellowship.

Extra slides

1D Example

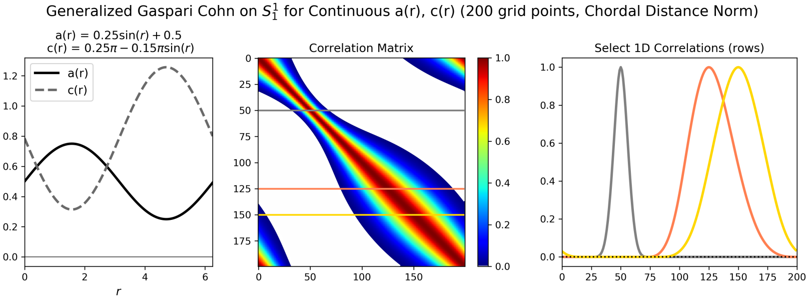


Figure 4: GenGC on the unit circle for continuous a and c (left), correlation matrix (middle), and selected correlations (right). White regions in the correlation matrix correspond to correlations between -0.003 and 0.003.

Correlation Reconstruction with GenGC: Preliminary Results 2

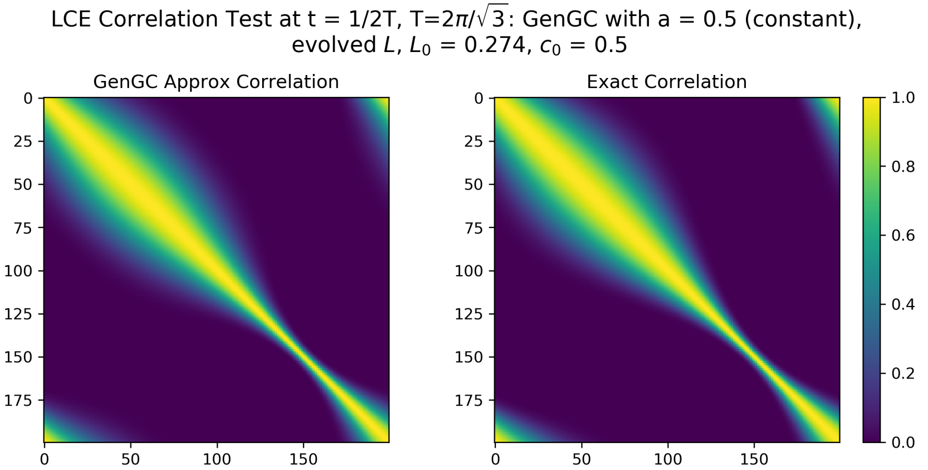


Figure 5: Comparisons of the correlation matrix approximated with GenGC using evolved correlation lengths L and $a = 1/2$ constant (left) with the exact correlation matrix (right). $1D$ experiments on S^1_1 with $v(x) = \sin(x) + 2$ and $c_0 = 0.5$. Errors in the GenGC approximation are between -0.05 and 0.05.

Continuity

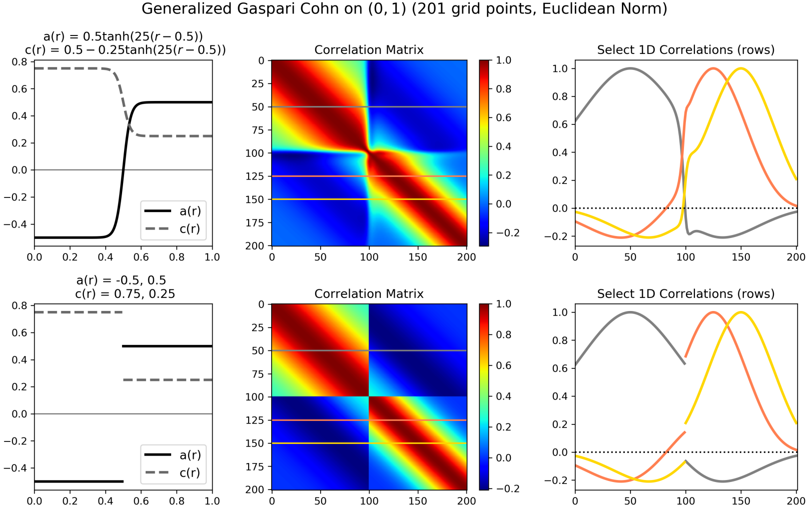


Figure 6: Example of GenGC on $(0, 1)$ comparing when each grid cell is its own subregion vs. the case where $(0, 1)$ is split into two subregions $(0, 1/2), (1/2, 1)$ (bottom row)