Adaptive Localization for the Ensemble-Based Observation Impact Estimate Using Regression Confidence Factors

NICHOLAS A. GASPERONI AND XUGUANG WANG
School of Meteorology, and Center for Analysis and Prediction of Storms, University of Oklahoma, Norman, Oklahoma

(Manuscript received 25 August 2014, in final form 9 January 2015)

ABSTRACT

The goal of this study is to improve an ensemble-based estimation for forecast sensitivity to observations that is straightforward to apply using existing products of any ensemble data assimilation system. Because of limited ensemble sizes compared to the large degrees of freedom in typical models, it is necessary to apply localization techniques to obtain accurate estimates. Fixed localization techniques do not guarantee accurate impact estimates, because as forecast time increases the error correlation structures evolve with the flow. Here a dynamical localization method is applied to improve the observation impact estimate. The authors employ a Monte Carlo “group filter” technique to limit the effects of sampling error via regression confidence factor (RCF). Experiments make use of the local ensemble transform Kalman filter (LETKF) with a simple two-layer primitive equation model and simulated observations. Results show that the shape, location, time dependency, and variable dependency of RCF localization functions are consistent with underlying dynamical processes of the model. Application of RCF localization to ensemble-estimated impact showed marked improvement especially for longer forecasts and at midlatitudes, when systematically verified against actual impact in RMSE and skill scores. The impact estimates near the equator were not as effective because of large discrepancies between the RCF function and the localization used at assimilation time. These latter results indicate that there exists an inherent relationship between the localization applied during the assimilation time and the proper localization choice for observation impact estimates. Application of RCF for automatically tuned localization is introduced and tested for a single observation experiment.

1. Introduction

Data assimilation of observations in both time and space improves a numerical weather prediction (NWP) forecast, in an average sense. However, it is important to determine the value added to a forecast from a specific subset of observations. In this way, we can investigate by instrument type, observation type, and location, which observations are the most impactful on a forecast. Additionally, we can avoid using observations that have negative impacts on a forecast. Evaluating the usefulness of observations is particularly important for operational NWP centers that operate under limited budgets and need to weigh the costs and benefits of adding more observations to an already large observational dataset.

There are a few basic approaches to quantifying the impact that assimilated observations have on a forecast. The first is the straightforward data-denial method, where parallel sets of analysis and forecast experiments are conducted with a “control” experiment assimilating all observations and experiments withholding subsets of observations (e.g., Zapotocny et al. 2002, 2007; Benjamin et al. 2010; Kutty and Wang 2015, manuscript submitted to Adv. Meteor.). This approach is popular; however, it is computationally expensive because of the number of experiments required. A second approach is an adjoint-based approach, first explored by Langland and Baker (2004). Different from data denial, this method can provide observation impact estimates for all observations simultaneously, without the need for separate denial experiments. The adjoint method has been applied successfully as an important diagnostic tool (e.g., Cardinali 2009; Langland et al. 2009; Gelaro et al. 2010; Weissmann et al. 2012; Hamill et al. 2013). However, adjoints are generally difficult to create, and because of the tangent corrections...
linear assumption their application is limited to shorter forecast lengths.

A third approach to evaluating observation impact on a forecast is the ensemble-based method. An approach analogous to the adjoint method of Langland and Baker (2004) was proposed by Liu and Kalnay (2008) with a minor correction in Li et al. (2010). Kunii et al. (2012) successfully applied their method by evaluating the impact of real observations in a forecast of Typhoon Sinlaku, using the Weather Research and Forecasting (WRF) Model together with the local ensemble transform Kalman filter (LETKF; Hunt et al. 2007). Kalnay et al. (2012) derived a simpler formulation that makes fewer approximations; it is more general and computationally efficient because it relies on readily available ensemble Kalman filter (EnKF) products and can be used with any deterministic EnKF method. Ota et al. (2013) successfully applied the Kalnay et al. (2012) formulation to the National Centers for Environmental Prediction (NCEP) Global Forecast System EnKF (GFS/EnKF; Whitaker et al. 2008), which is now part of the GFS hybrid data assimilation system (Wang et al. 2013).

The Kalnay et al. (2012) observation impact metric is appealing because ensemble perturbations take the place of the adjoint model in estimating sensitivities. However as with any ensemble method, it suffers from sampling error, which occurs when the number of ensemble members is small compared to the degrees of freedom in a model and observing system—predominantly the case in ensemble NWP. Sampling error results in spurious correlations and can lead to filter divergence in deterministic EnKF assimilation. Houtekamer and Mitchell (1998) showed that the effects of sampling error can be suppressed by excluding distant observations from influencing the analysis at a given grid point. They experimented filtering covariance estimates using a distance-dependent correlation function, referred to as covariance localization (Houtekamer and Mitchell 2001). Since then, much research has been done in developing localization methods to improve EnKF analyses using limited ensembles (e.g., Hamill et al. 2001; Houtekamer and Mitchell 2005; Anderson 2007; Kept 2011; Anderson 2012; Anderson and Lei 2013; Holland and Wang 2013).

In many cases, localization may be nontrivial, such as the case for applying localization to observations with complex forward operators (e.g., radar, satellite observations), or when localizing the impact of observations on variables not directly linked to the observations (e.g., localizing the impact of a temperature observation on the analysis of wind). Several adaptive localization methods have been developed to account for these issues. Anderson (2007) employed a Monte Carlo statistical method known as the group filter (GF) for evaluating sampling error using groups of ensemble members and calculating a weighting coefficient that minimizes root-mean-square (RMS) differences between group regression coefficients. Studies such as Chen and Oliver (2010), Bishop and Hodyss (2009a,b), and Anderson (2012) compute localization based on correlations between an observation and a state variable. More recently, Anderson and Lei (2013) developed an empirical localization function (ELF), which computes localizations from the output of an observing system simulation experiment in sets of pairs of observation and state variables binned by distance. Lei and Anderson (2014) compared the ELF to the GF method of Anderson (2007). In an ideal simulation where the true covariances are known, the ELF and GF show similar results, especially for larger ensembles. The ELF shows benefits over the GF in cases where there are biases in the spurious covariances and in the ability to automatically inflate covariances. However, extension of the ELF to real cases poses challenges, particularly because the ELF relies on the knowledge of a true state (Anderson and Lei 2013).

The problem of sampling error caused by small ensembles is a more serious issue for the ensemble-based observation impact metric. Localization can be applied to alleviate sampling error; however, a time-forecast component is added to the localization problem, such that a straightforward application of fixed localization techniques would not guarantee accurate impact estimates. To partially address the issue, Kalnay et al. (2012) proposed two methods of moving localization: 1) using a model-forecast nonlinear incremental evolution of the localization function, and 2) advecting the localization center using the climatological group velocity of dominant wavenumbers. Ota et al. (2013) applied a similar advected localization method, using the average forecast horizontal wind at each model vertical level. Both studies showed improvement relative to fixed GC localization; however, possible limitations exist with each method. The nonlinear evolution of the localization is computationally prohibitive for real NWP systems because a forecast is required for every observation. The advection methods are simpler to implement; however, they assume that the optimal localization is tied to the mean flow of the model and does not change in magnitude, size, or shape.

This study explores a dynamic localization method for ensemble-based observation impact estimate. There are two main purposes of this study. The first is to simply learn what a “proper” localization function looks like for the impact estimate, and how it evolves with increasing forecast time. Kalnay et al. (2012) and Ota et al. (2013) only considered moving the localization function, but it is possible that the shape and magnitude also need to evolve with forecast component, something that an adaptive method will be able to automatically
determine. The second purpose is to test the potential effectiveness of an adaptive method on the observation impact metric. Adaptive localizations generally have been shown to provide more accurate assimilations at increased computational cost relative to fixed localization; in this study it is explored whether the application of such adaptive methods initially developed for assimilation can be extended and applied successfully to the ensemble impact metric. Briefly, the dynamic localization function used here is obtained from confidence factors derived using groups of ensembles, first proposed for EnKF by Anderson (2007). The focus of this study is to extend the GF concept in the context of the observation impact estimate.

In section 2, the observation impact metric is described along with the dynamic localization method, referred throughout as regression confidence factor (RCF) localization. Section 3 describes the experiment setup. As an initial test of the method on observation impact estimates, an isentropic two-layer primitive equation model (Zou et al. 1993) under the perfect-model assumption is adopted. This model is coupled with the LETKF data assimilation system following Holland and Wang (2013). The RCF calculation settings are also explained in section 3. The resulting RCF localization functions are shown and applied in section 4 for both single-observation and full-domain observation assimilation experiments. The accuracy of the ensemble observation impact estimate using the dynamic localization is compared with that of using fixed Gaspari–Cohn [GC; see Eq. (4.10) in Gaspari and Cohn (1999)] localization for different observation locations (tropical vs midlatitude), forecast length, and differing state variables. The goal of comparison to static GC is to first determine if the adaptive method is better for the ensemble impact metric. More importantly, the GC function is used as a tool to provide a baseline for comparisons with RCF localization experiments, to help provide context for qualitative discussion. From this comparison, an important relationship between localization used during assimilation and for the impact estimate was discovered and is discussed in section 4. Another potential use for adaptive methods is in the ability to automatically tune GC localization. This concept is introduced and tested in section 4. A summary and discussion are given in section 5.

2. Methods

a. The observation impact metric

Following Kalnay et al. (2012), let \( \mathbf{x}_{t|0} \) represent the ensemble mean analysis and \( \mathbf{x}_{t|0}^f \) the deterministic forecast launched from the mean analysis (subscript “\( t|0 \)” can be read as “valid at time \( t \), initialized from analysis at time \( 0 \)”). A cost function \( J \) is defined in Langland and Baker (2004) to be the actual forecast error reduction (i.e., the difference in squared error between two adjacent forecasts):

\[
J = \mathbf{e}_{t|0}^T \mathbf{e}_{t|0} - \mathbf{e}_{t|n}^T \mathbf{e}_{t|n} = (\mathbf{e}_{t|0} - \mathbf{e}_{t|n})^T (\mathbf{e}_{t|0} - \mathbf{e}_{t|n}),
\]

(1)

where \( \mathbf{e}_{t|0} = \mathbf{x}_{t|0} - \mathbf{x}_{t|0}^f \) and \( \mathbf{e}_{t|n} = \mathbf{x}_{t|n} - \mathbf{x}_{t|n}^f \) are the errors from forecasts initialized at time \( t = 0 \) and \( t = n \), respectively. Here \( \mathbf{x}_{t}^f \) is the truth valid at time \( t \); in the absence of the true state, a verifying analysis can be used. The differences in forecast errors \( \mathbf{e}_{t|0} \) and \( \mathbf{e}_{t|n} \) are due to the assimilation of observations at time \( t = 0 \), so (1) represents the impact of assimilating observations on a forecast [see Fig. 1 in Langland and Baker (2004)]. When \( J \) is negative (positive), the magnitude of error in \( \mathbf{e}_{t|0} \) is less (greater) than the magnitude of error in \( \mathbf{e}_{t|n} \), which can be interpreted as positive (negative) impact.

It is shown in Kalnay et al. (2012) that (1) can be rewritten in ensemble form as

\[
J = \frac{1}{K-1} \delta \mathbf{y}_0^T \mathbf{R}^{-1} \mathbf{H} \mathbf{X}_{t|0}^f (\mathbf{e}_{t|0} + \mathbf{e}_{t|n}),
\]

(2)

where \( \delta \mathbf{y}_0 = \mathbf{y}_0 - \mathbf{H} (\mathbf{x}_{t|0}^f) \) is the observation innovation vector [i.e., the difference between the observations, \( \mathbf{y}_0 \), assimilated at time 0 and the mean background interpolated to observation space by forward operator \( \mathbf{H}(\cdot) \)]. The variable \( \mathbf{H} \) is the linearized forward observation operator, \( \mathbf{R} \) is the observation error covariance matrix, \( k \) is the ensemble size, and \( \mathbf{X}_{t|0}^f \) and \( \mathbf{X}_{t|0}^f \) are \( m \times k \) analysis and forecast perturbation matrices, respectively (\( m \) = model state variables, \( k \) = number of ensemble members). Despite the use of the tangent linear model approximation to obtain (2), each column in \( \mathbf{X}_{t|0}^f \) can be calculated using the full nonlinear model \( \mathbf{M}(\cdot) \), such that the \( i \)th column is \( \mathbf{M} [\mathbf{x}_{t|0}^{\text{true}}] - \mathbf{M} [\mathbf{x}_{t|0}^{\text{true}}] \). The expression in (2) is appealing as it can be applied using available assimilation products of any deterministic EnKF method.

As with any method involving the use of ensemble to estimate covariances, covariance localization is needed to suppress the effects of sampling error from too small ensembles. The matrix product \( \mathbf{Y}_t^f \mathbf{x}_{t|0}^f = \mathbf{H} \mathbf{X}_t^f \mathbf{x}_{t|0}^f \) is the ensemble estimate of model error covariance between the analysis in observation space and forecast valid at time \( t \). Localization of (2) is applied to this \( p \times m \) matrix (\( p \) = number of observations). Denoting localization matrix \( \mathbf{P}_t \), the observation impact estimate modulated by the localization function becomes

\[
J = \frac{1}{K-1} \delta \mathbf{y}_0^T \mathbf{R}^{-1} [\mathbf{P}_t]^T (\mathbf{Y}_t^f \mathbf{x}_{t|0}^f) (\mathbf{e}_{t|0} + \mathbf{e}_{t|n}),
\]

(3)

where the symbol “\( \ast \)” refers to the Schur product, an element-by-element multiplication of two matrices of
the same size. The localization matrix \( p_l \) must be an \( m \times p \) matrix, meaning that every grid point–observation pair can have a unique localization weight. Since localization in (3) is applied to \( Y_{0l}^i X_{i0}^j \), in addition to spatial and cross-variable components, a level of complexity is added in the time-forecast component. In (3), there is no requirement that \( p_l \) has to be the same localization as that used during EnKF assimilation \( p_A \). The choice of \( p_l \) should attempt to take the time-forecast component into account, in addition to spatial and cross-variable components.

b. The RCF method of computing localization for observation impact estimate

The GF method of Anderson (2007) operates using groups of ensembles to calculate regression sampling errors in the ensembles. Assume that \( g \) groups of \( k \) ensembles are available in an assimilation system. When computing the linear regression between the state variables and observations, there are \( g \) samples of the regression coefficient \( \beta \). A weighting factor \( \alpha \) is defined to minimize the expected RMS differences between all possible combinations of sample \( \beta \) pairs. So, \( \alpha \) is chosen to minimize

\[
\sqrt{\frac{1}{g} \sum_{i=1}^{g} \sum_{j=1}^{g} (\alpha \beta_i - \beta_j)^2}. \tag{4}
\]

A simple derivation [see Anderson (2007)] leads to the following expression for \( \alpha_{\text{min}} \):

\[
\alpha_{\text{min}} = \frac{\left( \frac{1}{m} \sum_{i=1}^{m} \beta_i \right)^2}{\frac{1}{m} \sum_{i=1}^{m} \beta_i^2} - 1 = \frac{g - Q^2}{(g - 1)Q^2 + g}. \tag{5}
\]

where \( Q \) is the ratio of the sample standard deviation to the absolute value of the sample mean of the group \( \beta_s \). The optimal weighting factor \( \alpha_{\text{min}} \) is also known as the RCF. A unique \( \alpha_{\text{min}} \) can be calculated for each observation–state pair. The set of RCFs for a given observation and all state variables is called a observation-state pair. The set of RCFs for a given observation impact estimate needed to compute the group \( \beta_s \) according to (6), so there is an inherent limit to the forecast length at which it can be successfully applied.

3. Experiment design

a. The assimilation and forecast system

To evaluate and explore methods to improve the ensemble-based observation impact, experiments with a simplified primitive equation model and simulated observations were done. The dry, global, two-layer primitive equation spectral model of Zou et al. (1993) was chosen, which has been used in several studies of perfect- and imperfect-model ensemble-based data assimilation experiments (e.g., Wang et al. 2007, 2009; Holland and Wang 2013). It is useful because of its low computational demands allowing for many experiments to be conducted. The model variables include two vertical layers of vorticity, divergence, and layer thickness coefficients. The layer thicknesses, \( \Delta \pi_1 \) and \( \Delta \pi_2 \), are described in terms of \( \pi \) (i.e., the Exner function). The model includes simple parameterization schemes for radiative heating and surface drag, with zonal wavenumber-2 terrain. A fourth-order Runge–Kutta scheme is used for forward integration.

The model was run using the same parameters as in Holland and Wang (2013). To isolate the impact of sampling error, the experiments were conducted in a perfect-model context. A model run with daily output over thousands of days of integration at T31 resolution served as the truth. An initial ensemble was generated by a random draw of the truth states. The assimilation-forecast cycles were run at the T31 resolution for 1000 cycles at 1 day \((n = 24 \text{ h})\) intervals because of the long error doubling time of 3.78 days (Hamill and Whitaker 2005). The first 100 cycles were discarded from the RCF and impact calculations to allow the system to stabilize. Observations in interface height (i.e., the height between layer 1 and layer 2) were generated from the truth by adding errors drawn from a distribution with zero mean and fixed standard deviation of 250 m, as in Wang et al. (2007). There are 362 equally spaced observations total. These observations were assimilated into the LETKF following the settings of Holland and Wang (2013) with multiplicative and additive inflation. The GC localization was applied during assimilation with a

\[
\beta_{ij} = \frac{(HX_{ij}^H X_{ij}^H)_{ij}}{|HX_{ij}^H (HX_{ij}^H)^{1/2}|_{ij}}. \tag{6}
\]
cutoff radius of 8000 km, which was optimally tuned but,
more importantly, provided a stable forecast-assimilation
system for the full 1000 cycles.

b. Settings for impact estimate experiments using RCF

Figure 1 shows a flowchart of RCF computations. An
initial set of ensembles is randomly split into groups of
ensembles each prior to assimilation. Cycled LETKF
analyses are done for each group of ensembles separately. Following analysis, an ensemble forecast is run to
some valid time \( t \). For each group, a regression co-
efficient \( \beta \) is then calculated according to (6), from
which RCF is calculated according to (5) and saved for
each assimilation cycle. Each grid point–observation
pair has a unique RCF. Since the RCF from each cycle
is noisy because of a small number of groups used in the
calculation, particularly for locations far away from an
observation, the next step is to take the mean of the
RCFs across all cycles to dampen out those effects. This
results in a “lookup table” of RCF functions for every
observation. There are up to 256 ensemble members
total, which are grouped randomly prior to assimilation.
Initial experiments tested the sensitivity of resulting
RCF functions with varying number of groups (2, 4, 8,
and 16) using 16-member groups, and varying the
number of ensembles per group (8, 16, 32, and 64) using
4 groups. For the subsequent impact experiments, RCF
functions were calculated using \( g = 4 \) groups of \( k = 16 \)
ensemble members each (64 ensemble members total)
and applied as localization functions for ensemble im-
 pact estimates using a 16-member ensemble.

Impact experiments were conducted varying the
forecast valid time \( t \) from 0 to 4 days. At \( t = 0 \), no
forecast is run so the method gives RCF functions for the
impact of the observations on the analysis. RCF func-
tions were calculated first for model interface height,
\( z_{\text{int}} \), and then in terms of the state variables:
\( \Delta \pi_1, \Delta \pi_2, u_1, u_2, v_1, \) and \( v_2 \). RCF functions for the latter are considered
cross-variable impacts, which will serve to examine the
effects of using the new localization method to estimate
the impact of observations on unobserved variables.

Once the RCF functions were calculated, they were
applied directly to the impact metric in (3) using one of
the 16 ensemble group assimilations and compared with
ensemble estimates with no localization and using a
static GC localization. The results were validated
against the actual impact, which is the impact calculated
directly from (1) using the truth. Additionally, sets of

---

**FIG. 1. Flowchart of RCF method.** After an LETKF ensemble analysis, an ensemble forecast is run to some time \( t \). Then the analysis and forecast ensembles are randomly split into four groups, and for each group \( \beta \) is calculated according to (6). The RCF is then computed according to (4) for all observation–state pairs. This RCF envelope, unique for each analysis cycle, is fed into a running average over all the analysis cycles. The mean RCF envelope then serves as the GF localization function for impact estimate experiments. For the experiments in this study, the mean RCF function is calculated over a total of 900 cycles.
single-observation experiments were conducted to better understand the results qualitatively, to aid in viewing specifically the time- and cross-variable components of the localization and the impact estimate.

4. Results

a. RCF localization

Prior to application of the dynamic RCF functions to the ensemble impact metric, we first examine the structure of the localization functions outputted from the procedure described in Fig. 1. At the analysis time (Fig. 2a), RCF appears to have a Gaussian-like shape to it, though there are some differences such as narrow peaks and heavier tails. The effect of averaging over 900 cycles has smoothed the functions, though some noisiness still remains particularly in the tails of the distributions. Increasing the number of ensemble members per group increases the width of the RCF function, mostly greater than 30° in longitude away from the observation. The increased RCF width is because larger ensembles are less prone to noise from spurious correlations until greater distances from the observation. This result is consistent with results of Anderson (2007, his Fig. 4), as well as studies examining changes in optimal GC length scale with changing ensemble size (e.g., Hamill et al. 2001).

RCF functions for a 2-day forecast are shown in Fig. 2b. The time-forecast dependency causes the main signal to dampen and shift downstream from the observation location. The diminished magnitude suggests less confidence in ensemble covariances at longer forecast lead times. With increasing ensemble member size per group, there is an increase in the strength of the RCF function across all longitudes. Each ensemble size is able to capture the same time-dependent shift away from the observation and generally the same shape. The ensemble size of 16 chosen for various impact experiments in the following sections has a maximum signal of about 0.55 for the 2-day forecast, in contrast to the maximum of 1.0 for the analysis RCF function in Fig. 2a.

Figure 3 examines the sensitivity of changing the number of groups used in the RCF computation. The differences in the number of groups results in small differences in the RCF functions for both the analysis and 2-day forecast. This insensitivity can be attributed to the process of taking a long-term average of RCF functions over all LETKF cycles. Interestingly, just two groups would be sufficient in capturing the general shape of the RCF function. The rest of the results consider RCF functions computed from four groups of 16 ensemble members per group.

Each observation within the domain has a unique RCF function, and because the method reveals dynamical features of the model not all observations have a Gaussian-like spatial correlation. The seven observations in Fig. 4 each show zonally stretched RCF distributions, consistent with the predominantly zonal flow of
the model. Midlatitude observations (observations 1, 2, 3, 6, and 7) tend to have a more Gaussian-like appearance, though some have a triple-peaked structure (observations 1 and 6) associated with the strongest westerly flow. The distance between the peaks of about 40° in longitude is likely a representation of underlying Rossby waves. For example, for an observation placed at a trough these additional peaks represent the adjacent ridges associated with a trough. As observations get closer to the tropics (e.g., observations 4 and 5), RCF begins to take a different, sometimes complex shape, including stretching eastward along the equator upstream of the main flow.

RCF functions from Fig. 4 are valid for the analysis time of model interface height. As illustrated in Fig. 5, RCF functions also reveal the time-forecast and cross-variable dynamics of the model. With increasing time from analysis to 3-day forecast, the RCF function in interface height (Figs. 5a,d,g) shifts downstream of the observation, expands in area, and reduces in magnitude. The shift in maximum amplitude is approximately 10° per day between analysis and 2-day forecast, which is roughly 1000 km day⁻¹. This is consistent with results of Torn and Hakim (2008) who found a 1000-km distance in their composite sensitivity 24-h patterns over Washington State. In terms of layer two zonal wind $u_2$ (Fig. 5b) and meridional wind $v_2$ (Fig. 5c), the RCF functions exhibit dual peak dynamical structures. These RCF functions together mimic the shape of geostrophic adjustment correlations (e.g., Schlatter 1975) for both analysis and 1-day forecast. As forecast time increases to 2 days and longer, the RCF functions for cross variables smooth out and lose definition in dynamical linkage, though they still show the time-forecast dependency. It is possible that the small 16-member ensemble is unable to resolve cross-variable correlations beyond 2 days because of the weaker correlations, or other dynamical processes beyond advection occur past the 2-day lead time forecast.

b. Single-observation impact experiment

A single-observation experiment was first conducted to examine the qualitative results of impact estimates with differing localizations, compared to the actual forecast error reduction. For the experiment, a 1-day forecast of one of the all-observation LETKF analyses was chosen as the background for a single-observation analysis. This single interface height observation has an observation innovation of 1500 m. Deterministic forecasts initialized from the ensemble mean analysis and background were run to calculate actual impact according to (1). Ensemble forecasts of the single-observation analysis were run to calculate the ensemble-estimated impact according to (3), with differing choices of localization. Two localizations were tested: the same optimal GC localization (8000 km) as was used during assimilation, and the dynamic RCF functions shown in Fig. 5.

Results for one single-observation experiment are shown in Fig. 6 for analysis, 2-day, and 4-day impact of model interface height. Note again that negative values (filled blue) imply positive impact. The actual impact at
analysis (Fig. 6a) follows flow-dependent structures of the background. Initially, three main centers are present at magnitudes above 500 m$^2$. The ensemble estimates (Figs. 6b,c) are qualitatively similar, as both localizations capture these three main centers well.

The actual impact on 2- and 4-day forecasts (Figs. 6e,i) shows many more impact centers due to forecast error growth. These impact areas propagate along the main westerly waves predominantly located within the tightly packed interface height contours. Actual impact centers span a much greater zonal distance and can be seen at distances exceeding 8000 km from the observation. The ensemble impact estimate using static GC localization (Figs. 6f,j) cannot capture these far away impact centers because of the limited length scale. Moreover, the magnitude of the estimated impacts of areas closest to the observation is much stronger than those shown in the actual impact, indicating the GC localization weight is too large there. Conversely, the largest area of actual positive impact in the analysis (60°N, 75°W) has now advected 30° and 75° eastward near the edge of the GC localization function in the 2- and 4-day forecasts, respectively, where the localization weight is nearly zero. As a result, the GC localized impact value is underestimated compared to actual impact. These effects combined results in a global RMSE that is about the same or greater than the RMS of the actual impact fields. The ensemble impact using dynamic RCF localization (Figs. 6g,k) shows improved estimates, and thus, is able to match the magnitudes of each center more closely with the actual error reduction. The localization functions spans a much greater distance than the GC function, with centers shifted downstream of the observation. Overall, the RMSE is much lower at less than half of global RMS of the actual impact.

This one case demonstrates both the qualitative nature of actual impact varying with forecast length, and how the RCF can outperform static GC localization and lead to improved estimates from the evolving RCF functions. In the next section, verification of the all-observation experiment is discussed to see if the RCF localization shows added overall skill in the case of homogeneous observation coverage.

Fig. 4. Examples of RCF functions for seven differing interface height observations (locations marked by white dots), calculated for analysis time ($t = 0$) of model interface height. The wind vectors are a 900-cycle average of ensemble mean layer-2 wind. For plotting purposes, each observation’s RCF function is displayed only for values $>0.3$. 

Unauthenticated | Downloaded 03/21/24 04:31 PM UTC
c. All-observation impact experiments

In all-observation impact experiments, each LETKF analysis ensemble was run for forecast lengths varied between 0 and 4 days. Actual error reduction was calculated using deterministic forecasts initialized from the analysis mean and background fields. Ensemble-estimated impact was calculated by summing at each gridpoint contribution of the impact from all observations, with varying localizations applied. Figure 7 shows the globally averaged skill score (SS) of the time-mean RMSE of ensemble impact estimations verified against actual forecast error reduction. The quantity SS is defined as

$$SS = 1 - \frac{\text{RMSE}}{\text{RMSE}_{\text{ref}}} = 1 - \left[ \sum_{k=1}^{N} \left( \frac{\Delta e^2_{\text{actual},k} - \Delta e^2_{\text{ens est},k}}{\sum_{k=1}^{N} (\Delta e^2_{\text{actual},k})} \right)^{1/2} \right],$$

where $N$ is the number of cycles in time considered (900 for this study). This SS is equivalent to that used in Kalnay et al. (2012) where the reference RMSE is equivalent to the time-mean RMS values of the actual forecast error reduction.

Figure 7 shows the time-mean SS averaged over the globe. With increasing forecast time, for observation impact estimation of $z_{\text{int}}$, differences in SS emerge between GC and RCF localizations, with RCF localization showing increasingly higher skill. While the GC experiment approaches the no-skill line (0.0) by day 4, the RCF experiment has a skill around 0.4, still higher than the skill of the GC experiment at day 2. At the analysis time, the SS of the fixed GC localization experiment is nearly the same as the RCF localization in impact estimates of model interface height. In terms of indirectly observed layer-2 meridional wind (Fig. 7b), at analysis time the RCF experiment shows slightly lower skill than the GC experiment, but with increasing forecast time the RCF experiment becomes increasingly skillful relative to the GC experiment. The slight degradation in skill at the analysis time can be attributed to the inconsistency between the localization used in data assimilation and the
This issue is discussed further in section 4d.

In addition to SS using time-mean RMSE, Fig. 8 considers SS where global-mean RMSE is calculated in (6), and \( N \) is now the number of grid points (4608). The result is a unique global SS for each cycle in the experiment. The bar chart in Fig. 8a shows the percentage of the 900 cycles that show positive global skill for interface height impact. At the analysis time and 1-day forecast, both RCF and GC localization have a high percentage of cycles showing positive global SS. But as in Fig. 7, at 2-day forecast and beyond the differences become larger. The RCF function still shows positive skill more than three-quarters of the time at the 4-day forecast length, which is still better than even the GC function at 2-day forecast impacts. Of course, positive skill may not necessarily mean much when the skill is close to zero, so Fig. 8b shows instead percentages of cycles with SS exceeding 0.5. With the increased threshold, at 2 days and beyond the RCF still outperforms the GC by more than 20%. By the 4-day forecast, with a stricter threshold the GC function is not any more favorable than a simple top-hat localization with the same length scale, whereas the RCF shows high skill for nearly a quarter of the cycles considered.

The previous results for GC localization operate under the assumption that the same localization applied during assimilation is applied toward the impact estimate. What if, instead, we were allowed to optimally tune the localization length of the GC function for each forecast length? The GC tuning of impact estimates is shown in Fig. 9 for the analysis, 1-day, and 4-day forecast lengths. At the analysis, it is verified that the same tuning (8000 km) used at assimilation also provides optimal ensemble impact estimates. With forecasts, potential improvement by tuning the GC width is relatively minor, especially compared to the improvement that the RCF localizations show (dots in Fig. 9). A similar set of experiments varying the GC radius was conducted in Sommer and Weissmann (2014), and they too concluded that simple adjustment of the GC radius did not lead to a large improvement in accuracy of the ensemble impact method.

Finally, we examine the effect of RCF localization varying by latitude, since RCF functions at the midlatitudes and equator are very different in shape (Fig. 4). Here the pattern or map correlation is considered, using the Pearson correlation calculated between the estimated ensemble and the actual impact at the same location over all 900 cycles. The result is a map of correlation values, and in Fig. 10 the zonal average of this correlation map is shown. For 1-day forecast impact, both fixed GC and RCF show correlations well above the no-localization case, as expected due to sampling error in the raw ensemble correlations. RCF shows slight improvement relative to GC at midlatitudes (30°–70°) and the same or slight degradation elsewhere. At the 3-day impact (Fig. 10b), the RCF shows improvement over the fixed GC localization for most latitudes with the largest improvement at the midlatitudes. On the other hand, GC shows low skill comparable to no localization used for the observation impact estimate.

**Fig. 6.** Contour maps of observation impact estimates from a single-observation (yellow dot) assimilation experiment. (left to right) The actual impact (i.e., actual forecast error reduction), the ensemble-estimated impact using a static GC function with 8000-km cutoff radius, the ensemble-estimated impact using the RCF localization functions as pictured in Fig. 5, and ensemble-estimated impact using automatically tuned elliptical GC functions (see section 4f), respectively. (a)–(d) Impacts at analysis \( t = 0 \), (e)–(h) impacts on a \( t = 2 \)-day forecast, and (i)–(l) impacts on a \( t = 4 \)-day forecast. Color-filled contours show impact values (m²), black contours are model interface height in 1000-m intervals, and magenta contour lines show the localization function applied to the impact estimate contoured in intervals of 0.2 starting at 0.2.
localization, an indication that while localization is needed, the GC localization is too restricted and centered in the wrong spot. A distinct shape is present at both 1- and 3-day lead times, showing high values at midlatitudes with a large dip within 30° of the equator.

The improvement of GC and RCF relative to no localization is much smaller in the tropics than the midlatitudes. In the next section we further investigate this issue.

d. Relationship between localizations for data assimilation and for observation impact estimate

The improvement in skill in the 1–4-day forecast range is due to the time-shift dependency. Two issues with the RCF localization applied to the impact estimate were noted. The first is why RCF did not show improved performance at the tropics, as seen in Fig. 10. The second, as seen in Fig. 7, is that the cross-variable dynamical linkage revealed within RCF did not offer any improvement on skill at analysis time. Both issues are related to the fact that localization used during assimilation was not consistent with the localization that the RCF suggested for the observation impact estimate.

First, it is important to note the uniqueness of the equatorially stretched RCF functions in the tropics (Fig. 11). The two-layer model used here, while able to well represent realistic midlatitude baroclinic instability, is not a realistic representation for the tropics due in part to the simplifying exclusion of moisture. Hendon and Hartmann (1985) analyzed the variability of a similar dry two-layer model and noted that the tropics are dominated by internal normal modes consistent with Matsuno (1966). Additionally, these waves are equatorially trapped, with minimal activity propagating from the tropics to the midlatitudes. The RCFs of an equatorial observation are also equatorially trapped and show east- and westward-propagating components (Fig. 11c). So tropical RCF functions represent the internal normal mode dynamics of the model used in this study.

The effect of inconsistent localization functions for the impact estimate compared to localization used during data assimilation can be shown within the derivation of Kalnay et al. (2012) together with a single-observation experiment for an equatorial observation using 8000-km GC localization during assimilation (Fig. 12). A basic EnKF mean update is of the form

\[ \mathbf{x} - \mathbf{x}^b = K \delta \mathbf{y} \]

Assuming that localization \( \mathbf{r} \) was applied to the EnKF in some way, then it is contained within \( K \) such that it and the mean update \( \mathbf{x} - \mathbf{x}^b \) both go to 0 as \( \mathbf{r} \) goes to 0 (i.e., no update of the background). This is evidenced by the actual impact field at analysis time shown in Fig. 12a, which covers the same extent as the 8000-km GC localization shown in Fig. 11a. In Kalnay et al. (2012), the derivation of observation impact was based on substitution of postanalysis expression for gain

\[ K = P^T \mathbf{H}^T \mathbf{R}^{-1} = (k - 1)^{-1} \mathbf{X}^T \mathbf{X}^T \mathbf{H}^T \mathbf{R}^{-1} \]

where \( \mathbf{M} \) is the tangent linear model. In order for the postanalysis formulation of gain \( K \) to be consistent with the EnKF update using some localization matrix \( \mathbf{r} \), then that localization function also has to be present in the postanalysis formation.
where $P_{loc}^a = (I - KH)(P_A \cdot P^a)$ is the “localized” analysis error covariance matrix. In other words, for the impact metric we need find a localization function $r_I$ so that

$$P_{loc}^a = r_I \cdot P^a,$$

where $P^a = (k - 1)^{-1} X a X a^T$. It is not straightforward to find a localization function $r_I$ to satisfy (9). However taken together, (8) and (9) indicate that we should expect the optimal localization choice for the impact metric $P_A$ to be related to the localization used during assimilation $P_A$. This is consistent with why the GC estimate (Fig. 12b) is qualitatively close to the actual impact (Fig. 12a) while the RCF estimate (Fig. 12c) is

---

**Fig. 8.** Bar charts showing percentage of cycles with skill scores of ensemble observation impact estimates greater than (a) 0.0 and (b) 0.5, for no localization (light blue), square or top-hat localization with 8000-km cutoff (orange), GC localization with 8000-km length scale (blue), and GF localization (red).

**Fig. 9.** Optimal tuning of the GC length scale for ensemble impact estimates at analysis (black), 1-day forecast (blue), and 4-day forecast (green), verified against actual error reduction and averaged over 48 cycles. Colored dots indicate RMSE of ensemble impacts using group filter localization, plotted vertically from its respective optimal RMSE using GC. Vertical dashed line indicates the GC length scale used for assimilation (8000 km).
not, since the GC function is the same as was used during assimilation whereas RCF is a completely different shape. For the case studied in Figs. 11–12, because little activity propagates from the tropics to midlatitudes, the 8000-km GC function used during assimilation may be inappropriate for this observation, incorrectly adjusting midlatitude locations. Instead, localization consistent with RCF should be used during assimilation.

The implicit relationship between $r_I$ and $r_A$ may also explain why the dynamical linking of cross variables in the RCF function did not show improved impact estimates at the analysis (Fig. 7), because it is simply a different shape than what was used at assimilation time. It also helps explain why the dynamic RCF method was successful at the midlatitudes, because the initial RCF function at the analysis time was similar to the 8000-km GC function used during assimilation, allowing for the added benefit of the time-evolving component for forecasts.

e. All-observation experiment using RCF localization during assimilation

The inherent relationship between $\rho_I$ and $\rho_A$ suggests that the best use of adaptive localization for the impact metric occurs when the same adaptive method is used during assimilation. To test this hypothesis, a new all-observation assimilation experiment is conducted where the analysis RCF functions (e.g., Figs. 4a–c) are used during the assimilation as localization, $\rho_A$. This experiment will be referred to as “RCF-assim.” Impact estimates are calculated with the RCF functions applied as localization and verified against the actual forecast error reduction, as in previous experiments. Pattern correlation and SS is

Fig. 10. Zonally averaged pattern–map correlation of ensemble-estimated impact of model interface height compared to actual impact for no localization (black), GC localization (blue), and dynamic RCF localization (red). Dashed red line shows correlations of estimated impact using RCF localization at analysis time during assimilation (RCF-assim, see section 4e) for (a) 1-day forecast impact and (b) 3-day forecast impact.
The SS in Fig. 7 shows that RCF-assim yields improved skill compared to the previous application of RCF to the GC assimilation, particularly for shorter (0–2 day) forecasts. This improvement converges to the level of skill from the previous RCF experiment at longer lead times for interface height. In terms of cross-variable $v_2$ impact, RCF-assim shows higher skill than the previous RCF experiment for 0–2-day forecasts, but actually becomes less skillful for 3–4 days. It is possibly an indication that RCF should also be re-calculated from the new RCF assimilation. Comparing pattern correlations of RCF-assim to the previous RCF experiment, there is clear improvement at all latitudes, including significant improvement in the tropics at both 1- and 3-day forecasts (Fig. 10). The results of RCF-assim confirm that the best use of adaptive localization methods for the ensemble impact metric occurs when the same adaptive method is used for localization during the assimilation.

**f. Using RCF to design automatically tuned elliptical GC localization functions**

Another potential use of adaptive methods such as the RCF is its derived shifting, magnitude change, and area coverage of the localization that could be used in the future to automatically tune a GC function for observation impact localization. For example, Fig. 13 shows four potential tuning parameters that may be calculated from RCF and how they evolve with increasing forecast time. Given predominantly zonal flow, each parameter was tuned for observations binned by latitude. One can see, for example, the differences in how far to advect a GC localization away from an observation due to differences in the zonal wind strength at each latitude (Fig. 13a) and the similarities at each latitude in reduction of maximum magnitude with increasing forecast time (Fig. 13b). It can also be seen that localization area should expand with increasing forecast time, but at different amounts depending on latitude of the observation (Fig. 13c). Complimentary to that, the localization function should have more of a two dimensional elliptical shape, since the zonal extent is greater than the meridional extent (Fig. 13d).

One question is how to apply these parameters with the preexisting definition of the GC function from Gaspari and Cohn (1999) to create an elliptically shaped localization. In the homogeneous or “circular” case, the GC function has the same cutoff distance in all directions; however, in the elliptical case, the cutoff distance changes based on the angle relative to the major and minor axes of the ellipse.

The steps to defining an elliptical GC localization for the impact metric are as follows:

1. Calculate parameters from the RCF functions as in Fig. 13. Parameters needed are the zonal span, meridional span, maximum magnitude, and offset distance from observation.
2. Use offset distance to define the center of the ellipse for a given observation and forecast time.
3. At each grid point, calculate the angle $\theta$ relative to the center of the ellipse using $\theta = \arctan(\Delta y/\Delta x)$, where $\Delta x$ and $\Delta y$ are the zonal and meridional component distances, respectively, between the grid point and the center of the ellipse.
4. The parametric form of an ellipse is given by $x = a \cos t$, $y = b \sin t$, where $a$ and $b$ are major and minor axis distances, respectively, and $t$ is the parametric angle. In this case, $a$ and $b$ are simply half of the zonal and meridional spans. Convert from polar angle $\theta$ to parametric angle $t$ using

$$t = \arctan(a/b) \tan \theta. \quad (10)$$

5. The cutoff distance for the GC functions as a function of angle is given by the distance formula:

$$c = \sqrt{a^2 \cos^2 t + b^2 \sin^2 t}. \quad (11)$$

6. Once the cutoff radius $c$ is determined, the GC function can be applied directly. To adjust for reducing magnitude with forecast time, multiply the output weights from the GC function by the maximum magnitude parameter. Figure 14 displays the resulting elliptical GC localization functions using the parameters from Fig. 13.

This new elliptical GC localization function was tested with the single-observation impact experiment. As shown in the rightmost column of Fig. 6, the estimates are similar in performance to the RCF in terms of overall structure and RMSE. More precise definitions and applications of these derived parameters may yield better results, but are beyond the scope of this paper. Here it is presented as a proof-of-concept to demonstrate how such parameters can be calculated from RCF and applied to tune the GC function.

5. Conclusions and discussion

The real-world application of the ensemble-based observational impact estimate needs to consider proper implementation of localization to reduce errors due to limited ensemble members. Such localization
FIG. 11. Localization functions for an equatorial observation (white dot): (a) GC localization function with 8000-km length scale (outermost ring is contoured at 0.01). RCF functions valid for (b) analysis ($t = 0$) and (c) 2-day forecast.
FIG. 12. Single-observation impact experiment for an assimilated observation located at the equator (yellow dot). (a) Actual impact, or actual error reduction, of observation at analysis time, (b) ensemble estimate of impact using GC localization (8000 km), and (c) ensemble estimate of impact using RCF localization. Color-filled contours show impact values (m²) and black contours are model interface height in 1000-m intervals.
needs to consider time-forecast dependency in addition to spatial and cross-variable dependencies. Kalnay et al. (2012) and Ota et al. (2013) have shown that initial attempts to “advect” the localization function downstream from an observation leads to more reliable and accurate impact estimates. Here we examined an alternative method, an adaptive localization that varies by location and state variable according to the underlying model dynamics. The method is based on confidence factors of grouped ensemble regression coefficients, first proposed for EnKF assimilation (Anderson 2007). In this study, the method was extended to work for the ensemble-based observation impact estimate. An envelope of regression confidence factors (RCFs) for each observation and all grid points was used directly as localization. The purpose of the study was to explore the potential effectiveness of this adaptive method on the observation impact estimate and to learn more about how to properly localize the ensemble-based impact metric.

Results of the technique within a simple two-layer isentropic model showed the ability of the RCF method to reveal underlying dynamics between variables and the time-forecast component. Applying RCF functions for impact estimates showed overall improvement when verified against the actual forecast error reduction, especially at longer forecast lead times compared to using fixed GC localization. Single-observation experiments displayed the evolving structures of the actual forecast error reduction with increasing forecast time and

![Fig. 13. Summary plot of potential localization tuning parameters for the impact estimation as a function of forecast time, derived from RCF functions in model interface height. (a) Offset distance from observation, or shift, defined as the distance between RCF maximum and the observation, (b) reduction in magnitude with increasing forecast time, (c) fractional surface area coverage, and (d) maximum span, or range, of RCF values in zonal direction (solid) and meridional direction (dashed). Values are averaged for all observations within 40°–55°N (black) and 55°–70°N (gray).](image-url)
showed that the dynamic localization was able to simulate the evolved error reduction much more closely than the static localization especially at longer forecast lead times. For an LETKF using all 362 observations, skill scores of the observation impact estimate applying RCF localization significantly improved upon those using fixed GC for increasing forecast lead times. For example, the skill of the estimates of the impact using RCF on a 4-day forecast beats the skill of the estimate using static GC at the 2-day forecast, essentially doubling the forecast length of accurate impact estimates. This improvement was attributed mainly to the time dependency of RCF shifting downstream, diminishing in magnitude, and expanding in areal coverage.

Further diagnostics also found that optimal use of localization from dynamical methods relies on the assumption that consistent localization was used during the assimilation as well. In other words, there is a fundamental relationship between the localization applied at assimilation time and the localization used for the impact estimate. Therefore, further improvement in skill of using the RCF in observation impact estimate is possible if a consistent adaptive localization was used at the assimilation time in place of the static GC function. This was confirmed in an additional experiment using RCF localization both during assimilation and for the impact estimates. Skill scores were higher and significant improvement was made in the estimates at the topics.

Another potential utility of adaptive methods such as the RCF is to automatically tune an elliptical GC function for observation impact localization. As a proof-of-concept, the application of RCF-derived parameters to tune the GC function shows promising results that were comparable to the RCF experiment for a single observation test.

For real-time application, a drawback of the RCF technique is computational cost of additional groups of ensemble members. To partially alleviate this issue, one could use increased ensembles only during a “training period” for a given weather regime to create RCF lookup tables offline, similar to the suggestion of Anderson (2007). Additionally, Fig. 3 suggests that just two groups can yield RCF functions consistent with RCF using more groups, due to the averaging process over many cycles to reduce noise. Figure 3 also suggests another idea, which is to split available ensembles into subgroups without increasing the total number of ensemble members and then artificially “inflate” the resulting RCF functions, because with increasing groups the RCF functions retain similar shape and shift but differ only in magnitude. Alternatively, dividing existing ensembles into subgroups the RCF method can still be utilized to derive parameters for automatically tuning localizations.

More challenges may be found for application to complex models, especially down to convective scales, which have many different kinds of day-to-day extreme weather patterns that we would not want the averaging process to smooth out. One idea to address this issue is an “online smoothing” process, or averaging over composites of weather patterns for a given day. Another idea is to

---

**Fig. 14.** Automatically tuned elliptical GC localization functions for (a) analysis, (b) 1-day, (c) 2-day, and (d) 3-day forecasts. The parameters used for the tuning are shown in Fig. 13.
use a different adaptive approach such as the ensemble correlations raised to a power (ECO-RAP) technique of Bishop and Hodyss (2009a,b), which can be applied to day-to-day ensemble correlations to automatically localize sampling error. The current study extends the RCF method for an observation impact study and explored its effectiveness compared to fixed GC. Comparing the RCF method with other adaptive methods at convective scales will be a topic in future work.

Acknowledgments. The authors thank the Oklahoma Supercomputing Center for Education and Research for providing the computational resources that were used for simulations throughout this study. The research was primarily supported through the Department of Commerce and NOAA Grant NA110AR4320072 and NASA NIP Grant NNX10AQ78G. The authors would also like to thank Brian Holland for providing his EnKF system for our use and Govindan Kutty for providing the subroutine to calculate the observation impact. Comments from three anonymous reviewers and the editor were very helpful in improving the manuscript and are greatly appreciated.

REFERENCES


