

Ensemble-Based Sensitivity Analysis

RYAN D. TORN* AND GREGORY J. HAKIM

University of Washington, Seattle, Washington

(Manuscript received 2 January 2007, in final form 18 May 2007)

ABSTRACT

The sensitivity of forecasts to observations is evaluated using an ensemble approach with data drawn from a pseudo-operational ensemble Kalman filter. For Gaussian statistics and a forecast metric defined as a scalar function of the forecast variables, the effect of observations on the forecast metric is quantified by changes in the metric mean and variance. For a single observation, expressions for these changes involve a product of scalar quantities, which can be rapidly evaluated for large numbers of observations. This technique is applied to determining climatological forecast sensitivity and predicting the impact of observations on sea level pressure and precipitation forecast metrics. The climatological 24-h forecast sensitivity of the average pressure over western Washington State shows a region of maximum sensitivity to the west of the region, which tilts gently westward with height. The accuracy of ensemble sensitivity predictions is tested by withholding a single buoy pressure observation from this region and comparing this perturbed forecast with the control case where the buoy is assimilated. For 30 cases, there is excellent agreement between these forecast differences and the ensemble predictions, as measured by the forecast metric. This agreement decreases for increasing numbers of observations. Nevertheless, by using statistical confidence tests to address sampling error, the impact of thousands of observations on forecast-metric variance is shown to be well estimated by a subset of the $O(100)$ most significant observations.

1. Introduction

Forecast sensitivity analysis provides an objective means of evaluating how changes to an initial condition affect a forecast. Typically the analysis applies to linear changes as measured by a scalar metric of the forecast variables. In a predictability context, sensitivity analysis provides a basis for understanding the dynamics of forecast errors, and also the locations for which additional observations may be gathered to reduce errors, as measured by the forecast metric.

Previous studies on initial condition sensitivity have involved using the adjoint of a linearized forecast model. Adjoint sensitivity and singular vector analyses for extratropical cyclones emphasize structures in the lower troposphere, which have large vertical tilts and

are not always obviously related to the major synoptic features (e.g., Errico and Vukicevic 1992; Langland et al. 1995; Rabier et al. 1996; Zou et al. 1998; Hoskins et al. 2000). Difficulties with these techniques include coding the adjoint of a tangent linear model, which is especially challenging for on-off processes within boundary layer and microphysical parameterizations, and assumed linearity.

Here we consider an ensemble approach to sensitivity analysis, where sample statistics are used to estimate relationships between forecast metrics and initial conditions. Such an approach was proposed by Anderson (2001) to construct an adaptive observing system by using ensemble data to provide estimates of the joint distribution of the model state at earlier times with the state at the present time. Hamill and Snyder (2002) applied a similar technique to estimate the impact of observations on analysis variance, but did not consider forecast impact. Ensemble sensitivity was formally applied to an extratropical cyclone by Hakim and Torn (2008), while Ancell and Hakim (2007) compared ensemble sensitivity with adjoint sensitivity analysis for a wintertime flow pattern. For the case examined by Ancell and Hakim (2007), they found that ensemble sensitivity provides accurate estimates of the impact of ini-

* Current affiliation: National Center for Atmospheric Research, Boulder, Colorado.

Corresponding author address: Ryan D. Torn, Department of Atmospheric Sciences, University of Washington, Box 351640, Seattle, WA 98195-1640.
E-mail: torn@atmos.washington.edu

tial condition changes to a forecast metric. Furthermore, their results indicate that the technique is useful for identifying a target region for additional observations because, unlike adjoint sensitivity, the analysis-error statistics are included in the ensemble calculation.

Ensemble sensitivity analysis is applied here to a six-month sample of ensemble analyses and forecasts generated by a pseudo-operational ensemble Kalman filter. One goal is to illustrate how ensemble sensitivity can easily be used to determine climatological sensitivity for a given forecast metric. This can be done “offline” without further model integrations, provided ensemble analyses and forecasts are available. This may be useful for observing-network design, with a goal of constructing an optimal network to minimize errors in a particular forecast metric (Khare and Anderson 2006). A second goal is to test the accuracy of the ensemble sensitivity predictions of the impact of observations on a forecast metric. Having the ability to estimate observation impact on a metric may prove useful for thinning a large set of observations to a smaller one during data assimilation, and also for providing rapid real-time updates to a forecast metric without having to wait for completion of the full assimilation and forecast process. Whitaker et al. (2008) propose an ensemble-based observation thinning algorithm based on analysis-error variance reduction, which we extend here to forecast metrics and to statistically significant changes in the metric mean value.

The outline of the paper is as follows. An overview of the pseudo-operational ensemble Kalman filter and ensemble sensitivity analysis are given in section 2. Sensitivity results for a 6-month period, and the most sensitive cases in the period, are discussed in sections 3 and 4, respectively. The accuracy of the ensemble sensitivity predictions for ensemble forecasts are tested through observation denial experiments in section 5, and for the full data assimilation cycle over a large sample of cases in section 6. In section 7, we apply this method to link observations to forecast verification. A concluding summary is given in section 8.

2. Experiment setup

Ensemble-based initial condition sensitivity for the west coast of North America is evaluated using data drawn from the University of Washington ensemble Kalman filter (UW EnKF) system (Torn and Hakim 2007, manuscript submitted to *Mon. Wea. Rev.*) during 1 January–30 June 2005. This system assimilates observations every 6 h (0000, 0600, 1200, and 1800 UTC) using a square root version of the EnKF (Whitaker and Hamill 2002) for a 90-member ensemble. Observations assimilated include Automated Surface Observing Sys-

TABLE 1. Observation types and average number of observations assimilated during each forecast cycle by the UW EnKF system during January–July 2005. There are 30 rawinsonde launches at 0000 and 1200 UTC.

Observation	Type	Analysis time (UTC)			
		0000	0600	1200	1800
Surface	Alt, u , v	430	420	420	440
Rawinsonde	u , v , T , RH	1000	0	1000	0
ACARS	u , v , T	1650	1390	740	1860
Cloud winds	u , v	2030	1740	1670	1510
Total		5110	3550	3830	3810

tem (ASOS) stations, ships, buoys, rawinsondes, Aircraft Communications Addressing and Reporting System (ACARS), and cloud-motion vectors (Velden et al. 2005); Table 1 summarizes the type and average number of observations assimilated during each analysis time. To minimize spurious long-distance covariances, the influence of observations is localized using the Gaspari and Cohn (1999) fifth-order piecewise rational function given by their Eq. (4.10), which in our implementation reduces to zero 2000 km from the observation location; vertical covariance localization is not applied. Moreover, the tendency for small ensembles to underestimate covariance magnitude is treated by inflating the deviations from the ensemble mean by replacing the posterior perturbations with a linear combination of the prior and posterior perturbations where the prior (posterior) is weighted by 0.80 (0.20) (Snyder and Zhang 2003).

We use the Advanced Research version (ARW) of the Weather Research and Forecasting (WRF) model (Skamarock et al. 2005) on a numerical grid with 45-km horizontal grid spacing and 33 vertical levels over a domain that includes the Gulf of Alaska and western North America. The model uses the WRF three-class microphysics scheme (Hong et al. 2004), Kain–Fritsch cumulus parameterization (Kain and Fritsch 1990), Mellor–Yamada–Janjic boundary layer scheme (Janjic 2002), and the Noah land surface model (Ek et al. 2003). An ensemble of lateral boundary conditions are generated using the fixed covariance perturbation (FCP) technique of Torn et al. (2006) with a scaling factor and autocorrelation coefficient of 1.6 and 0.4, respectively. Ensemble-mean forecasts on the lateral boundaries are obtained from the National Centers for Environmental Prediction (NCEP) Global Forecasting System (GFS) forecast from the previous forecast cycle valid at the appropriate time. At 0000 and 1200 UTC, 24-h ensemble forecasts are generated by advancing all 90 ensemble members with FCP ensemble boundary conditions.

The sensitivity of a forecast metric to the initial con-

ditions is computed using the ensemble sensitivity technique first outlined in Hakim and Torn (2008) and further explored by Ancell and Hakim (2007). For an ensemble of size M , the sensitivity of the ensemble-mean value of the forecast metric J to an analysis state variable x is determined by

$$\frac{\partial J}{\partial x} = \frac{\text{cov}(\mathbf{J}, \mathbf{x})}{\text{var}(\mathbf{x})}. \quad (1)$$

Here \mathbf{x} and \mathbf{J} are $1 \times M$ ensemble estimates of the state variable and forecast metric, respectively, with the ensemble mean removed; cov denotes the covariance between the two arguments; and var is the variance. A derivation of (1) and its relationship to adjoint sensitivity analysis is found in Ancell and Hakim (2007). The above equation represents linear regression where the independent variable is an analysis grid point and the dependent variable is the forecast metric. In the following sections, initial condition sensitivities are determined for the 24-h forecast of average sea level pressure (SLP) and average precipitation within a box that includes the western half of Washington State (“western Washington”). This region is often impacted by short-term forecast failures resulting from initial condition errors (McMurdie and Mass 2004) and is of interest to the authors by proximity. We note that, in general, the ensemble sensitivity technique is not limited to the metrics and forecast lead hour we describe here.

3. Climatological results

Data drawn from the UW EnKF system are used to determine the climatological sensitivity of pressure and precipitation averaged in a box over western Washington. Climatological sensitivity is defined here as the percentage of analysis cycles for which the ensemble sensitivity of the forecast metric with respect to an analysis grid point is different from zero at a certain level of confidence. Specifically, a state variable can produce a statistically significant change in the forecast metric if

$$\left| \frac{\partial J}{\partial x} \right| > \delta_s, \quad (2)$$

where δ_s is the confidence interval on the linear regression coefficient (e.g., Wilks 2005, section 6.2.5). For example, taking δ_s to be the 95% confidence interval (the value used for the climatological results given below), if (2) is satisfied, we may reject the null hypothesis that changes to x do not change the forecast metric with 95% confidence. Regions with a high percentage of

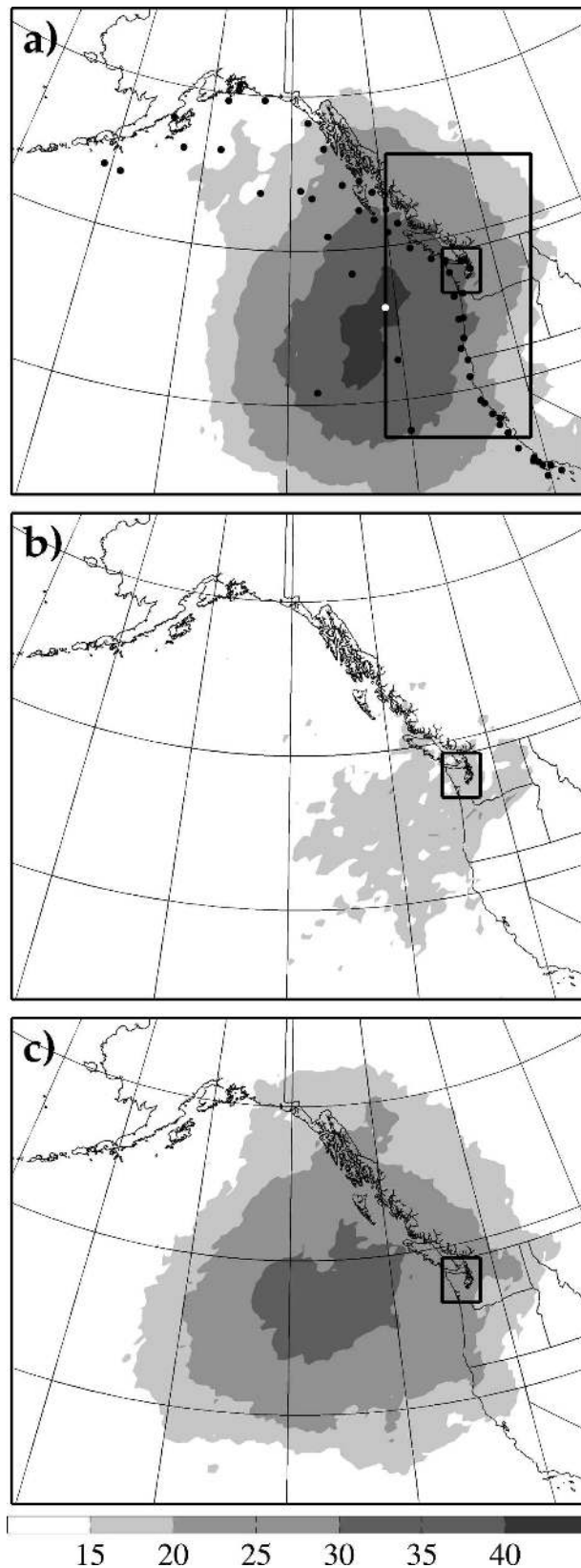
sensitive forecast cycles may be regarded as potential locations for siting new observations.

Figure 1a shows results for sensitivity of the 24-h forecast of average SLP in the box over western Washington to SLP analyses. The region with the largest percentage of sensitive forecast cycles is located over the Pacific Ocean, with a maximum value of 44% of cycles at (45°N, 132°W). This pattern qualitatively reflects the progression of weather systems from west to east at a mean translation speed of 9 m s^{-1} , which is roughly consistent with the average speed of individual eddies in the Northern Hemisphere midlatitude flow (e.g., Hakim 2003). For shorter lead times, the region of maximum sensitivity is located closer to western Washington (not shown).

Regions of consistent sensitivity in Fig. 1a predict where additional SLP observations would most frequently change the SLP forecast metric. The location of maximum sensitivity is close to buoy 46005 (white dot). In fact, this buoy failed on 26 December 2004 and therefore observations from this location were not available during the time period of this experiment. This suggests that the absence of observations from buoy 46005 may have adversely affected forecasts over western Washington during these six months. In light of this possibility, we will revisit this problem in section 5, where the change in the 24-h SLP forecast associated with a missing buoy is quantified by withholding a nearby buoy from the analysis and comparing the predicted and actual changes in the forecast metric.

The forecast SLP metric is also frequently sensitive to analyses of 850-hPa temperature and 500-hPa geopotential height. For 850-hPa temperature, there are two main sensitive regions, one to the southwest of Washington State near 43°N, 130°W, and another to the east of the metric box (Fig. 1b). For 500-hPa geopotential height (Fig. 1c), the forecast metric is sensitive more than 20% of the time to the region bounded by 40°–60°N and 120°–160°W. We note that this region is located a few hundred kilometers upstream of the region of maximum SLP sensitivity, reflecting a moderate upstream tilt typical of baroclinic waves in the westerlies.

The second metric we consider is precipitation averaged in the box over western Washington. Since sensitivity can only be determined when the forecast metric has nonzero variance, the percentage of sensitive cycles is computed with respect to the total number of cycles where the area-averaged precipitation in the box exceeds 1 mm in the ensemble-mean forecast for hours 18–24; 58% of all forecasts exceed this threshold. Sensitivity of this metric to SLP shows a maximum of 40% over the Pacific Ocean in a meridionally elongated re-



gion near 132°W (Fig. 2a). Whereas the northern half of this region is relatively well observed by the near-shore buoy network, the southern half is observed by fewer buoys, especially since buoy 46005 was not functioning.

For 850-hPa temperature analyses, the precipitation metric is sensitive more than 20% of the time to the southwest of western Washington, with maximum sensitivity near 132°W (Fig. 2b). Precipitation forecast sensitivity to 500-hPa geopotential height is similar to other fields, with maximum values along 132°W (Fig. 2c). The main sensitive region for 500-hPa height is slightly upstream of the region of maximum SLP sensitivity; thus the sensitivity fields for this metric are also tilted westward with height.

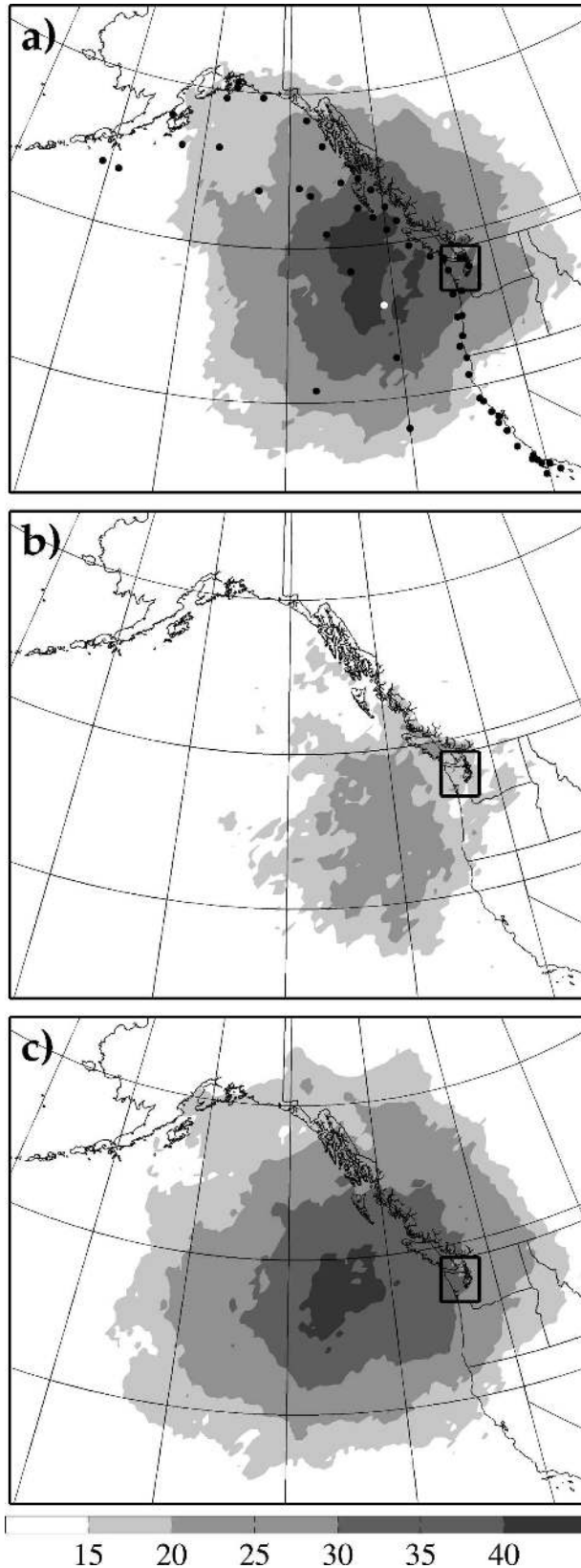
In summary, 24-h forecasts of SLP and precipitation averaged over western Washington are consistently sensitive to analysis errors to the west of the forecast box in an area characterized by few in situ observations. The relatively close proximity of the main sensitivity regions for both SLP and precipitation metrics with respect to a variety of state variables suggests that western Washington short-term forecasts may benefit from the introduction of new regular observations over a relatively small region. An important caveat to these conclusions is that these experiments do not include satellite radiance observations. Adding these observations could change the sensitive regions and “whiten” analysis errors, thus making it more difficult to find sensitive regions with small ensembles (e.g., Hamill et al. 2002). Nevertheless, our main point is that ensemble sensitivity analysis provides a simple, well-defined method for observing network design for which the results are tied to the forecast model, but the method is not (cf. adjoint sensitivity).

4. Most sensitive cases

Whereas the previous section showed how often a forecast metric is sensitive to changes in a state variable, composite averages are used here to determine the spatial pattern of sensitivity that occurs for forecasts having the largest sensitivity. These distributions show locations where small errors can, on average, lead to

←

FIG. 1. Percentage of forecast cycles with gridpoint sensitivity statistically significant at the 95% confidence level for western WA 24-h SLP forecasts sensitivity for (a) SLP, (b) 850-hPa temperature, and (c) 500-hPa height. Forecasts are initialized at 0000 and 1200 UTC from 1 Jan to 30 Jun. The forecast SLP is averaged over the region indicated by the smaller box in (a). Dots in (a) indicate the position of fixed buoys, and the larger box denotes the NAC metric region used in Figs. 8 and 9.



large changes in the metric. The most sensitive western Washington SLP and precipitation forecasts are determined by computing a domain-average forecast sensitivity (DAS):

$$DAS = \frac{1}{N_h} \sum_{i=1}^{N_h} \left| \frac{\partial J}{\partial x_i} \right|, \quad (3)$$

where x_i is the SLP at a grid point i , and N_h is the number of horizontal grid points. This norm is used to determine the most sensitive cycles because SLP is a column-integrated quantity and is expected to have the largest sensitivity values since the forecast metric is area-averaged SLP. Composite patterns of forecast sensitivities for the 30 cycles with the largest DAS values are calculated by

$$\overline{\Delta J} = \frac{1}{N_t} \sum_{t=1}^{N_t} \left(\frac{\partial J}{\partial x_i} \right) \sigma_{x_t}, \quad (4)$$

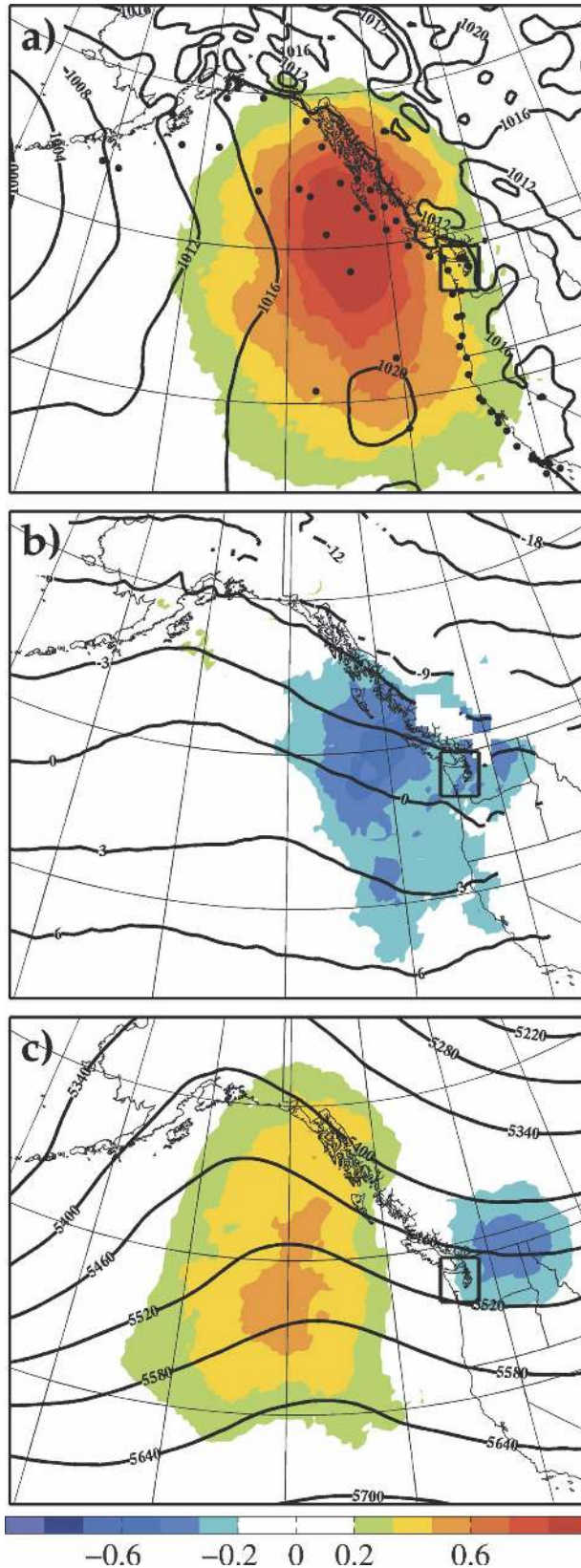
where σ_{x_t} is the standard deviation of x_i at time t , and N_t is the number of cycles used in the composite, which is 30.¹ Multiplying $\partial J/\partial x$ by σ_{x_t} gives the change in J brought about by a one standard deviation change in x , and thus a quantitative comparison of how perturbations in various analysis fields change J . Regions of high composite sensitivity indicate where additional observations could have the largest impact during the most sensitive forecasts.

Figure 3a shows the composite sensitivity of the 24-h forecast of average SLP in the box over western Washington to SLP analyses. Increasing (decreasing) x at one grid point by one standard deviation within regions of largest sensitivity values implies a 0.9 hPa increase (decrease) in the forecast metric. The region of largest sensitivity is in an area characterized by few buoys at (47°N, 135°W) and is located north of the region having the largest percentage of sensitive cycles (Fig. 1a).

For 850-hPa temperature, the average sensitivity is less coherent than for SLP, although sensitivity appears both east and west of the forecast-metric box (Fig. 3b).

¹ We chose 30 cycles so that a few forecast cycles with large sensitivity values do not bias the horizontal distribution and magnitude of the average sensitivity.

FIG. 2. As in Fig. 1, but for the 24-h forecast of precipitation averaged over the western WA region. Here the percentage of forecast cycles is computed with respect to the number of cycles where the precipitation in the box exceeds 1 mm for forecast hours 18–24.



Furthermore, this region of high sensitivity is to the north of the region of consistent sensitivity in Fig. 1b; thus while the average SLP forecast is more often sensitive to the 850-hPa temperature southwest of the metric box, the largest magnitude sensitivities are to the northwest of Washington. Increasing (decreasing) the temperature in the regions with the largest values by one standard deviation only leads to a 0.5-hPa decrease (increase) in the SLP in the box 24 h later. For 500-hPa height, the SLP forecast is sensitive to a meridionally elongated region near 140°W ; a one standard deviation change in x within the regions of largest sensitivity is associated with a 0.6-hPa change in the forecast metric (Fig. 3c). This region is located a few hundred kilometers upstream of the region of maximum SLP sensitivity, indicating that the moderate baroclinic tilt of the sensitivity field is a common property among these results.

Average sensitivity for the 30 most sensitive precipitation forecasts is determined in a manner similar to the 30 most sensitive SLP forecasts described above. Although one should expect precipitation to have a non-Gaussian distribution since it is bounded from below by zero, Gaussian statistics are nevertheless assumed in these calculations. Results for the precipitation metric show a more pronounced composite-average low pressure system in the Gulf of Alaska and maximum sensitivity to SLP near (44°N , 133°W), just south of the results for the SLP metric (Fig. 4a). Sensitivity to 850-hPa temperature (Fig. 4b) falls within a relatively small region near a thermal ridge to the southeast of the composite cyclone. Sensitivity to 500-hPa height (Fig. 4c) exhibits largest sensitivity a few hundred kilometers west of the region of maximum sensitivity to SLP and downstream of a composite trough in the height field. For periods when the average precipitation in the box is greater than 2 mm, a one standard deviation change to the SLP and 500-hPa height field in the region of largest sensitivity is predicted to change the precipitation metric by 0.4 mm, and for 850-hPa temperature by 0.3 mm; thus we conclude that, as for SLP forecasts, precipitation forecasts are less sensitive to 850-hPa temperature than SLP or 500-hPa height.

←

FIG. 3. Composite sensitivity patterns for western WA 24-h SLP forecasts (shading; hPa) to analyses of (a) SLP (hPa), (b) 850-hPa temperature (K), and (c) 500-hPa height (m). Each field represents the sensitivity multiplied by the analysis std dev at each analysis grid point for the 30 most sensitive western WA SLP forecasts between 1 Jan and 30 Jun 2005. Contours denote the composite-average ensemble-mean analysis for these 30 cases (hPa).

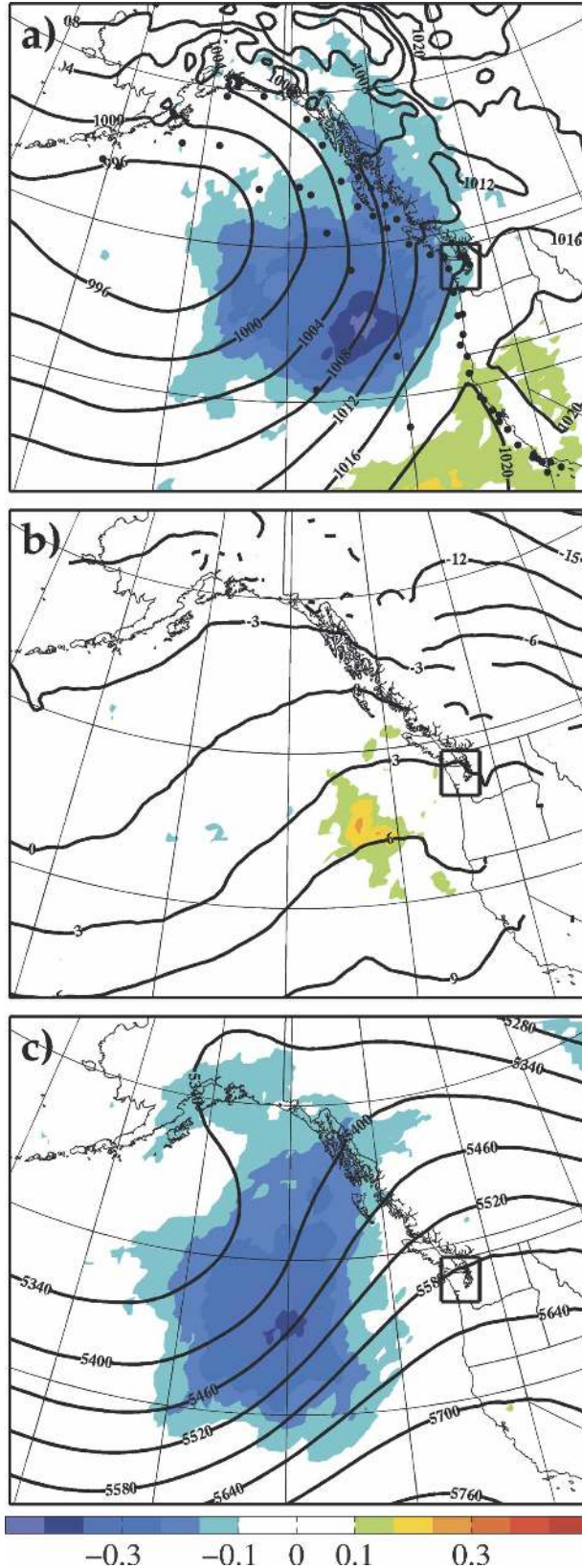


FIG. 4. As in Fig. 3, but for the 24-h forecast of precipitation (mm) for the western WA metric box.

5. Observation denial experiments for single observations

Recall from section 3 that buoy 46005 is located in a region of frequent sensitivity, but was not functional during the period considered. The change in western Washington 24-h SLP forecasts due to a missing buoy within the consistently sensitive region is assessed by withholding a nearby reliable buoy (buoy 46036, 42.3°N, 133.8°W) from the analysis and comparing the resulting forecasts. In addition to providing an estimate of the importance of offshore buoy observations, these data-denial experiments are used to quantify the accuracy of ensemble-based calculations of the change in a forecast metric.

The change in the 24-h western Washington SLP forecast associated with assimilating buoy 46036’s SLP observation is assessed for the 30 forecast cycles for which this metric is most sensitive to this buoy’s observation using the following method. A “control” analysis is generated by the assimilation method described in section 2, with the exception that buoy 46036’s SLP observation is assimilated without applying covariance localization. When covariance localization is applied, the magnitude of the ensemble estimated change is consistently larger than the actual change in the metric obtained from nonlinear forecasts. Since the forecast ensemble has no information on analysis localization, observation increments that are localized may not properly project onto the forecast metric. This contrasts with the results of Hamill and Snyder (2002), who found that predictions of analysis-error variance reduction matched the actual reduction including localization; presumably this result is due to the fact that the analysis increments are not propagated with a model.

A “no buoy” analysis is generated by the identical procedure as the “control,” but without buoy 46036’s SLP observation; therefore, the differences between these two analyses is due solely to the assimilation of buoy 46036. The change in the mean value of the forecast metric due to observation assimilation is estimated by

$$\delta J = \mathbf{J}(\mathbf{H}\mathbf{X}^b)^T(\mathbf{H}\mathbf{P}^b\mathbf{H}^T + \mathbf{R})^{-1}[\mathbf{y} - \mathcal{H}(\mathbf{x}^b)], \quad (5)$$

where \mathcal{H} is an operator that maps from state space to observation space, \mathbf{H} is a linearized version of \mathcal{H} , \mathbf{x}^b is the background ensemble-mean state vector ($N \times 1$, where N is the number of degrees of freedom in the model), \mathbf{X}^b is the $N \times M$ ensemble state matrix with the mean removed, \mathbf{P}^b is the background error covariance matrix, \mathbf{R} is the observation error covariance matrix, and \mathbf{y} is the observation values (Ancell and Hakim 2007). This equation represents linear regression,

where the independent variable is the innovation, $\mathbf{y} - \mathcal{H}(\mathbf{x}^b)$, the dependent variable is the forecast metric, and the “slope” is given by the covariance between the forecast metric and the model estimate of the observation, $\mathbf{J}(\mathbf{H}\mathbf{X}^b)^T$, divided by the covariance of the independent variables (innovation covariance). For a single observation, the innovation, innovation covariance, and slope are all scalars, and the calculation can be evaluated rapidly. When the forecast metric is a function of the forecast state vector, we shall refer to δJ as the *change in the forecast metric* associated with the observation, and when the forecast metric refers to a forecast error, we shall refer to δJ as the *observation impact*.

In addition to assessing the change in the expected value of the metric, we also assess the change in the forecast-metric variance due to observation assimilation via (Ansell and Hakim 2007)

$$\delta\sigma = -\mathbf{J}(\mathbf{H}\mathbf{X}^b)^T(\mathbf{H}\mathbf{P}^b\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{H}\mathbf{X}^b\mathbf{J}^T. \quad (6)$$

For a single observation, this expression can be evaluated as a product of two scalars: the inverse of the innovation variance, $(\mathbf{H}\mathbf{P}^b\mathbf{H}^T + \mathbf{R})^{-1}$, and the forecast-metric-observation-estimate covariance, $\mathbf{J}(\mathbf{H}\mathbf{X}^b)^T$. Furthermore, we observe that (6) is negative definite since the right-hand side is proportional to the square of the forecast-metric-observation-estimate covariance.

These predictions of δJ and $\delta\sigma$ are computed from the ensemble without the buoy and are verified against perturbed forecasts generated from an analysis where a single buoy pressure observation is withheld. We proceed by describing the change in the average SLP due to assimilating the buoy during one case characterized by an eastern Pacific cyclogenesis event, and then summarize all 30 cases.

Figure 5a shows the UW EnKF ensemble-mean SLP analysis and forecast sensitivities for 1200 UTC 5 February 2005. A frontal wave is situated on the eastern edge of a deeper cyclone near the international date line; during the next 24 h, this wave deepens as it moves east toward the North American coast. Forecast sensitivities are maximized along the eastern edge of the frontal wave near buoy 46036 (dot). Increasing (decreasing) the SLP in this region of the analysis by 1 hPa, which amounts to shifting the frontal wave to the northwest (southeast), leads to a 1.5-hPa increase (decrease) in the forecast metric.

The difference between the control and no-buoy analysis and their resulting 24-h forecast differences are shown in Figs. 5b and 5c, respectively. For the control analysis, the SLP is 0.4 hPa lower to the south of the wave and 0.2 hPa higher to the north of the wave; thus

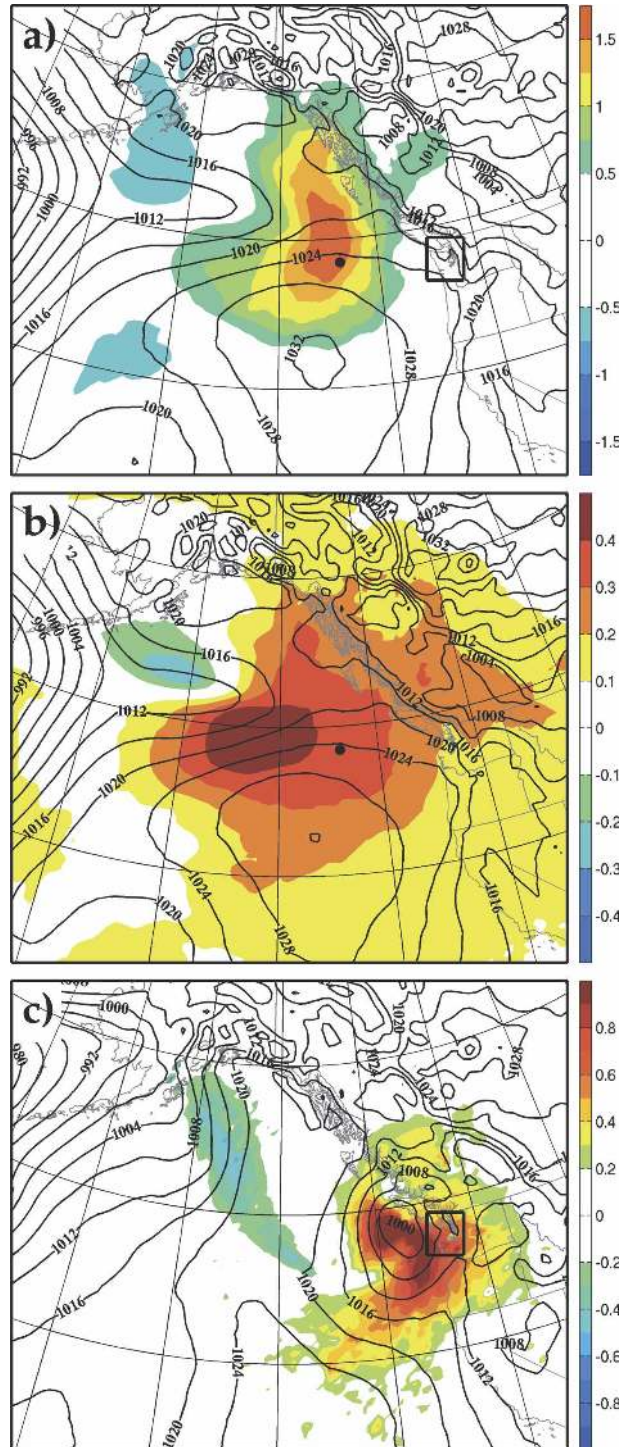


FIG. 5. (a) Sensitivity of the western WA 24-h SLP forecast to the SLP analysis (shading; hPa hPa⁻¹) and the UW EnKF ensemble-mean analysis of SLP (contours; hPa) for the forecast initialized at 1200 UTC 5 Feb 2005. (b) Difference between the no-buoy ensemble-mean analysis SLP field and the control ensemble-mean analysis SLP field at 1200 UTC 5 Feb 2005 (shading; hPa). The no-buoy ensemble-mean analysis of SLP is given by the solid lines (hPa). (c) As in (b), but for the 24-h forecast of SLP valid at 1200 UTC 6 Feb 2005.

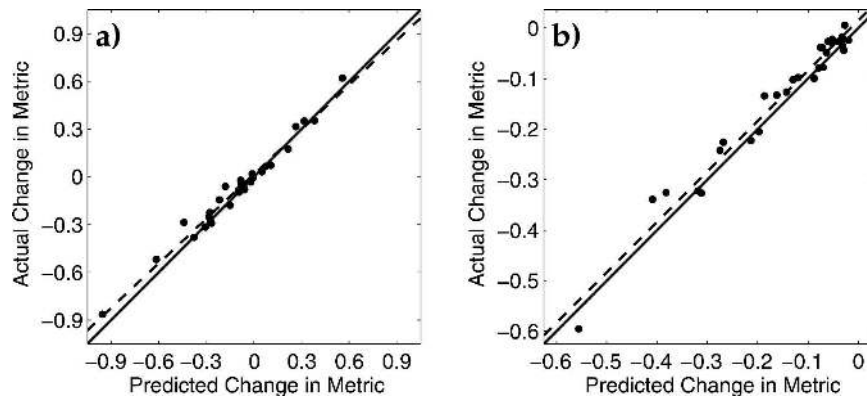


FIG. 6. Change (hPa) in the (a) expected value and (b) spread of the 24-h western WA SLP forecast due to the assimilation of buoy 46036's SLP observation as determined by the difference between two nonlinear forecasts (ordinate) and the ensemble-based sensitivity prediction (abscissa) for the 30 most sensitive forecast cycles during January–July 2005. The dashed line is the linear least squares fit to the data. Values on the main diagonal (solid line) indicate perfect agreement between the ensemble-based prediction and the WRF model solutions.

the buoy's observation shifts the wave to the south. The largest 24-h forecast differences are associated with the resulting cyclone along the Washington coast; the forecast initialized from the control analysis has SLP values that are up to 0.8 hPa lower. The ensemble-based prediction of a 0.60-hPa (0.15-hPa) decrease in the expected value (spread) of the metric compares closely with the 0.63-hPa (0.18-hPa) change obtained from the nonlinear forecasts.

Repeating the above process for the remaining 29 forecast cycles indicates that ensemble-based predictions provide accurate estimates of the changes in the forecast metric. Figure 6 shows that the ensemble-based prediction of the change in the expected value and spread is in good agreement with the actual change obtained from the nonlinear model ($R^2 = 0.985$); in 90% of cases considered, the error in the expected value (spread) is less than 0.1 hPa (0.05 hPa). Moreover, these results indicate that a buoy within the most sensitive region could produce up to a 0.8-hPa change in the expected value and a 0.5-hPa reduction in the spread of the 24-h western Washington area-averaged SLP forecasts.

6. Observation denial experiments for multiple observations

Here we extend the results of the previous section from a single observation to larger sets. These experiments are performed to quantify the accuracy of ensemble-based estimates of how multiple observations change a forecast metric and to evaluate the value of

various observation platforms in the UW EnKF system. We note that the objective here is similar to that of Langland and Baker (2004), who use an adjoint-based technique to estimate the impact of observations on the error in global forecasts.

For the remainder of this section, we consider 12-h forecast cycles and compare forecast metrics for the test case where observations are assimilated at hour 6, with the control case where no observations are assimilated at hour 6. Specifically, the change in the forecast metrics valid 6 h later due to observations assimilated at 0600 and 1800 UTC is assessed during March 2005. Observations are processed one at a time using the serial assimilation procedure described in section 2. Furthermore, rather than solve (6), the ensemble forecast-metric values are updated in the same manner as the analysis state variables. The change in ensemble spread may then be evaluated from the updated ensemble metric values, which reflect the influence of all prior observations. This method of updating the forecast-metric values with observations is similar to what is used for parameter estimation with an EnKF (e.g., Evensen 2003; Aksoy et al. 2006; Tong and Xue 2008).

Before assessing the changes in western Washington SLP forecasts due to the assimilation of all observations, we first evaluate the change in this metric due to the assimilation of select surface observations. These experiments are meant to be intermediary between the single observation experiments of the previous section and experiments using all observations that will be described later in this section. For each analysis time, the sensitivity of the forecast-metric expected value to the

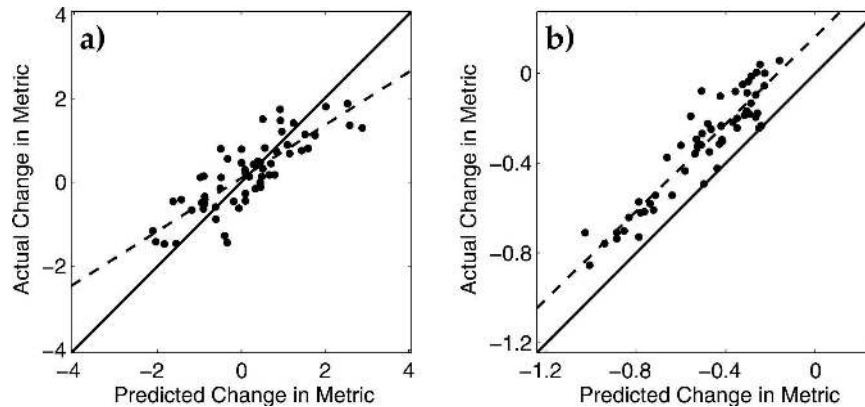


FIG. 7. Change (hPa) in the (a) expected value and (b) spread of 6-h forecasts of western Washington SLP due to assimilating all statistically significant (at the 99% confidence level) surface observations. Ensemble predictions (abscissa) and compared with results for differences between perturbed WRF forecasts (ordinate) during March 2005. Dashed lines give the linear least squares fit, while the solid line indicates perfect agreement between the ensemble-based prediction and the WRF model solution.

model estimate of each surface observation is computed using (1) and tested for statistical significance at the 99% confidence level using (2). Significant observations are assimilated and the change in the forecast metric is evaluated (≈ 20 observations per cycle). Ensemble-based predictions of δJ and $\delta\sigma$ are verified by advancing the resulting analysis ensemble forward 6 h using the WRF model.

Results show good agreement between the ensemble predictions and WRF verification of δJ and $\delta\sigma$ (Fig. 7). The correlation between the predicted and actual δJ and $\delta\sigma$ is 0.82 and 0.87, respectively. Whereas the bias in δJ is small, the ensemble-based estimate of $\delta\sigma$ is consistently larger than the actual value by 0.19 hPa. Differences between the predicted and actual change are due to sampling error and nonlinearity; sampling error is discussed further in the conclusions section.

The change in the western Washington SLP forecast metric due to the of all observations available to the UW EnKF system is now assessed. For each analysis time, all observations are assimilated (≈ 3700 per cycle); however, the estimated change due to an individual observation is computed only if the sensitivity to the model estimate of the observation is significant at the 99% confidence interval (≈ 100 per cycle). This confidence interval is found to give the best agreement between the predicted and actual values for δJ and $\delta\sigma$. When statistically insignificant observations are included in the calculation, spurious covariances increase the RMS difference between the predicted and actual values by a factor of 2 or more (not shown). For these calculations, covariance localization is achieved using the Gaspari and Cohn localization, which reduces to

zero 5000 km from the observation. This broad localization function prevents observations from adjusting state variables at unreasonable distances; similar results are obtained when covariance localization is not applied. For simplicity, covariance inflation is not considered in these experiments.

Figures 8a and 8b indicate that the predicted change in western Washington SLP due to observations is in good agreement with the actual difference. On average, observations change the expected value of this metric by 0.86 hPa and reduce the spread by 0.59 hPa. Whereas the correlation between the predicted and actual δJ and $\delta\sigma$ is 0.49 and 0.93 respectively, there is more scatter about the line of perfect agreement when compared to Fig. 7. In contrast to the select surface observation results, the ensemble-based prediction of $\delta\sigma$ here exhibits little bias.

We performed a second test to address how the size of the forecast-metric box affects the results obtained above. The change in the forecast of average SLP in a box over the western North American coast (NAC; region given by the larger box in Fig. 1a) due to observations is determined by repeating the procedure used for the western Washington SLP metric. Ensemble-based predictions of the change in NAC show comparable skill to the results obtained for western Washington SLP (Figs. 8c,d). The correlation between the predicted and actual δJ and $\delta\sigma$ is 0.42 and 0.71, respectively. Observations produce a slightly smaller change in the expected value (0.75 hPa) and spread (0.37 hPa) of the NAC SLP metric as compared to the western Washington SLP metric because there is less variability in SLP when averaged over a larger area.

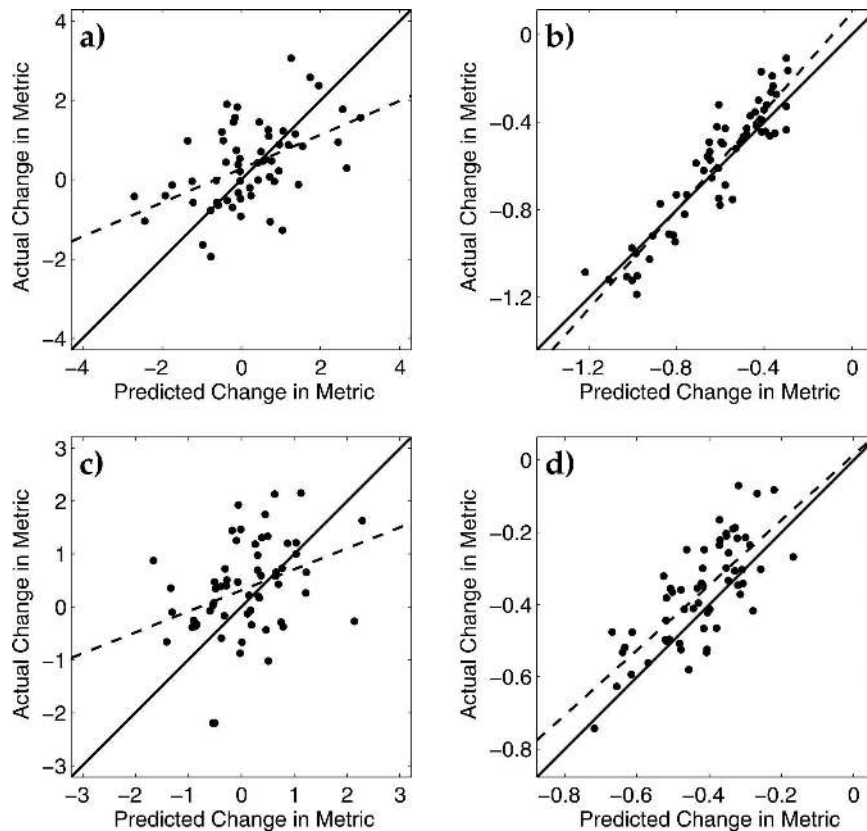


FIG. 8. Change (hPa) in the (a) expected value and (b) spread of 6-h forecasts of western Washington SLP forecasts due to assimilating all available observations. Ensemble predictions (abscissa) are compared with results for differences between perturbed WRF forecasts (ordinate) during March 2005. Dashed lines give the linear least squares fit, while the solid line indicates perfect agreement between the ensemble-based prediction and the WRF model solution. (c), (d) Similar to (a) and (b), but applied to the average SLP within the larger NAC box (see Fig. 1).

7. Impact of observations on forecast verification

Although the impact of observations on a forecast metric is well predicted by this technique, this does not guarantee that the observations actually improve the forecasts. As a consequence, we repeat the ensemble sensitivity procedure to assess how observations impact the RMS error in SLP within the western Washington and NAC regions. In this case, J is the RMS error in the box and, unlike the previous calculations, this metric can only be evaluated when an analysis is available for verification. An ensemble of RMS error values within each box is determined based on each ensemble member's forecast verified against the appropriate ensemble-mean analysis; negative values of δJ indicate that observation assimilation decreases the RMS error.

Figures 9a and 9b indicate that ensemble-based impact predictions for western Washington SLP error have skill comparable to the western Washington aver-

age SLP forecast metric; the correlation between the predicted and actual δJ and $\delta\sigma$ is 0.48 and 0.79, respectively. On average, assimilating observations reduces the RMS error in Washington SLP forecasts by 0.67 hPa. Comparable results are obtained for the RMS error in SLP over the NAC region. In this case, the correlation between the predicted and actual δJ (0.42) and $\delta\sigma$ (0.71) are similar to the results obtained from the average NAC SLP metric. For both regional metrics, there are multiple cycles where the ensemble-based estimate of δJ is off by at least 1 hPa. Each of these forecasts is characterized by a cyclone undergoing rapid cyclogenesis or cyclolysis near the edge of the respective box, which suggests sensitivity to the cyclone position.

The ensemble-based observation impact estimates are partitioned by observation type to determine which observations have the largest impact on the RMS error in SLP forecasts. Figure 10 shows the probability den-

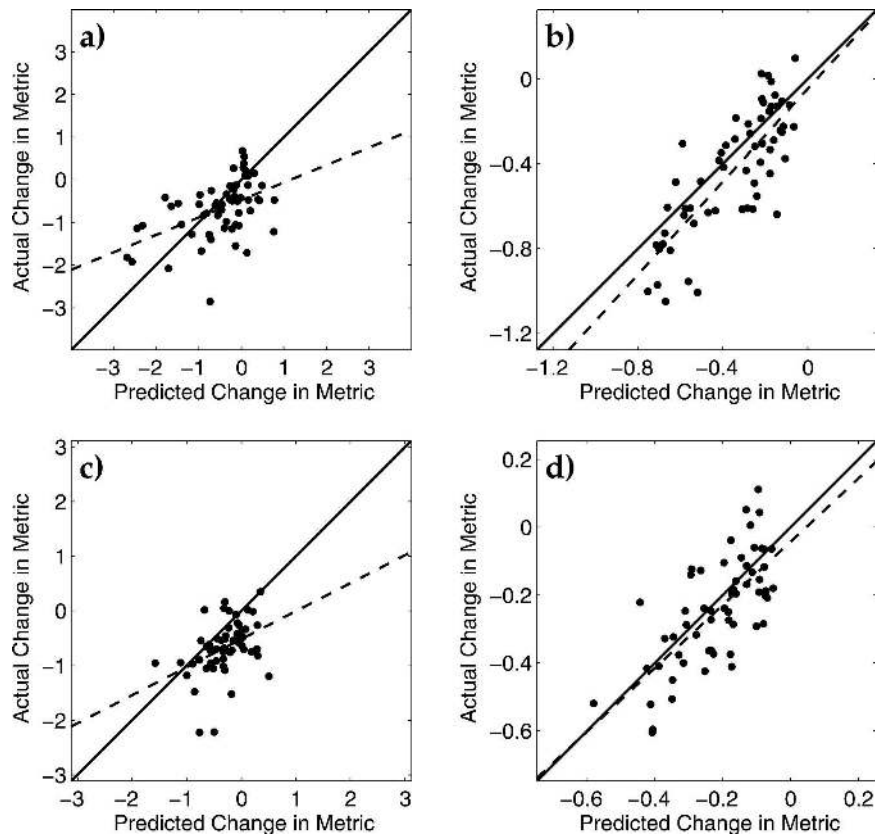


FIG. 9. As in Fig. 8, but for the RMS error in SLP (hPa) forecasts within the (a), (b) western WA region and (c), (d) NAC region valid 6 h later.

sity functions (PDFs) for the impact of the individual statistically significant observations (99% confidence) on the expected value and spread of the RMS error in western Washington SLP forecasts. The results for the RMS error in NAC SLP forecasts are qualitatively similar and are not shown. For all observation types, the PDFs are sharply peaked near zero, which indicates that most observations have little impact on SLP errors within this region. These PDFs are qualitatively similar to histograms for observation impacts on forecast errors in a quasigeostrophic channel model discussed in Morss and Emanuel (2002; see their Fig. 8). The long tails in the surface observation PDFs indicate that these observations are more likely to have a large impact on the SLP error when compared to ACARS and cloud winds. In Figs. 10c and 10e, the PDFs are symmetric about zero, which implies that ACARS and cloud-wind observations are equally likely to have a positive or negative impact on the RMS error in SLP. In contrast, the surface observation PDF (Fig. 10a) is skewed toward negative values, indicating that surface observations are more likely to reduce, rather than increase, the RMS error in SLP. Moreover, the surface observa-

tions having highest impact are taken by buoys located approximately 500 km offshore, which is consistent with earlier results indicating this is a region of high sensitivity (not shown). It should be noted that the impact of different observation types may change depending on the metric chosen. While ACARS and cloud-wind observations may not have a significant impact in the SLP metrics, these observation types may have larger impact on metrics that measure upper-tropospheric forecasts.

8. Discussion and conclusions

The ensemble sensitivity technique described by Hakim and Torn (2008) and Ancell and Hakim (2007) is tested using data drawn from a pseudo-operational ensemble Kalman filter. Ensemble analyses and forecasts from January to July 2005 are used to determine locations of frequent sensitivity for selected forecast metrics near western North America. The skill of ensemble sensitivity analysis in predicting the change in a forecast-metric mean, δJ , and variance, $\delta\sigma$, due to observation assimilation is also tested. Although this

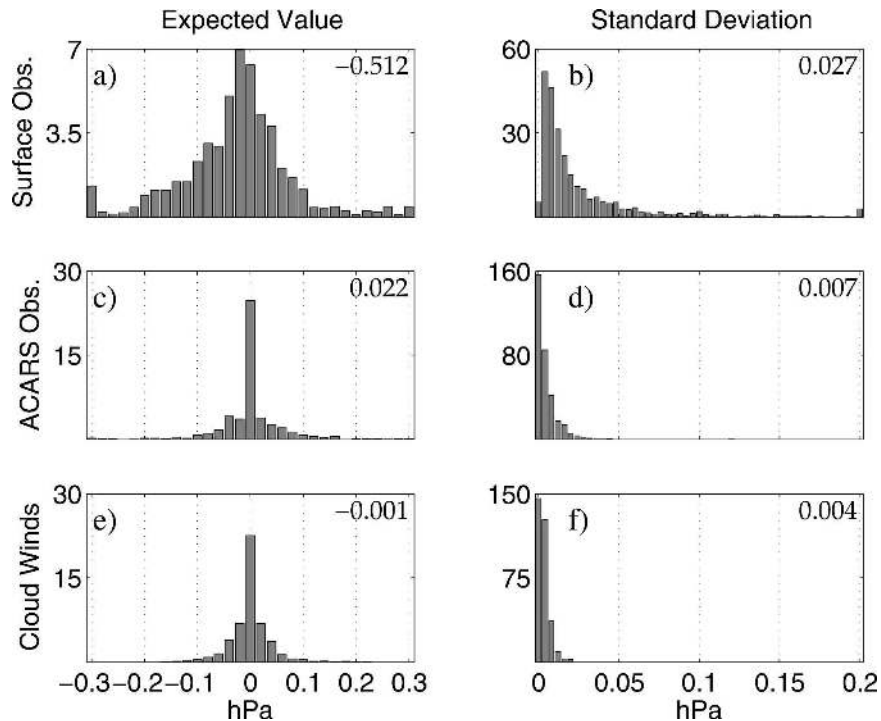


FIG. 10. PDFs (hectopascals per cycle) of the impact of individual statistically significant (99% confidence) (top) surface, (middle) ACARS, and (bottom) cloud-wind observations assimilated at 0600 and 1800 UTC on the (left) expected value and (right) spread on the RMS error in SLP forecasts within the western WA region valid 6 h later during March 2005. The value at the top of each panel indicates the average impact of each observation type during a data assimilation cycle.

study focuses mainly on metrics near the west coast of North America, we emphasize that the technique is general and may be applied to any scalar forecast metric.

Climatological sensitivity fields for 24-h western Washington SLP and precipitation forecasts are most often sensitive to the upstream mass field and to a lesser extent the temperature field. While a large fraction of the frequently sensitive region is observed by the fixed buoy network, the buoy closest to the sensitivity maximum was not functioning during the period and thus could have adversely affected western Washington forecasts. Composite patterns for the most sensitive forecasts indicate that the region of largest sensitivity for 24-h western Washington SLP and precipitation forecasts is located approximately 1000 km west of the metric box and exhibits modest upshear tilt in the vertical. A one-standard-deviation change in the most sensitive region of the mass field would produce a larger change in both SLP and precipitation metrics as compared to a one standard deviation change in the most sensitive region of the temperature field, which suggests that targeted buoy and ship SLP observations

could produce the largest change in short-range, surface-based forecast metrics in this area.

The change in western Washington SLP forecasts due to removing a buoy from the region of frequent sensitivity is evaluated for the 30 most sensitive cases. Removing the buoy pressure observation from the assimilation process yields a perturbed forecast metric that is compared with the prediction from ensemble sensitivity analysis. For all forecast cycles, the ensemble-based estimate is in good agreement with the actual change obtained from perturbed nonlinear model forecasts. These results indicate that a single SLP observation within the region of consistent sensitivity can change 24-h western Washington area-averaged SLP forecasts by up to 0.8 hPa and reduce the forecast spread by 0.5 hPa.

The single observation calculations are extended to estimate the change in 12-h forecasts associated with assimilating a large number of observations. Approximately 100 observations per analysis time produce a statistically significant change on the forecast-metric mean value at the 99% confidence level. Therefore, this approach attempts to predict the change in the forecast-

metric mean and variance with approximately 100 observations from the several thousand observations that are assimilated. An attractive attribute of this approach is that it can be applied “offline” to an existing dataset of ensemble analyses and forecasts without running the model or cycling a data assimilation system. Results show that the ensemble-based estimates provide a relatively accurate prediction of observation changes to the spread of the forecast metric, although agreement is not as good as for the case of a single observation. Similar results are found for a forecast metric that covers a much larger portion of western North America, suggesting that the results are not limited to metrics covering small geographical areas.

We also tested ensemble sensitivity for predicting forecast error in a manner similar to the case just summarized. For an error metric defined by the root-mean-square error in a box over western Washington state, results again show that ensemble sensitivity provides accurate estimates of the reduction in error spread, and to a lesser extent the error expected value. Partitioning the error estimates by observation type indicates that surface observations are more likely to reduce the error in SLP forecasts when compared to ACARS and cloud-wind observations. In particular, the fixed-position buoys 500 km from the North American coast have the greatest impact because they are in a region of consistent sensitivity. We note that the impact of these observation platforms may vary depending on the forecast metric, season, model, and particular observation set.

For the experiments and metrics considered here, it appears that ensemble sensitivity analysis is more reliable in predicting changes in the spread (variance) of a forecast metric when compared to the mean value of the metric. We propose that, in the absence of significant nonlinearity and model error, this difference may be understood through the effect of sampling error on these calculations. The key quantity in Eqs. (5) and (6) is the covariance between the forecast metric and the model estimate of an observation, $\mathbf{J}(\mathbf{H}\mathbf{X}^b)^T$, which affects the predicted changes in the metric mean and spread differently. If sampling error for this covariance has zero mean, then the predicted changes to the metric mean value will also be unbiased, and the scatter about the line of perfect prediction will be proportional to the sampling-error variance. In contrast, the error in the predicted change in the metric *spread* is proportional to the sampling error *variance*, which introduces a bias even when the sampling error itself is unbiased (see Fig. 7b); essentially, sampling error leads to an overprediction in the variance reduction. Using a confidence test on the covariance, as we have done here, reduces the effect of sampling error, but also limits the number of

observations that are included in the calculation. A point of diminishing returns is reached when, for confidence levels approaching 100%, important observations are excluded and thus the ability to predict the impact on a forecast metric is adversely affected. For the experiments considered here, this point of diminishing returns is reached around the 99% confidence level.

The results presented here suggest that ensemble sensitivity analysis provides an attractive alternative to adjoint sensitivity analysis. In addition, the results suggest that this technique may prove useful for observation thinning, where a large sample of observations is reduced to a set that is expected to decrease forecast-metric spread the most, while also producing a statistically significant change in the forecast-metric mean value. Unlike previously proposed observation thinning algorithms (e.g., Liu and Rabier 2002; Ochotta et al. 2005), this technique selects observations based on the forecast metric of interest. Similarly, ensemble sensitivities may be useful for observation targeting because the impact of a hypothetical observation on the forecast-metric variance can be determined prior to knowing the observation value and, unlike other targeting techniques, it considers the data assimilation scheme and analysis-error statistics (Berliner et al. 1999; Langland 2005). Ensemble sensitivity analysis may also prove useful for selecting observations based on their predicted effect on forecast-error variance in previous forecasts.

Acknowledgments. We thank Chris Velden and Dave Stettner of CIMSS/SSEC for providing cloud winds for the UW EnKF system and Rolf Langland of the U.S. Naval Research Laboratory (Monterey) for discussions on forecast error. We are also grateful to Tom Hamill and Jeff Anderson for constructive comments in review that improved the manuscript. This study was made possible in part because of the data that were made available to the Earth System Research Laboratory/Global Systems Division by the following commercial airlines: American, Delta, Federal Express, Northwest, United, and United Parcel Service. This work was supported by NSF Grant ITR-0205648, NOAA CSTAR Grant NA17RJ1232, and ONR Grant N00014-06-1-0510.

REFERENCES

- Aksoy, A., F. Zhang, and J. W. Nielsen-Gammon, 2006: Ensemble-based simultaneous state and parameter estimation in a two-dimensional sea-breeze model. *Mon. Wea. Rev.*, **134**, 2951–2970.
- Ancell, B., and G. J. Hakim, 2007: Comparing adjoint- and ensemble-sensitivity analysis with applications to observation targeting. *Mon. Wea. Rev.*, **135**, 4117–4134.

- Anderson, J. L., 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.*, **129**, 2884–2903.
- Berliner, M. L., Z.-Q. Lu, and C. Snyder, 1999: Statistical design for adaptive weather observations. *J. Atmos. Sci.*, **56**, 2536–2552.
- Ek, M. B., K. E. Mitchell, Y. Lin, E. Rodgers, P. Grunmann, V. Koren, G. Gayno, and J. D. Tarpley, 2003: Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta Model. *J. Geophys. Res.*, **108**, 8851, doi:10.1029/2002JD003296.
- Errico, R. M., and T. Vukicevic, 1992: Sensitivity analysis using an adjoint of the PSU-NCAR mesoscale model. *Mon. Wea. Rev.*, **120**, 1644–1660.
- Evensen, G., 2003: The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dyn.*, **53**, 343–367.
- Gaspari, G., and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.*, **125**, 723–757.
- Hakim, G. J., 2003: Developing wave packets in the North Pacific storm track. *Mon. Wea. Rev.*, **131**, 2824–2837.
- , and R. D. Torn, 2008: Ensemble synoptic analysis. *Sanders Symposium Monograph, Meteor. Monogr.*, No. 55, Amer. Meteor. Soc., in press.
- Hamill, T. M., and C. Snyder, 2002: Using improved background-error covariances from an ensemble Kalman filter for adaptive observations. *Mon. Wea. Rev.*, **130**, 1552–1572.
- , —, and R. E. Morss, 2002: Analysis-error statistics of a quasi-geostrophic model using three-dimensional variational assimilation. *Mon. Wea. Rev.*, **130**, 2777–2790.
- Hong, S.-Y., J. Dudhia, and S.-H. Chen, 2004: A revised approach to ice microphysical processes for the bulk parameterization of clouds and precipitation. *Mon. Wea. Rev.*, **132**, 103–120.
- Hoskins, B. J., R. Buizza, and J. Badger, 2000: The nature of singular vector growth and structure. *Quart. J. Roy. Meteor. Soc.*, **126**, 1565–1580.
- Janjic, Z. I., 2002: Nonsingular implementation of the Mellor–Yamada level 2.5 scheme in the NCEP Meso model. NCEP Office Note 437, National Centers for Environmental Prediction, Camp Springs, MD, 61 pp.
- Kain, J. S., and J. M. Fritsch, 1990: A one-dimensional entraining detraining plume model and its application in convective parameterization. *J. Atmos. Sci.*, **47**, 2784–2802.
- Khare, S. P., and J. L. Anderson, 2006: A methodology for fixed observational network design: Theory and application to a simulated global prediction system. *Tellus*, **58A**, 523–537.
- Langland, R. H., 2005: Issues in targeted observing. *Quart. J. Roy. Meteor. Soc.*, **131**, 3409–3425.
- , and N. L. Baker, 2004: Estimation of observation impact using the NRL atmospheric variational data assimilation adjoint system. *Tellus*, **56A**, 189–201.
- , R. L. Elsberry, and R. M. Errico, 1995: Evaluation of physical processes in an idealized extratropical cyclone using adjoint sensitivity. *Quart. J. Roy. Meteor. Soc.*, **121**, 1349–1386.
- Liu, Z.-Q., and F. Rabier, 2002: The interaction between model resolution, observation resolution and observation density in data assimilation: A one-dimensional study. *Quart. J. Roy. Meteor. Soc.*, **128**, 1367–1386.
- McMurdie, L., and C. Mass, 2004: Major numerical forecast failures in the northeast Pacific. *Wea. Forecasting*, **19**, 338–356.
- Morss, R. E., and K. A. Emanuel, 2002: Influence of added observations on analysis and forecast errors: Results from idealized systems. *Quart. J. Roy. Meteor. Soc.*, **128**, 285–321.
- Ochotta, T., C. Gebhardt, D. Saupe, and W. Wergen, 2005: Adaptive thinning of atmospheric observations in data assimilation with vector quantization and filtering methods. *Quart. J. Roy. Meteor. Soc.*, **131**, 3427–3437.
- Rabier, F., E. Klinker, P. Courtier, and A. Hollingsworth, 1996: Sensitivity of forecast errors to initial conditions. *Quart. J. Roy. Meteor. Soc.*, **122**, 121–150.
- Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, W. Wang, and J. G. Powers, 2005: A description of the Advanced Research WRF Version 2. NCAR Tech. Note 468+STR, National Center for Atmospheric Research, Boulder, CO, 88 pp.
- Snyder, C., and F. Zhang, 2003: Assimilation of simulated Doppler radar observations with an ensemble Kalman filter. *Mon. Wea. Rev.*, **131**, 1663–1677.
- Tong, M., and M. Xue, 2008: Simultaneous estimation of microphysical parameters and atmospheric state with radar data and ensemble square-root Kalman filter. Part II: Parameter estimation experiments. *Mon. Wea. Rev.*, in press.
- Torn, R. D., G. J. Hakim, and C. Snyder, 2006: Boundary conditions for limited-area ensemble Kalman filters. *Mon. Wea. Rev.*, **134**, 2490–2502.
- Velden, C., and Coauthors, 2005: Recent innovations in deriving tropospheric winds from meteorological satellites. *Bull. Amer. Meteor. Soc.*, **86**, 205–223.
- Whitaker, J. S., and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, **130**, 1913–1924.
- , —, X. Wei, Y. Song, and Z. Toth, 2008: Ensemble data assimilation with the NCEP Global Forecast System. *Mon. Wea. Rev.*, **136**, 463–482.
- Wilks, D. S., 2005: *Statistical Methods in the Atmospheric Sciences*. Elsevier Academic, 648 pp.
- Zou, X., Y.-H. Kuo, and S. Low-Nam, 1998: Medium-range prediction of an extratropical oceanic cyclone: Impact of initial state. *Mon. Wea. Rev.*, **126**, 2737–2763.